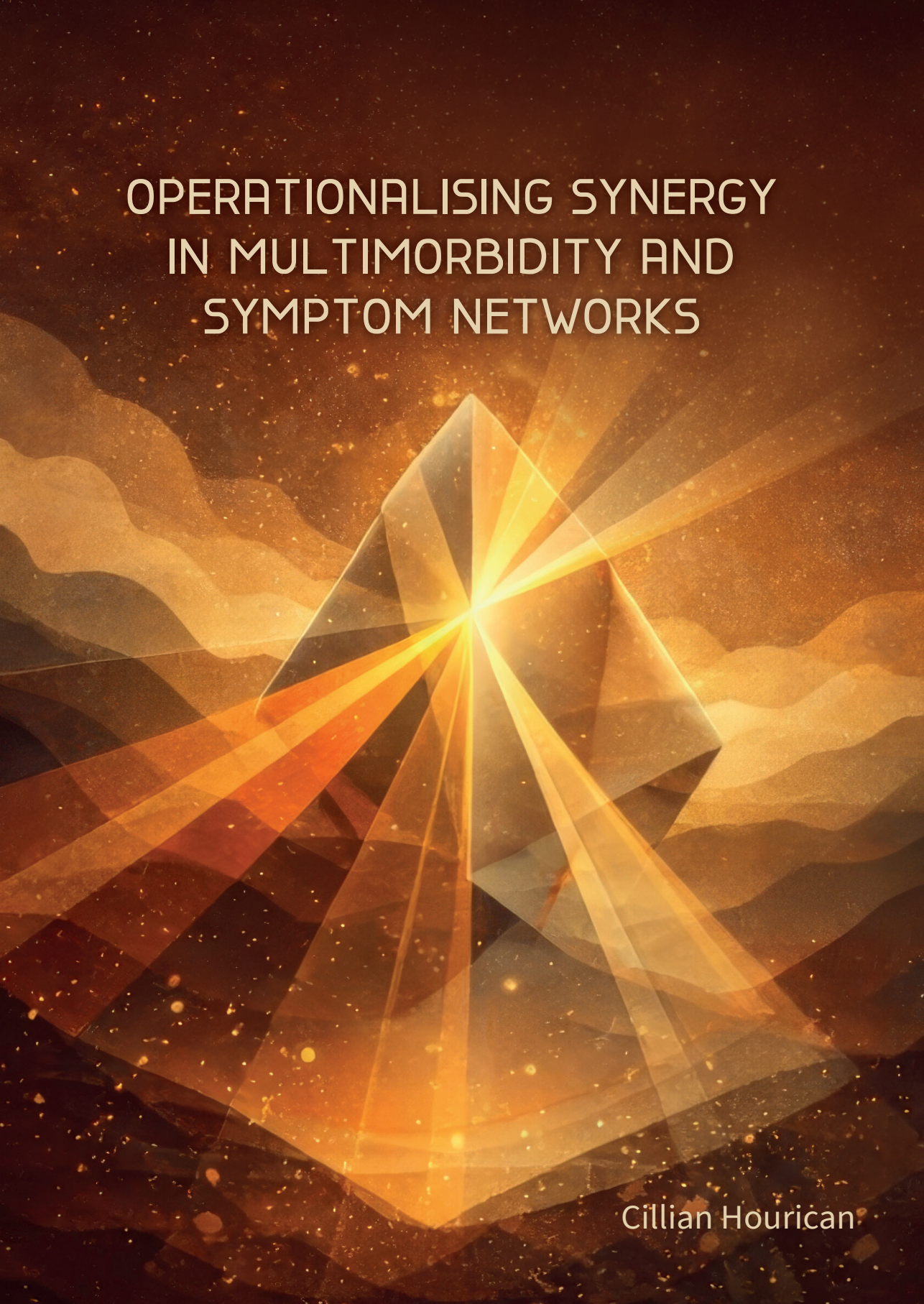
OPERATIONALISING SYNERGY IN MULTIMORBIDITY AND SYMPTOM NETWORKS



Cillian Hourican

Operationalising synergy in multimorbidity and symptom networks

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Aula der Universiteit
op vrijdag 27 november 2026, te 11.00 uur

door

Cillian Joseph Hourican

geboren te Dublin

Promotiecommissie

Promotores	dr. M.H. Lees	University of Amsterdam
	prof. dr. M.G.M. Olde Rikkert	Radboudumc
Copromotores	dr. R. Quax	University of Amsterdam
	dr. G.M.E.E. Peeters	Radboudumc
Overige leden	dr. J.A. Bosch	University of Amsterdam
	prof. dr. ir. C.I. Sánchez Gutiérrez	University of Amsterdam
	dr. F.A. Nóbrega Santos	University of Amsterdam
	prof. dr. J.A. Ellis-Monaghan	University of Amsterdam
	dr. D.L. Vetrano	Karolinska Institutet
	prof. dr. D. Marinazzo	Ghent University
Faculteit	Faculteit der Natuurwetenschappen, Wiskunde en Informatica	

This work was supported by the Netherlands Organisation for Health Research and Development (ZonMw) under Open Competition Grant 09120012010063. R. Quax was also supported by the EU project TO_AITION, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 848146.







Copyright © 2026 by C. Hourican
Cover and chapter opening pages designed by Herik Media.
ISBN 978-94-6534-814-8
An electronic version of this dissertation is available at `https://dare.uva.nl/`.

Contents

Summary **vii**

Samenvatting **xi**

1 Introduction **5**
1.1 From dyads to higher-order structure . . . 8
1.2 Synergy vs redundancy . . . 10
1.3 Research challenges and questions . . . 13
1.4 Thesis outline . . . 15
1.5 Broader impact of this thesis . . . 15

2 A Case for Higher-Order Interactions in Multimorbidity Networks **21**
2.1 Introduction . . . 21
2.2 Methods . . . 22
2.3 Results . . . 26
2.4 Discussion . . . 29

3 Efficient Search Algorithms for Identifying Synergistic Triplets in Health Data **35**
3.1 Introduction . . . 35
3.2 Background . . . 37
3.3 Methods . . . 38
3.4 Results . . . 45
3.5 Discussion . . . 63
3.6 Conclusion . . . 65

4 Decomposing the Informational Roles of Symptoms, Signs, and Diseases in Functional Limitation **69**
4.1 Introduction . . . 69
4.2 Methods . . . 70
4.3 Results . . . 72
4.4 Discussion . . . 74

5 From Triplet Catalogues to Interpretable Archetypes of Synergistic Dependence **81**
5.1 Introduction . . . 81
5.2 Results . . . 83
5.3 Discussion . . . 90
5.4 Summary of findings . . . 97
5.5 Methods . . . 98

6 Embedding-Based Partial Information Decomposition for Symptom Networks at System Scale **115**
6.1 Introduction . . . 115
6.2 Methods . . . 119
6.3 Results . . . 132
6.4 Discussion . . . 142
6.5 Conclusion . . . 151

7 Discussion **155**
7.1 What synergy and redundancy can (and cannot) tell us . . . 156
7.2 Implications and clinical translation. . . . 160
7.3 Limitations and boundary conditions . . . 167
7.4 Future directions . . . 171
7.5 Toward adaptive, data-rich multimorbidity research . . . 173
7.6 Conclusion . . . 174

A Supplementary Material for Chapter 2 **179**

B Supplementary Material for Chapter 3 **187**

C Supplementary Material for Chapter 4 **207**

D Supplementary Material for Chapter 5 **215**

E Supplementary Material for Chapter 6 **241**

Bibliography **277**

Acknowledgements **307**

List of Publications **313**

Summary

Multimorbidity is the co-occurrence of two or more chronic conditions in the same person. Multimorbidity has become the norm rather than the exception in ageing populations. More than half of adults over sixty-five now live with multiple diagnoses. On paper, two patients carrying the same number of conditions may appear comparable; in daily life they often do not. One remains independent while the other struggles with basic tasks. The difference is not because a single disease dominates. The difference is due to symptoms, signs, behaviours, and conditions interacting differently depending on context. A diagnosis list states what a patient has. It does not state what a patient can do, how their burden will evolve, or which factors drive their current limitations. This mismatch between the tidy categories of disease classification and the messy reality of lived function is the starting point of this thesis: Multimorbidity is not merely an accumulation of diseases. It is an interacting system. Clinically, the connections among diseases may matter more than any single condition.

Most analytic tools used to study these systems still rely on pairwise summaries: association networks, conditional-dependence graphs, and additive models. Each factor contributes independently in such representations. These representations capture a great deal. However, they are structurally unable to express when informative patterns emerge only when variables act together. Consider fatigue, poor sleep, and low mood: each symptom on its own may be only weakly predictive of functional decline, yet their co-occurrence can be strongly predictive in a way that no pair of these symptoms achieves alone. Information of this kind, which is available only from the combination and not recoverable from any subset, is called *synergy*. It stands in contrast to *redundancy*, in which multiple variables overlap and partly repeat the same signal about an outcome. *Partial Information Decomposition* (PID) is a framework from information theory. PID makes this distinction formal and quantitative. Given a set of source variables (e.g. symptoms, signs, and diagnoses) and a target (daily functioning), PID separates the joint information into components that are redundant, unique to each source, or synergistic. This thesis develops and applies PID-based methods to identify, interpret, and scale the analysis of higher-order interactions in multimorbidity and symptom systems.

The first contribution of this thesis is to establish why higher-order analysis is needed and how it can be made computationally feasible. Chapter 2 uses synthetic Bayesian networks. Systems are constructed so that their true higher-order structure is known in advance. The network serves as a reference standard. These networks demonstrate that pairwise association networks can both miss critical dependencies and mischaracterise the expected consequences of interventions when the underlying system contains higher-order interactions. The same approach is then applied to clinical data from the

EASYcare study. Symptoms, signs, and diseases are combined within a single network. This identifies 184 synergistic triplets represented using a hypergraph. A hypergraph is a network in which edges can connect more than two nodes. Chapter 3 addresses a practical barrier: combinatorial explosion, the fact that the number of possible variable combinations grows unmanageably with system size. This scaling problem makes exhaustive evaluation infeasible for realistic datasets. The chapter develops stochastic search heuristics, namely simulated annealing and particle swarm optimisation. These search heuristics discover high-synergy sets one to two orders of magnitude faster than exhaustive enumeration. It also identifies a previously underappreciated artefact, termed the echo problem, in which redundancy and overlap among variables inflate the apparent number of higher-order findings. A filtering step is proposed to control this inflation before results are interpreted.

The second contribution turns from method to clinical application. It first considers a single endpoint and then extends the analysis across a whole variable set. Chapter 4 applies PID to symptoms, signs, and diseases measured in the Longitudinal Aging Study Amsterdam (LASA) with functional limitations as the target outcome. In a four-variable decomposition, no domain contributed unique information in the total sample. In the multimorbidity subgroup, symptoms were the only domain to contribute unique information. Symptoms accounted for 34.1% of the total information about functional limitations. This information was not carried by signs or diseases. This is consistent with the view that disease labels become informationally saturated as conditions accumulate. Symptom-level assessment may therefore add value precisely where it is most needed. Chapter 5 scales this logic to a set of seventeen LASA variables. It introduces the *BUST framework*: Breadth, Uniformity, Synergy, and Total. The framework characterises each variable pair along four dimensions. *Breadth* captures whether a pair's synergy is distributed across many outcomes (generalist) or concentrated on a few (specialist). *Uniformity* captures whether synergy remains stable across age, sex, and multimorbidity strata or is context-dependent. *Synergy* quantifies the magnitude of the information available only when the variables are considered together. *Total* records overall dependence. The BUST map was applied across 2,040 variable-pair and target triplets in LASA. It summarised the 135 source pairs that carried synergy. This reveals three findings: First, the ten pairs ranked highest by the synergy-based composite share no members with the ten highest-ranked pairs by total mutual information. The two quantities are nevertheless weakly and positively correlated across all pairs ($r = 0.31$). Second, synergy is concentrated in clinically recognisable cross-domain constellations. These constellations link sleep, pain, mood, physical activity, and body composition. Third, synergy profiles are context dependent. For some pairs, synergy is distributed differently between participants with and without multimorbidity. For other pairs, the distribution differs across sex or age strata. Building on these profiles, the thesis assigns clinical features distinct informational roles. These roles are general indicators, disambiguators, or synergy gates. Each role reflects how a feature's information is structured relative to different targets and contexts.

The third contribution extends PID from local variable sets to full symptom networks.

It does so through approximation rather than exact decomposition. When every other symptom serves as context, a complete multivariate decomposition becomes infeasible. The number of information atoms grows faster than exponentially with the number of sources. Chapter 6 addresses this barrier through embedding-based PID (ePID). The method compresses all other symptoms into a single summary variable that preserves information most relevant to the target. A tractable two-source PID is then applied to the focal symptom, the compressed summary, and the target. Compression uses an Agglomerative Conditional Information Bottleneck. This groups the joint states of the remaining symptoms into a small number of categories that are chosen to retain target-relevant structure. The method is benchmarked against synthetic networks with three to five sources, where the ground-truth decomposition is known. The embedding tracks synergistic information closely ($r = 0.92$). The empirical applications embed a larger remainder than this validated range. The reported atom fractions should therefore be read as approximate. In UK Biobank and the Xinxiang student sample, the PHQ-9 depression data show that most pair-to-target dependence is shared with the remainder. This occurs through two channels: redundancy and remainder-unique information. Redundancy alone accounts for 45 to 51 per cent of the decomposed information. Synergy constitutes a small component. This highlights specific symptom configurations where interaction-aware modelling may be needed rather than additive approaches alone. The composition of information channels replicates across cohorts. Spearman correlations are 0.59 for redundancy, 0.68 for remainder-unique information, and 0.78 for synergy. This supports the generalisability of the decomposition. It also suggests that higher-order structure in depressive symptom networks is not an artefact of a single sample. The same decomposition protocol is applied to a contrasting instrument, the 28-item Interpersonal Reactivity Index. Here, the pipeline returns a synergy-dominated decomposition. This indicates that embedding-based PID registers synergy dominance where it is present rather than imposing either pattern.

These three contributions form an integrated framework whose progression from motivation through discovery, application, interpretation, and scaling reflects the practical arc of making higher-order dependence a computable object rather than a conceptual curiosity. The common thread is an *information-theoretic weighting and interpretation programme* for clinical systems. Each chapter adds a layer: a demonstration that pairwise representations are structurally incomplete, search heuristics that make synergy discoverable without exhaustive enumeration, structured summaries that compress high-dimensional interaction landscapes into interpretable maps, and embedding-based approximations that extend the decomposition to full symptom networks. The thesis begins from the mismatch between disease lists and lived function; the framework developed here offers a way to make that hidden structure visible, quantifiable, and clinically interpretable. Rather than replacing existing analytic tools, the approach complements them by exposing a mode of dependence that additive and pairwise representations cannot express by construction. Important limitations remain: all analyses are cross-sectional and observational, PID decompositions depend on the choice of redundancy measure and

the included variable set, and treatment processes, which shape the symptom landscape in multimorbidity populations, are not included as sources or conditioning variables.

The broader significance of this work lies in making context visible and computable in precisely those settings where measurements span multiple clinical domains and decisions must be made for individual patients. In such settings, the operative question is not whether a variable is associated with an outcome, but whether additional measurements provide genuinely new information for a given person or subgroup, or whether they largely repeat existing signal. By providing a principled way to separate redundancy from synergy, and by developing computational scaffolding that makes these quantities discoverable and interpretable at scale, the thesis contributes a step towards a more operational notion of personalised medicine: not merely predicting outcomes, but understanding which combinations of measurements are informative, for whom, and under what conditions. The methods introduced here have been tested on ageing-cohort, cardiovascular-risk, metabolomic, psychiatric-symptom, and psychometric data across several independent samples, namely LASA, the Cardiovascular Risk in Young Finns Study, the UK Biobank metabolomics panel, UK Biobank mental-health data, and the Xinxiang student sample, together with an online sample completing the Interpersonal Reactivity Index; whether they generalise to other clinical domains and data modalities remains an empirical question. The present work provides the conceptual and computational foundation on which that broader validation can proceed.

Samenvatting

Multimorbiditeit is het gelijktijdig aanwezig zijn van twee of meer chronische aandoeningen bij dezelfde persoon. In vergrijzende bevolkingen is het eerder regel dan uitzondering geworden: meer dan de helft van de vijf-en-zestig-plussers leeft met meerdere diagnoses. Op papier lijken twee patiënten met evenveel aandoeningen vergelijkbaar; in het dagelijks leven zijn ze dat vaak niet. De een blijft zelfstandig, de ander komt de dag nauwelijks door. Dat verschil ontstaat niet doordat één ziekte de overhand heeft, maar doordat symptomen, klinische tekenen, gedrag en aandoeningen anders op elkaar inwerken afhankelijk van de context. Een diagnoselijst vermeldt wat een patiënt heeft. Ze zegt niet wat iemand kan, hoe de ziektelast zich zal ontwikkelen, of welke factoren de huidige beperkingen veroorzaken. Die kloof tussen de ordelijke categorieën van de ziekte-classificatie en de complexe werkelijkheid van het dagelijks functioneren is het vertrekpunt van dit proefschrift. Multimorbiditeit is niet enkel een opeenstapeling van ziekten. Het is een systeem van interacties. Klinisch gezien kunnen de verbanden tussen aandoeningen zwaarder wegen dan elke afzonderlijke aandoening.

De meeste methoden waarmee zulke systemen worden bestudeerd, steunen nog altijd op paarsgewijze samenvattingen: associatienetwerken, conditionele-afhankelijkheidsgrafen en additieve modellen. Daarin levert elke factor een eigen, losstaande bijdrage. Zulke representaties leggen veel vast, maar ze kunnen per constructie niet uitdrukken dat een patroon pas informatief wordt wanneer variabelen samen optreden. Neem vermoeidheid, slecht slapen en somberheid. Elk symptoom apart voorspelt maar zwak functionele achteruitgang; samen voorspellen ze sterk, op een manier die geen enkel tweetal eruit haalt. Informatie die alleen uit de combinatie beschikbaar is en niet uit enige deelverzameling valt te herleiden, heet *synergie*. Daartegenover staat *redundantie*: meerdere variabelen overlappen en herhalen deels hetzelfde signaal over een uitkomst. *Partial Information Decomposition* (PID) is een framework uit de informatietheorie dat dit onderscheid formeel en kwantitatief maakt. Gegeven een set bronvariabelen, bijvoorbeeld symptomen, klinische tekenen en diagnoses, en een doelvariabele zoals dagelijks functioneren, splitst PID de gezamenlijke informatie in componenten die redundant zijn, uniek voor één bron, of synergetisch. Dit proefschrift ontwikkelt PID-methoden en past ze toe om hogere-orde-interacties in multimorbiditeits- en symptoomsystemen te identificeren, te interpreteren en op schaal te analyseren.

De eerste bijdrage laat zien waarom hogere-orde-analyse nodig is en hoe ze rekenkundig haalbaar wordt. Hoofdstuk 2 gebruikt synthetische Bayesiaanse netwerken: systemen die zo zijn gebouwd dat hun werkelijke hogere-orde-structuur vooraf vaststaat, zodat ze als ijkpunt kunnen dienen. Daarmee blijkt dat paarsgewijze associatienetwerken kritieke afhankelijkheden kunnen missen en de verwachte gevolgen van interventies

verkeerd kunnen inschatten zodra het onderliggende systeem hogere-orde-interacties bevat. Dezelfde aanpak is vervolgens toegepast op klinische gegevens uit de EASYcare-studie, waarin symptomen, klinische tekenen en ziekten in één netwerk samenkomen. Dat leverde 184 synergetische geordende drietallen op, weergegeven als hypergraaf: een netwerk waarin één verbinding meer dan twee knopen kan koppelen. Hoofdstuk 3 pakt een praktisch obstakel aan, de combinatorische explosie. Het aantal mogelijke variabelencombinaties groeit onbeheersbaar met de omvang van het systeem, waardoor volledig zoeken bij realistische datasets onhaalbaar is. Het hoofdstuk ontwikkelt stochastische zoekheuristieken, namelijk simulated annealing en particle swarm optimisation, die sets met hoge synergie één tot twee ordes van grootte sneller vinden dan volledige enumeratie. Daarnaast beschrijft het een tot dan toe onderbelicht onderwerp, het echo-probleem, waarbij redundantie en overlap tussen variabelen het schijnbare aantal hogere-orde-bevindingen opblazen. Een filterstap houdt die inflatie in toom voordat de resultaten worden geïnterpreteerd.

De tweede bijdrage verschuift van methode naar klinische toepassing: eerst voor één uitkomstmaat, daarna over een volledige variabelenset. Hoofdstuk 4 past PID toe op symptomen, klinische tekenen en ziekten uit de Longitudinal Aging Study Amsterdam (LASA), met functionele beperkingen als doelvariabele. In een decompositie met vier variabelen droeg in de totale steekproef geen enkel domein unieke informatie bij. In de subgroep met multimorbiditeit deden symptomen dat als enige. Zij waren goed voor 34,1% van de totale informatie over functionele beperkingen, informatie die klinische tekenen en ziekten niet droegen. Dat sluit aan bij het idee dat diagnoselabels informationeel verzadigd raken naarmate aandoeningen zich opstapelen. Beoordeling op symptoomniveau kan dus juist daar meerwaarde hebben waar die het hardst nodig is. Hoofdstuk 5 schaalt deze redenering op naar zeventien LASA-variabelen met het *BUST-framework*: Breadth, Uniformity, Synergy en Total. Het karakteriseert elk variabelenpaar langs vier dimensies. *Breadth* geeft aan of de synergie van een paar over veel uitkomsten is verspreid (generalist) of zich op enkele concentreert (specialist). *Uniformity* geeft aan of de synergie stabiel blijft over leeftijd, geslacht en multimorbiditeit, of juist context-afhankelijk is. *Synergy* kwantificeert hoeveel informatie pas beschikbaar komt wanneer de variabelen samen worden beschouwd. *Total* registreert de totale afhankelijkheid.

Het BUST-framework is toegepast op 2.040 combinaties van een variabelenpaar en een doelvariabele in LASA en vat de 135 bronparen samen die synergie droegen. Daaruit komen drie bevindingen naar voren. Ten eerste delen de tien paren die het hoogst scoren op de synergiemaat geen enkel lid met de tien hoogst scorende paren op totale wederzijdse informatie. Toch hangen beide grootheden over alle paren zwak positief samen (r = 0,31). Ten tweede concentreert synergie zich in klinisch herkenbare, domeinoverschrijdende constellaties rond slaap, pijn, stemming, lichamelijke activiteit en lichaamssamenstelling. Ten derde zijn synergieprofielen context-afhankelijk. Bij sommige paren verschilt de verdeling tussen deelnemers met en zonder multimorbiditeit, bij andere tussen geslachts- of leeftijdsgroepen. Op basis van deze profielen kent het proefschrift klinische kenmerken verschillende informationele rollen toe: algemene

indicatoren, onderscheidende indicatoren en synergie-poorten. Elke rol weerspiegelt hoe de informatie van een kenmerk zich verhoudt tot uiteenlopende doelvariabelen en contexten.

De derde bijdrage breidt PID uit van losse variabelensets naar volledige symptoomnetwerken, en doet dat bij benadering in plaats van via exacte ontleding. Zodra elk ander symptoom als context dient, wordt een volledige multivariate decompositie onhaalbaar: het aantal informatie-atomen groeit sneller dan exponentieel met het aantal bronnen. Hoofdstuk 6 neemt dat obstakel met embedding-gebaseerde PID (ePID). De methode comprimeert alle overige symptomen tot één samenvattende variabele die de voor de doelvariabele meest relevante informatie behoudt. Daarna volgt een hanteerbare PID met twee bronnen: het focale symptoom, de samenvattende variabele en het doel. De compressie verloopt via een Agglomerative Conditional Information Bottleneck, die de gezamenlijke toestanden van de overige symptomen samenvoegt tot een klein aantal categorieën, gekozen om doelrelevante structuur te behouden. De methode is vervolgens getoetst aan synthetische netwerken met drie tot vijf bronnen, waarvan de werkelijke decompositie bekend is. De embedding volgt de synergetische informatie daar nauwgezet ($r = 0{,}92$). De empirische toepassingen bedden een grotere restverzameling in dan dit gevalideerde bereik omvat. De gerapporteerde atoomfracties zijn daarom bij benadering. In UK Biobank en de Xinxiang-studentensteekproef laten de PHQ-9-gegevens over depressiesymptomen zien dat het grootste deel van de afhankelijkheid tussen paar en doel wordt gedeeld met de rest van het netwerk. Dat verloopt via twee kanalen: redundantie en rest-unieke informatie. Redundantie alleen al is goed voor 45 tot 51 procent van de ontlede informatie. Synergie vormt een klein onderdeel. Ze wijst wel specifieke symptoomconfiguraties aan waar modellering die rekening houdt met interactie nodig kan zijn, in plaats van uitsluitend additieve benaderingen. De samenstelling van de informatiekanalen repliceert over cohorten heen, met Spearman-correlaties van 0,59 voor redundantie, 0,68 voor rest-unieke informatie en 0,78 voor synergie. Dat ondersteunt de generaliseerbaarheid van de decompositie en wijst erop dat de hogere-orde-structuur van depressieve symptoomnetwerken geen toevalsbevinding is van één steekproef. Hetzelfde protocol is toegepast op een sterk afwijkend instrument, de Interpersonal Reactivity Index met 28 items. Daar levert de werkwijze juist een door synergie gedomineerde decompositie op. Dat laat zien dat embedding-gebaseerde PID synergiedominantie registreert waar die aanwezig is, in plaats van het ene of het andere patroon op te leggen.

Samen vormen deze drie bijdragen een geïntegreerd framework. De opbouw, van motiveren via ontdekken, toepassen en interpreteren naar opschalen, weerspiegelt de praktische weg waarlangs hogere-orde-afhankelijkheid zich van een conceptuele notie tot een berekenbaar object ontwikkelt. De rode draad is een *informatietheoretisch weeg- en interpretatieprogramma* voor klinische systemen. Elk hoofdstuk voegt een laag toe: de demonstratie dat paarsgewijze representaties structureel incompleet zijn, zoekheuristieken die synergie vindbaar maken zonder volledige enumeratie, gestructureerde samenvattingen die hoogdimensionale interactielandschappen tot leesbare kaarten comprimeren, en embedding-gebaseerde benaderingen die de ontleding uitbreiden naar volledige symp-

toomnetwerken. Het proefschrift begint bij de kloof tussen diagnoselijsten en dagelijks functioneren. Het hier ontwikkelde framework maakt die verborgen structuur zichtbaar, meetbaar en klinisch interpreteerbaar. De benadering vervangt bestaande analytische methoden niet, maar vult ze aan door een vorm van afhankelijkheid bloot te leggen die additieve en paarsgewijze representaties per constructie niet kunnen uitdrukken. Er blijven belangrijke beperkingen: alle analyses zijn cross-sectioneel en observationeel; PID-decomposities hangen af van de gekozen redundantiemaat en van de opgenomen variabelen; en behandelingen, die het symptoomlandschap bij multimorbiditeit mede vormgeven, zijn niet als bron- of conditioneringsvariabele meegenomen.

De bredere betekenis van dit werk ligt in het zichtbaar en berekenbaar maken van context, juist daar waar metingen meerdere klinische domeinen beslaan en er beslissingen moeten worden genomen voor individuele patiënten. De vraag is dan niet of een variabele samenhangt met een uitkomst, maar of een extra meting werkelijk nieuwe informatie oplevert voor deze persoon of subgroep, of anders vooral herhaalt wat al bekend is. Door redundantie en synergie op principiële wijze te scheiden, en door de rekeninfrastructuur te bouwen die deze grootheden op schaal vindbaar en interpreteerbaar maakt, zet dit proefschrift een stap naar een meer operationele opvatting van gepersonaliseerde geneeskunde: niet alleen uitkomsten voorspellen, maar begrijpen welke combinaties van metingen informatief zijn, voor wie, en onder welke omstandigheden. De methoden zijn beproefd op gegevens over verouderingscohorten, cardiovasculair risico, metabolomica, psychiatrische symptomen en psychometrie, in meerdere onafhankelijke steekproeven: LASA, de Cardiovascular Risk in Young Finns Study, het metabolomicapanel en de gegevens over psychische gezondheid van UK Biobank, de Xinxiang-studentensteekproef en een online steekproef die de Interpersonal Reactivity Index invulde. Of ze ook generaliseren naar andere klinische domeinen en datatypen blijft een empirische vraag. Dit werk levert het conceptuele en computationele fundament waar die bredere validatie op kan voortbouwen.



Chapter 1

INTRODUCTION

A symptom becomes meaningful only in context. As a patient's subjective report of bodily change, its significance emerges only in relation to other clinical information. Fatigue can reflect sleep disruption, medication side-effects, inflammation, low mood, cardiopulmonary limitation, or deconditioning; dizziness can signal vestibular disease, orthostatic intolerance, anxiety, or polypharmacy. Crucially, the plausibility of each interpretation shifts the moment a second symptom, a medication list, or a diagnosis enters the picture. A symptom becomes clinically informative not only by what it is, but by *what it appears alongside* (another symptom, a medication, a diagnosis), and at times, a particular combination can carry clinical meaning that none of its parts carries alone.

This relational view of symptoms is historically recent. Clinical practice began with the ailment: a collection of subjective experiences such as *the vapors* or *the shakes*. When physiology was largely inaccessible, symptoms were not merely clues to disease; they *were* disease [1]. The underlying pathophysiology remained a black box, and the clinician's only instrument was the patient's own account.

Nosology (the systematic classification of diseases into named categories) fundamentally transformed the medical landscape [2]. By grouping recurring symptom patterns into named syndromes, early classifiers turned a heterogeneous clinical environment into a structured programme of study [3].

The subsequent laboratory revolution added measurable signs, biomarkers, and pathological findings to this framework [4]. This shifted the *clinical gaze* from what the patient felt (subjective symptoms) to what the clinician could observe and quantify [5].

This drive for precision produced the International Classification of Diseases (ICD)—a taxonomy so successful that it became the "grammar" through which clinical reality is expressed [2]. However, as critics of the traditional biomedical model note, it remains a grammar of isolated *entities* rather than a system that accounts for the complex *interactions* among them [6, 7].

For much of the twentieth century, this framework served clinical practice well: one patient, one dominant condition, one diagnostic pathway. However, in ageing populations where multimorbidity—the co-occurrence of two or more chronic conditions in the same individual—has become the norm rather than the exception, affecting more than half of adults over 65 [8, 9], the mapping becomes many-to-many once more: the same symptom can be incidental in one constellation of symptoms and diagnoses, and consequential in another. In the Netherlands, 2024 primary-care data show that prevalence rises steeply with age: from 65% among those aged 65–69 to over 90% among those aged 85 and older [10]. A diagnostic label remains necessary but often insufficient: it does not tell us whether this particular patient can still climb stairs, whether their fatigue will worsen, or whether adjusting one medication will destabilise an otherwise stable condition.

The success of diagnosis-driven medicine has also shaped healthcare systems around separable disease entities. The gains have been real and substantial: identifiable mechanisms, targeted therapies, and reproducible care protocols. Yet it also means that symptoms are often treated primarily as pointers to a single culprit, rather than as integrative signals that reflect interactions among conditions, treatments, and behaviours. For an older adult living with multiple different conditions, a symptom such as breathlessness or weight loss is rarely attributable to one dominant driver; it is more often the product of an interaction. In geriatric medicine, this is known as *multicausality*: the principle

that functional decline in the presence of multimorbidity is almost never attributable to a single pathway [11].

Multimorbidity is most commonly operationalised as simple arithmetic: a count of two, three, or even five diagnoses [12, 13]. Weighted indices and disease clusters refine that count, but they still summarise the disease list rather than the interactions within it. For the older patient, such a summary is an abstraction: what determines whether they can walk, socialise, or remain independent is rarely a single label but rather the way symptoms, signs, and behaviours interact across those labels [9, 14]. Consider two older adults with the same number of chronic conditions. On paper, they look similar; in life, they do not. One remains independent while the other struggles with daily functioning, and the difference is contingent on context (e.g., symptoms, signs, diseases) rather than on any dominant diagnosis.

This everyday mismatch between diagnosis lists and lived function is the starting point of this thesis. Multimorbidity is not an accumulation of diseases but an interacting system, or network, in which symptoms, signs, and behaviours are nodes alongside diagnoses [8, 15]. What matters clinically may lie in the connections that cross diagnostic boundaries rather than in any single condition. This view is increasingly recognised in geriatric and multimorbidity research, but the quantitative tools for expressing and analysing such interactions beyond pairwise associations remain underdeveloped [15, 16]. This thesis takes steps towards closing that gap.

This structural perspective is not new to clinical practice. Clinical care has long recognised that diseases do not simply add up. Disease-specific guidelines were largely built for single conditions, and their direct combination in the presence of multiple conditions can have undesirable or even contradictory consequences, especially when treatment burden, polypharmacy, and competing outcomes enter the picture [17–19]. For older adults with multimorbidity, expert guidance therefore emphasises patient-centred goals and functional outcomes; an implicit acknowledgment that the disease list alone does not determine what matters or what to do [20, 21].

This clinical perspective motivates the computational one: if function and wellbeing are the targets of care, then the analytic representations should be able to express how multiple signals combine, overlap, and condition one another, not merely list them as independent risk factors. The tools introduced in this thesis are designed to fill that gap: to quantify which combinations carry information that none of their parts carry alone, and to do so at the scale of modern cohort data.

Translating that clinical intuition into a formal framework requires a language for multivariate dependence. In computational modelling, a variable can be informative in isolation; it can repeat information already present in other variables; or it can become informative only in combination with other variables [22]. When the latter occurs, the relevant signal is not located in any single clinical indicator but emerges from a specific configuration: no subset of clinical indicators is sufficient, yet the combination is. If the goal is to make person- and subgroup-specific inferences, then the value of an additional clinical indicator is conditional: what it adds depends on what is already known. The

methodological question is how to describe such multivariate dependence in a way that is quantitative, model-agnostic (not tied to any specific statistical or machine-learning model, though the estimation step requires discretisation choices that can influence results), and interpretable, while still remaining tractable at cohort scale.

Answering this question matters because the direction of contemporary healthcare is towards individualised decision-making under complexity [23, 24]. Personalised medicine is often framed as "the right treatment for the right patient at the right time" [25], but achieving that goal requires more than simply adding more measurements. It requires recognising when and how information is distributed across combinations of variables—and, consequently, when the value of an additional measurement depends on what has already been observed. Much of the necessary data already exists in principle, since large cohorts and biobanks routinely capture symptoms, diagnoses, biomarkers, and behavioural factors at scale, often longitudinally [26, 27]. The empirical chapters draw on several such resources: the EASYcare cohort of older persons in Dutch general practice (Chapter 2), UK Biobank NMR metabolomics and the Cardiovascular Risk in Young Finns Study [28] (Chapter 3), the Longitudinal Aging Study Amsterdam (LASA), which measures symptoms, signs, diseases, and daily functioning in community-dwelling older adults (Chapters 4–5), and item-level depressive-symptom responses from UK Biobank and a large Chinese student sample (Chapter 6). Their coverage and granularity vary, and they were typically designed around single-disease paradigms rather than multi-domain interaction. The difficulty lies in the summaries built from them, which remain dyadic, formed from variables taken two at a time. Pairwise association networks, additive models, and conditional-dependence graphs—where an edge means two variables remain associated after adjusting for all other observed variables—dominate the interpretive layer. Downstream prediction models (the prediction models that are ultimately fitted to the data) may capture interactions implicitly, but the summaries we report and act on rarely do [29, 30]. Flexible prediction models can *exploit* higher-order structure for accuracy, but they do not *reveal* it in a form that supports clinical interpretation, comparison across subgroups, or identification of which combinations are informative versus redundant. Bridging that gap—providing not just predictive power but interpretive resolution—is the contribution this thesis seeks to make. When relationships are irreducibly multivariate, dyadic summaries can miss key dependencies, misattribute effects to individual variables, and suggest interventions that behave unexpectedly because the system reacts through higher-order pathways that no pairwise summary captures. To give a concrete example: in pharmacoepidemiology, a drug–drug interaction may appear safe when each medication is evaluated pairwise with patient characteristics, yet become hazardous in the presence of a third factor. For instance, co-trimoxazole and ACE inhibitors each cause mild potassium retention on their own, but their co-prescription in a patient with reduced renal function can produce fatal hyperkalaemia [31]. Pairwise screening can detect the drug pair, but the hazard is conditional on renal function, and no pairwise summary can express that three-way dependency. Such scenarios are not hypothetical, because polypharmacy and impaired renal function are both common among older

adults [19].

The gap is not only conceptual but methodological and operational; researchers and clinicians alike recognise that interactions matter [32]. We lack a practical, scalable, and interpretable approach for quantifying higher-order dependence in health data and for separating overlapping information from together-only information at cohort scale. Extending dyadic analysis to higher orders runs into combinatorial explosion (the number of possible variable combinations grows unmanageably with dimension), estimator fragility (estimates become statistically unreliable when data are sparse relative to the number of combinations tested), and an interpretation bottleneck: even when higher-order effects can be computed, they do not naturally compress into a clinically legible picture. This thesis addresses that gap by operationalising higher-order dependence through information-theoretic synergy and by developing computational methods that make synergy discoverable, interpretable, and scalable in multimorbidity and symptom networks.

The thesis begins by demonstrating that dyadic representations can fail in clinically consequential ways, then develops scalable strategies for discovering higher-order structure. Subsequent chapters establish the clinical relevance of synergy for daily functioning and map higher-order dependence across multiple outcomes and population strata. The final empirical chapter extends Partial Information Decomposition (PID—a framework for separating shared, unique, and together-only information) to full symptom networks through computationally efficient approximations.

The specific research questions are formulated in Section 1.3, following the introduction of synergy and redundancy. Before presenting these tools, however, we must examine the current representational landscape of multimorbidity. This reveals where existing approaches become structurally incomplete, demonstrating that the shift from pairwise to higher-order analysis is not merely a modelling preference. Rather, it is a necessary response to a specific class of dependencies that dyadic summaries cannot express.

1.1. From dyads to higher-order structure

The preceding sections argued that multimorbidity turns clinical reasoning into a problem of interpreting interactions rather than separating independent conditions. A patient may present with breathlessness, cognitive complaints, chronic pain, and reduced appetite, while carrying several diagnostic labels that each offer a partial explanation of the combination of these symptoms. In such cases, the practical question moves beyond the traditional "which single disease causes this symptom?" to "which configuration of signs, symptoms, behaviours, and diseases explains the current functional burden and predicts what will happen next?" That shift in perspective is subtle but important: it moves the object of interest from a list of conditions to an interacting system, in which context determines what a feature means.

Most established approaches to multimorbidity and symptom burden still operate with dyads as their basic unit. Even when downstream predictive models are interaction-

capable, the summary we report back to clinicians is still usually dyadic [33]. In epidemiology and gerontology, multimorbidity is frequently summarised using summary indices (such as weighted disease counts) or clustering approaches that compress heterogeneous disease constellations into a small number of risk strata [12, 34]. Network approaches add structure by representing diseases (or symptoms) as nodes and connecting them by pairwise co-occurrence or association, while symptom network analysis in psychiatry often uses conditional-dependence graphs to separate direct from indirect dyadic relations [29, 30, 35]. These tools are useful because they are estimable at scale, easy to visualise, and often stable under noise and missingness [30, 36]. They have shaped how the field describes multimorbidity at the population level.

The limitation is not that dyadic representations are "wrong", but that they are structurally incomplete for a class of questions that matter in multimorbidity research. A dyadic edge can tell us that two variables tend to co-occur or remain associated after adjustment, but it cannot represent situations in which the informative structure only appears when variables are considered together. This is the interaction-only regime, and it is particularly common in multimorbidity, where symptoms become less specific, pathways overlap, and different combinations can lead to similar overall burden. In that regime, a purely dyadic picture can be misleading in two ways: it can miss interaction-only structure entirely, and it can encourage interpretations in which risk and impairment are attributed to individual variables rather than to the combination of co-occurring conditions, symptoms, and treatments in which they operate.

Complex-systems research offers natural representational targets for such regimes. Higher-order formalisms generalise standard graphs by allowing multi-body relations, making it possible, in principle, to represent that a set of variables forms a meaningful unit of dependence [32, 37]. In a hypergraph, a single edge can connect three or more nodes, while simplicial complexes are the subclass in which every subset of an included set must itself be included. In clinical terms, a hyperedge represents a synergistic observation: for example, that fatigue, poor sleep, and low mood jointly predict functional decline in a way that no constituent pair can. However, the primary obstacle in biomedical applications is not a lack of formalisms, but the absence of a weighting and interpretation logic that is (i) data-driven rather than hand-specified, (ii) computationally feasible without exhaustive enumeration, and (iii) interpretable at the system level rather than only for isolated motifs.

A hyperedge that connects multiple variables is not yet an explanation; it becomes scientifically useful only when it is attached to a quantitative definition of what the joint structure contributes. This should then be embedded in a workflow that can discover, validate, and summarise such structures under cohort-scale constraints.

This is where information theory becomes a natural bridge between clinical motivation and computational feasibility. Information-theoretic dependence measures are model-agnostic and can capture nonlinear relationships without committing to a particular regression form [38]. More importantly for the present thesis, they provide a principled way to ask what is gained by combining variables, which is exactly the question

that dyadic summaries cannot answer directly. The practical challenge is that multivariate information analysis naively scales poorly and can be fragile in finite samples, so adopting an information-theoretic framing is not sufficient on its own [39]. What is needed is a computational pipeline that makes higher-order dependence discoverable without brute-force search, interpretable in a way that supports comparison across outcomes and subgroups, and credible through explicit robustness and validation logic.

Why use information-theoretic decomposition and not an alternative framework? Shapley-value approaches can attribute contributions to individual variables, but they do not separate redundant from synergistic contributions. Topological data analysis can detect higher-order structure, but its outputs are geometric rather than information-theoretic and do not naturally decompose into interpretive components tied to a specific outcome. Higher-order regression terms can be added, but their interpretation depends on scale and functional form. Partial Information Decomposition is preferred here because it provides a semantically interpretable decomposition that is target-relative—matching the clinical need to ask "informative *for what?*"—and because it connects naturally to both estimation theory and the existing network literature [22, 40].

The focus of this thesis is, therefore, complementary to the broader development of higher-order network formalisms. Rather than introducing higher-order objects as a modelling choice and leaving their meaning implicit, the thesis develops an information-theoretic weighting and interpretation programme—centred on separating overlap from complementarity—that turns higher-order structure into a measurable and clinically interpretable object in multimorbidity and symptom systems. With this motivation in place, the next section introduces the core interpretive distinction that runs throughout the thesis: how redundancy and synergy capture two different ways in which multivariate information can be organised, and why that distinction matters for both scientific interpretation and personalised, context-aware modelling.

1.2. Synergy vs redundancy

In day-to-day clinical reasoning, the value of an additional measurement is rarely judged by whether it is "associated" with an outcome in isolation. It is judged by whether it *changes* what can be inferred for a particular patient once everything else that is already known is taken into account. Sometimes a new symptom report, biomarker, or diagnostic label is essentially a second view of the same underlying signal: it may confirm, stabilise, or simplify decision-making, but it does not fundamentally add new information. At other times, the extra measurement is only useful *in combination* with what is already known: it sharpens a risk estimate, disambiguates two plausible mechanisms, or separates subgroups that would otherwise be conflated. This distinction—between repeated signal and context-dependent complementarity—is what motivates the concepts of *redundancy* and *synergy* as organising principles in this thesis.

1.2.1. What is synergy?

In this thesis, *synergy* refers to information about a chosen *target* (for example, daily functioning, a symptom, or a clinical state) that becomes available only when multiple sources are considered *jointly* and is not recoverable from any source alone. In other words, synergy formalises a "together-only" regime: each variable on its own may be weakly informative, yet their combination is strongly informative about the target. This is a higher-order phenomenon in a precise information-theoretic sense: it is defined with respect to a target and the joint distribution of the observed variables, not with respect to any particular parametric model [22, 41, 42]. Formally, given sources X_1 and X_2 and a target Y, the synergistic component is the portion of $I(X_1, X_2; Y)$ that exceeds what can be attributed to redundancy or either source's unique contribution.

This information-theoretic notion should be distinguished from more familiar uses of the word "interaction." A regression interaction term is tied to a chosen functional form and scale, and it can be present or absent depending on modelling choices. In contrast, synergy is model-agnostic, as it is grounded in information theory: it asks whether the joint observation of sources reduces uncertainty about the target in a way that cannot be reconstructed from the sources separately [40, 41]. Synergy is not a mechanistic or causal claim. In observational health data, it is best treated as a descriptor of dependence structure—a way of locating where context is part of the signal—and therefore as a guide for representation, interpretation, and model choice.

Synergy provides a quantitative language for context-dependent informativeness: it highlights situations in which the informational value of a second measurement depends on the first, so that "what to measure next" is not a universal rule but a context-sensitive one. In that sense, synergy can be read as a value-of-information concept (a formalisation of the question: how much would obtaining this additional measurement be worth?): it points to combinations of symptoms, signs, behaviours, or diseases that are jointly informative for outcomes that matter to patients, and it encourages analysis that asks not only *what predicts*, but *what predicts for whom and under what conditions*.

1.2.2. What is redundancy?

Redundancy captures the complementary case: overlap in what multiple sources convey about the target. If two variables are redundant with respect to an outcome, then observing both provides less additional information than one might expect from looking at each in isolation, because they partly "say the same thing" about the target [43, 44]. Redundancy is not synonymous with the correlation between sources. Two sources can be correlated but still provide different information about a target, and two sources can be weakly correlated yet nevertheless overlap strongly in the particular aspect of the target they predict. Redundancy is therefore inherently target-dependent: it is overlap in *predictive content about a specific target*, not a generic similarity measure between variables [41, 45].

In clinical and cohort settings, redundancy is not merely a nuisance. It can reflect meaningful measurement structure: repeated indicators of the same underlying physio-

logical domain, robustness through multiple proxies, or substitutability in assessment design. From a personalised-medicine viewpoint, redundancy can be beneficial because it supports parsimony and resilience: if two measures are largely redundant for a target, one may be sufficient in settings where burden, cost, or feasibility constrain what can be collected. The key point is that redundancy and synergy answer different operational questions. Redundancy speaks to substitutability and confirmation; synergy speaks to complementarity and context. Both are required if the aim is not only to build predictive models but also to understand how the information in a multimodal (combining different measurement types such as symptoms, biomarkers, and behavioural assessments) health dataset is organised—what overlaps, what complements, and what is jointly indispensable.

1.2.3. Making synergy operationalisable

Making this distinction quantitative starts from the standard information-theoretic quantities of entropy, mutual information, and conditional mutual information. Mutual information asks how much knowing one variable reduces uncertainty about another, for example how much sleep quality tells us about daily functioning. What it cannot say is whether two sources carry largely overlapping information about a target, or information that is unlocked only when they are combined. That is the interpretive gap PID addresses [22, 40]. In the bivariate setting, PID decomposes the total information that two sources (X_1, X_2) carry about a target Y into redundant, unique, and synergistic components:

$$I(X_1, X_2; Y) = \mathrm{Red}(X_1, X_2; Y) + \mathrm{Unq}(X_1; Y \setminus X_2) + \mathrm{Unq}(X_2; Y \setminus X_1) + \mathrm{Syn}(X_1, X_2; Y). \quad (1.1)$$

Here $\mathrm{Unq}(X_1; Y \setminus X_2)$ is the information that X_1 carries about Y and X_2 does not. It is not a conditional mutual information: $I(X_1; Y \mid X_2) = \mathrm{Unq}(X_1; Y \setminus X_2) + \mathrm{Syn}(X_1, X_2; Y)$, so conditioning on X_2 leaves the unique and synergistic parts bundled together. The decomposition makes the "added value of combining measurements" explicit: it separates what is duplicated, what is genuinely new from each source, and what is only present jointly.

Redundancy and synergy are not uniquely determined by Shannon-theoretic quantities alone. Different PID constructions adopt different axioms and operational motivations, and these choices can lead to different numerical decompositions even when they agree qualitatively [40, 43, 44, 46]. This thesis does not aim to determine the best candidate or improve upon these constructions, but treats the choice among them as part of the challenge of making synergy usable in real data. Any pipeline built on them should be agnostic to the particular measure of synergy, and the definitions on which it relies must be stated explicitly. Estimation must be robust under the sample-size and sparsity constraints of cohort data, and interpretations must remain stable under reasonable sensitivity analyses. Sensitivity to the choice of decomposition is therefore treated as an empirical question in the relevant chapters rather than settled in advance. Chapter 5 repeats its decompositions under three redundancy definitions, with near-identical pair rankings ($\rho \geq 0.986$), and Chapter 6 runs its own robustness checks under four alternative

measures. Agreement in these settings is not a guarantee of agreement elsewhere.

Finally, when moving beyond pairs to larger sets, fully enumerating and decomposing all combinations becomes infeasible, and different tools are required for screening and system-level summarisation. Multivariate extensions of mutual information can serve as scalable indicators of whether a system is predominantly redundancy-dominated or synergy-dominated, without committing to a full decomposition for every subset [41, 42]. Recent developments in information decomposition frameworks also emphasise that multi-way interactions are not a single monolithic phenomenon but a family of informational modes, reinforcing the need for careful operational definitions when claims are made about "higher-order" effects in complex systems [47].

These concepts — synergy, redundancy, and their operationalisation through PID — form the interpretive backbone of the thesis. The remainder of the thesis treats them as computational objects: defining them clearly, making them discoverable under combinatorial constraints, and translating them into interpretations that remain meaningful for multimorbidity and symptom systems.

1.3. Research challenges and questions

The overall aim of this thesis is to operationalise information-theoretic synergy in health data by developing computational methods that identify, interpret, and scale higher-order dependence in multimorbidity and symptom systems. The thesis-level claim is that clinically relevant structures in these systems are often context-dependent in a way that dyadic summaries cannot faithfully represent; that synergy and redundancy provide a principled vocabulary for describing this regime; and that, with appropriate computational and robustness scaffolding, these quantities can be analysed at the cohort and symptom-network scale in a way that improves inference beyond dyadic representations.

Making that claim scientifically useful requires addressing several hurdles. The first is computational: higher-order search grows combinatorially with dimension, so exhaustive enumeration becomes infeasible in modern cohort and biobank settings. The second is interpretive: even when higher-order sets can be computed, raw triplet-level output does not naturally compress into a picture that can be compared across outcomes, domains, or clinically meaningful strata. The third is statistical: multivariate information-theoretic estimates can be sensitive to discretisation choices, finite-sample bias, and multiple testing, so an interaction-aware analysis must include explicit stability and sensitivity logic if it is to avoid overstating higher-order effects. Across chapters, the thesis therefore treats synergy not only as a definition, but as a computational object with constraints, failure modes, and validation requirements.

Central research question: *How can information-theoretic notions of higher-order dependence (especially synergy) be operationalised in health data to produce scalable, interpretable, and clinically relevant insights for multimorbidity and symptom networks?*

Sub-question 1: Representation and motivation. *Do multimorbidity and symptom burden require models beyond disease-only and beyond pairwise associations? What goes wrong when higher-order interactions are ignored and how can this be shown?* This question establishes the representational problem: the object of interest is an interacting clinical system, and there are regimes in which dyadic summaries are structurally unable to express the dependence that matters for outcomes and intervention reasoning.

Sub-question 2: Efficient discovery. *Given the combinatorial explosion, how can we find the most informative synergistic sets efficiently and robustly in high-dimensional biomedical data, and which discovery artefacts must be controlled?* This question focuses on feasibility: if synergy is to be more than a conceptual language, there must be scalable strategies for discovering candidate higher-order structures while managing redundancy-driven artefacts and screening effects that would otherwise inflate apparent interactions.

Sub-question 3: Clinical relevance. *Do symptoms provide predictive information about daily functioning beyond signs and diseases, and does this differ in people with multimorbidity?* This question ties our computational goals to clinical substance by asking whether synergy-aware analysis changes what is learnt about an outcome that matters to patients, and whether multimorbidity marks a regime in which cross-domain combinations become especially informative.

Sub-question 4: Interpretation at scale. *Once synergy is detectable, how can PID-based higher-order dependence be summarised and interpreted across many targets and subgroups in a way that remains clinically legible rather than producing a catalogue of triplets?* This question addresses the interpretive bottleneck: the goal is not merely to detect higher-order motifs, but to build an interpretation-first workflow that supports comparison, stratification, and reuse.

Sub-question 5: Symptom-network scalability. *How can we approximate network-context PID in symptom systems where full multivariate decompositions are infeasible, while preserving the meaning of redundancy versus synergy, and does the resulting decomposition replicate across cohorts and discriminate between psychometric instruments?* This question focuses on generality and scale, treating the remainder of a symptom system as "context" and asking how that context can be incorporated without losing interpretability.

Several important boundaries delimit the scope of this work. The thesis does not make causal claims from observational data; it characterises the dependence structure but does not identify causal mechanisms. It does not develop new PID axiomatics or propose novel decomposition measures; rather, it builds computational and interpretive infrastructure around existing constructions. Finally, it does not deliver clinical decision tools; the outputs are scientific analyses that inform how multimorbidity and symptom systems can be represented, not prescriptive instruments for bedside use.

1.4. Thesis outline

Each of Chapters 2–6 addresses one of the sub-questions above, moving from motivation and method to application and generalisation. Chapters 2–3 establish higher-order structure and cross-domain integration as necessary objects in multimorbidity research and make them computationally reachable. Chapters 4–6 then show what a synergy-aware analysis reveals in cohort data, how its output can be kept interpretable across targets and strata, and how the same logic scales to symptom networks. Chapter 7, the General Discussion, answers the five sub-questions in turn and sets out the limitations and future directions of the programme.

Chapter 2 establishes the representational motivation using a controlled synthetic setting. It asks when pairwise association networks are structurally insufficient for reasoning about multimorbidity systems and intervention consequences and uses this setting to clarify why higher-order structure is not merely a modelling choice but, in some regimes, a necessary descriptive object.

Chapter 3 addresses the computational bottleneck that would otherwise prevent higher-order analyses from being used routinely. It develops stochastic search heuristics (including simulated annealing and particle-swarm optimisation) to discover high-synergy candidate sets without exhaustive enumeration and formalises practical failure modes that arise when redundancy and overlap generate superficially large numbers of higher-order candidates.

Chapter 4 demonstrates the clinical relevance of synergy-aware analysis, focusing on daily functioning and the multimorbidity subgroup. It shows that symptoms contribute predictive information beyond signs and diseases, framing the result in terms of informational roles—what is duplicated, what is genuinely added, and what is only available through combinations—to connect the computational objects of the thesis to a clinically meaningful endpoint. This analysis is constrained to a system of four variables.

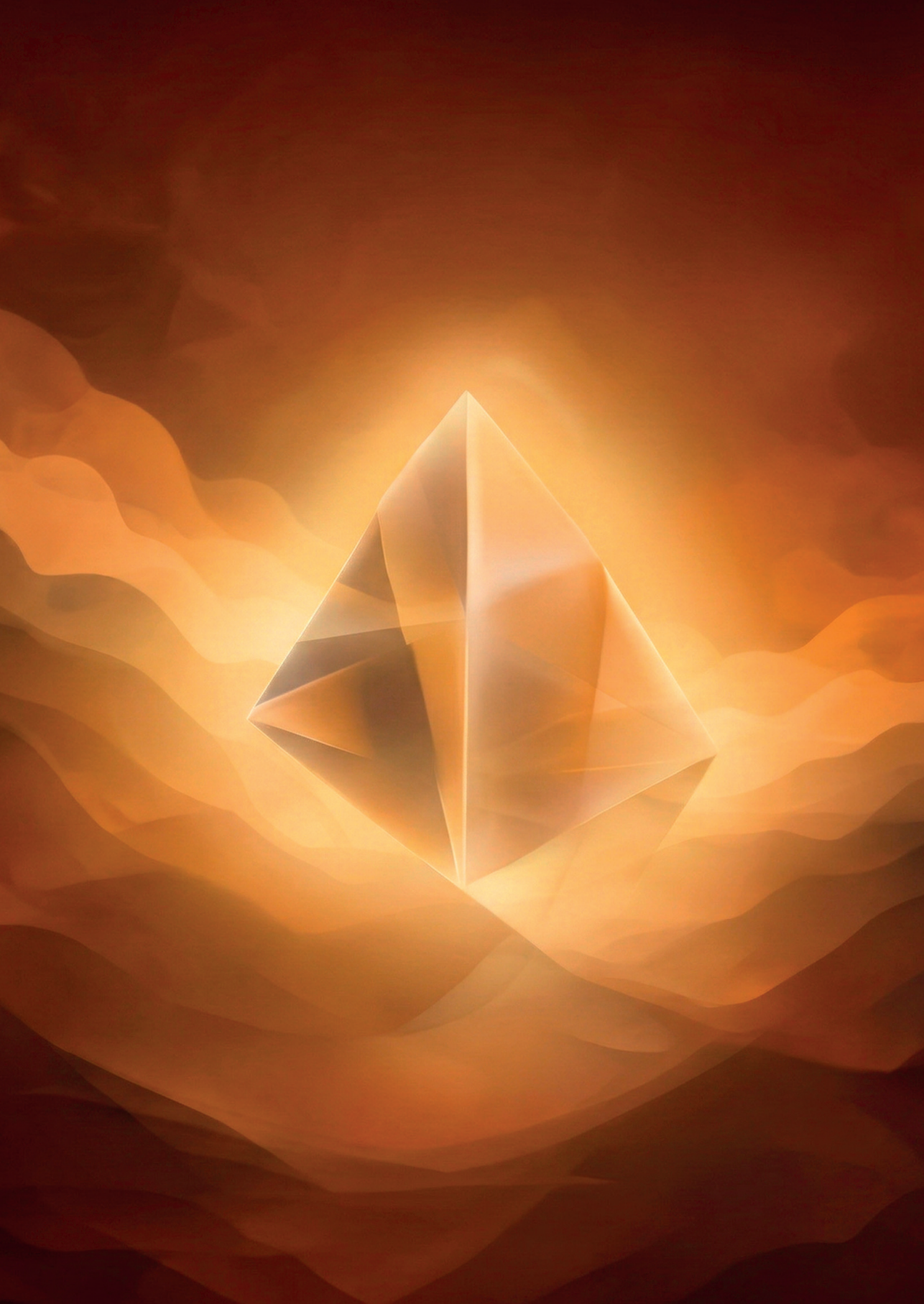
Chapter 5 confronts the interpretation bottleneck that follows from higher-order discovery. It operationalises interpretability through a structured compression of PID outputs (BUST: Breadth–Uniformity–Synergy–Total), allowing higher-order findings to be compared across targets and strata without reducing interpretation to a ranked list of motifs.

Chapter 6 extends the thesis logic to symptom networks, where the relevant "context" for a focal relation is the remainder of the system. It introduces a scalable approximation that retains a redundancy–synergy separation while incorporating multivariate context through tractable compression, enabling network-scale PID-style interpretation and supporting comparison across cohorts and instruments.

1.5. Broader impact of this thesis

The broader relevance of this work lies in making context visible and computable in settings where measurements span multiple clinical domains and decisions must be made for individual patients. In these settings, the scientific and practical question

is often not whether a variable is associated with an outcome, but whether additional measurements provide genuinely new information for a given person or subgroup, or whether they largely repeat existing signal. By providing a principled way to separate redundancy from synergy, and by developing computational scaffolding that makes these quantities discoverable and interpretable at scale, the thesis contributes a step toward a more operational notion of personalised medicine: going beyond predicting outcomes on a group level to understanding which combinations of measurements are informative, for whom, and under what conditions.



Chapter 2

A CASE FOR HIGHER-ORDER INTERACTIONS IN MULTIMORBIDITY NETWORKS

Based on:
Hourican, C., Peeters, G., Melis, R.J.F., Wezeman, S.L., Gill, T.M., Olde Rikkert, M.G.M., & Quax, R. (2023). Understanding multimorbidity requires sign-disease networks and higher-order interactions, a perspective. Frontiers in Systems Biology, 3, 1155599. *https://doi.org/10.3389/fsysb.2023.1155599*

In Chapter 1, we argued that multimorbidity demands reasoning about interactions, not independent conditions, and that dyadic summaries are structurally incomplete for key dependencies in this regime. Specifically, sub-question 1 asked what goes wrong when higher-order interactions are ignored. This chapter addresses that question with a synthetic Bayesian network in which both pairwise and synergistic associations are known by design. It shows that interventions based solely on pairwise information can miss optimal targets or produce unexpected consequences at both population and individual levels. Cross-sectional data from the EASYcare cohort then confirm these concerns empirically. Synergistic associations are not merely a theoretical possibility but a prevalent feature of sign-disease data. The chapter was published as a perspective article. Its framing of synergy is accordingly more informal than Chapter 1's information-theoretic definition. The core argument, that pairwise representations are insufficient and higher-order structure is needed, motivates the programme developed in subsequent chapters.

2.1. Introduction

Multimorbidity—the co-occurrence of two or more chronic conditions—affects over 65% of adults aged 65 and older [16, 49]. With ageing populations, its prevalence is set to rise [8, 50, 51], bringing worse health outcomes and reduced quality of life [34].

As argued in Chapter 1, the single-disease paradigm treats each condition independently, linking symptoms and signs to a single diagnosis and managing diseases in parallel [52]. For patients with multimorbidity this approach is not merely inefficient but potentially harmful, since parallel treatment leads to high treatment burden, over-treatment, and unintended treatment interactions [20].

Beyond count scores and clustering—which summarise burden or co-occurrence without describing how conditions interact—the field of 'network medicine' has brought disease networks to the study of multimorbidity [53, 54]. These networks capture pairwise associations between diseases and have been used to investigate structure, dynamics, and disease trajectories. The limitation, as Chapter 1 argued, is not that pairwise representations are wrong but that they are structurally incomplete: they miss possible higher-order (synergistic) associations. Existing works that employ hypergraphs do not use them to represent non-linear multivariate associations but rather to count co-occurrences of diseases [55, 56].

The first argument of this chapter is that symptoms, signs, and disease states should be part of one and the same network. In practice, however, the single-disease paradigm has dominated, so symptoms and signs have not been the primary focus. Tripp-Reimer et al. [57] note that symptoms are not widely used, given their subjectivity and the difficulty of extracting them from health records. Combining them in one network nonetheless matters. Symptoms and signs form plausible pathways between diseases. How symptoms, signs, and diseases should be defined remains debated, with no consensus to date [15]. Throughout this chapter we use the term *sign* to refer to an objective indication of some medical fact, *symptom* to a subjective experience of a patient, and *disease* to the diagnosis of disease states.

The second argument is that we must go beyond pairwise interactions by also including so-called 'synergistic associations' in the network models. Synergistic interactions are often referred to as effect modification or mediation, and traditionally captured by including interaction terms in regression models (more specifically: polynomial series). There are two shortcomings with this approach. The first is that it is parametric: a functional form must be specified in advance, usually a multiplication of variables. Multiplication captures certain non-linear effects, such as AND-gates, however it may miss other non-linear effects such as XOR-gates. No sequence of AND-gates can represent an XOR-gate [58]. The second shortcoming is that, depending on the choice of functional form, many terms may be needed to represent a given non-linear association. Making matters worse, for a single choice of functional form, different loadings of the interaction terms may lead to the same prediction power. This is referred to as the multiplicity problem [59, 60]. This makes the interpretation of the interaction terms, representing the synergistic interaction, difficult and ambiguous. The analysis therefore uses a non-

This chapter is based on C. Hourican, G. Peeters, R. J. Melis, S. L. Wezeman, T. M. Gill, M. G. Olde Rikkert, and R. Quax. Understanding multimorbidity requires sign-disease networks and higher-order interactions, a perspective. *Frontiers in Systems Biology*, 3:1155599, 6 2023. ISSN 2674-0702. doi: 10.3389/fsysb.2023.1155599

parametric approach, based on Shannon's information theory. More precisely, it employs the O-information metric because of its computational efficiency [42].

We first demonstrate our two arguments in a synthetic Bayesian network (BN) model constructed using both pairwise and synergistic associations, which we use to generate synthetic datasets with 100,000 realisations. In healthcare, BNs are used in aiding decision-making processes for diagnosis, prognosis, and treatment selection [61]. Binary nodes of diseases, disorders and risk factors, along with background information, have been used in BNs to study the multimorbidity rate, and co-occurrences of multiple diseases [62–64]. However, to our knowledge, no BNs appear to study multimorbidity using symptoms, signs and diseases in one network, even though several works mention the need to combine these elements [65–67]. Then we use *EASYcare Twostep Older persons Screening data* [68] to identify synergistic associations and highlight their prevalence. We show that interventions guided solely by pairwise associations can miss the most efficient target or produce unexpected downstream effects, and we identify 184 statistically significant synergistic triplets in the EASYcare data—evidence that the higher-order structure uncovered in the synthetic setting is not merely a theoretical curiosity but a pervasive feature of empirical sign-disease systems. The code and generated data are provided at `https://github.com/CillianHourican/Synergistic-Networks`.

2.2. Methods

2.2.1. A primer on synergistic interactions

Synergistic effects are common in complex systems and can take many forms. The clearest evidence of a synergistic effect is when independent random variables individually cannot predict a target variable (zero mutual information) but when considered together they can (non-zero mutual information). This is not the only situation where synergistic effects take place: synergy is also present when the sum of the pairwise dependencies between the independent variables and the target variable is less than the multivariate dependence of the independent variables on the target variable. Throughout this chapter, 'dependence' is measured using Shannon's mutual information (introduced in Chapter 1) [69], which is non-parametric and captures both linear and non-linear associations [70].

A classic example of a purely synergistic interaction is when predicting the output Y of an XOR-gate with two binary inputs S_1, S_2. An XOR gate effectively indicates whether the inputs are equal (output 0) or unequal (output 1). Intuitively, this is a type of relationship where the output cannot be predicted from one input alone. To see this, let an XOR gate output $Y = 1$ whenever exactly one input equals 1 and the other 0, and $Y = 0$ otherwise. Furthermore, let each input be 50/50 distributed. Without observing any input values, the target Y has a marginal probability distribution which is maximally uncertain (also 50/50 distributed):

$$P(y = 1) = P(y = 0) = 0.5.$$

This distribution does not change if we condition on either one of the input variables:

$$P\left(y=0\middle|S_1\right)=0.5=P\left(y=1\middle|S_1\right),$$

$$P\left(y=0\middle|S_2\right)=0.5=P\left(y=1\middle|S_2\right).$$

In words, observing either input alone does not improve our information about (ability to predict) the output. Indeed, using mutual information we see that separately each input tells us zero information about the response:
$MI(S_1;Y)=0$, $MI(S_2;Y)=0$.

$$MI(S_1,S_2;Y)=1.$$

This is because observing both inputs fully determines the output Y (100/0 or 0/100 distributed). Overall, the information about the target in an XOR-gate is not stored in either single input (pairwise dependence) but is stored completely synergistically in the combination of both inputs.

This is important because many networks are constructed using pairwise mutual information. In our example, though, the pairs (S_1,Y), (S_2,Y) have zero mutual information. No algorithm that depends on pairwise associations can find the synergistic relationship. This is the essence of our argument against focusing purely on pairwise association networks and for explicitly including synergistic interactions. This argument remains true when building association networks based on conditional dependencies or conditional independences [71].

Quantifying the exact amount of synergy between variables remains an open problem, with frameworks such as partial information decomposition aiming to decompose information between variables into unique, redundant, and synergistic information [40, 41, 72, 73]. In our current study, we use the *O-information* heuristic when analysing data to assess if synergy is indeed present [42]. This heuristic is easy to compute but is conservative, i.e., some interactions may be at least partly synergistic while the O-information measure fails to detect them [42]. The converse is not possible: if O-information indicates net synergy, then the system's synergistic dependencies exceed its redundant ones, though the absolute magnitude may still be small. This metric has been utilised in neuroscience, in fMRI signals to characterise higher-order communication between different regions of the brain, and to capture neural spiking dynamics [74, 75]. This information-theoretic approach captures synergy by computing (multivariate) mutual information and is model free.

Beyond multimorbidity, capturing synergistic interactions is relevant in many domains, such as synergistic drug-drug combinations, increased co-occurrences of diseases, and functional activity in the brain [41, 76–78]. Prediction models are becoming increasingly complex. Feature attribution methods have therefore become essential tools for explainability and interpretability. This usually involves a sensitivity analysis or breaking the system into components and analysing how these components interact to produce its

overall behaviour. Approaches such as Integrated Gradients, Layer-wise Relevance Propagation, and Gradient-weighted Class Activation Mapping have been developed for neural networks, while Sobol' Indices and Shapley values are used for a much broader range of models [79–83]. These approaches require two choices to be made: what model should be fit to the data, and what attribution technique should be utilised. This contrasts with the information-theoretic approach, which is applied directly to data. Further discussion on different approaches to capture synergy, O-information, and mutual information variants is provided in Appendix A.2.

2.2.2. Creating a synthetic network model with synergistic interactions

Our synthetic model (Figure 2.1(A)) is a BN model of sign and disease variables. In this example model we do not differentiate between symptoms and signs and refer to them simply as signs. BNs are causal graphs which are directed and acyclic, and whose edges are defined by conditional distributions among stochastic variables, which are shown as nodes [84].

Our running example follows Lucas et al. [61] and Borsboom [85], where signs can cause (the diagnosis of) diseases, but diseases do not directly cause sign activation. In other words, while signs are causally linked and can have both incoming and outgoing connections, diseases are inferences derived from (a combination of) signs. Notwithstanding this, others do consider causal relations from diseases to signs [86]. We expect our main conclusions to hold if such relations were added, since the synergistic triplets are constructed among signs alone.

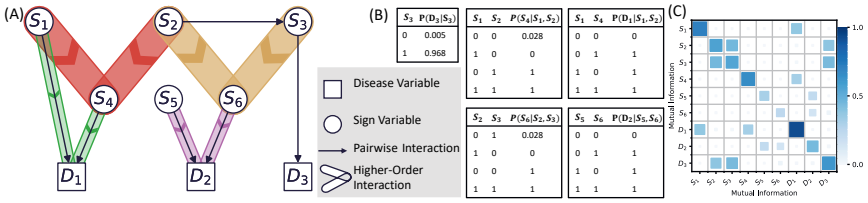


Figure 2.1: **(A)** The synthetic network model with pairwise and synergistic interactions. The narrower hyperedges indicate interactions less synergistic than the wider hyperedges. **(B)** The conditional probability tables that define the dependencies in the model. The table for $P(S_4 \mid S_1, S_2)$ reads as follows: The first row shows a 2.8% chance of S_4 being on when S_1 and S_2 are off. If S_1 is on and S_2 is off, S_4 is always off. If S_1 is off and S_2 is on, S_4 is always on. And if both S_1 and S_2 are on, S_4 is always on. **(C)** Mutual information scores for each pair of variables from a synthetic dataset, with 100,000 realisations, generated by the model. Larger and darker squares show stronger pairwise associations.

The model is built in three steps: node types and priors, pairwise edges, and the synergistic triplets that feed the disease nodes. Our synthetic network has three independent nodes (S_1, S_2, S_5), without incoming links, each with a Dirichlet prior distribution [87]. For further simplicity, our model has only binary nodes ('on/off'). This is sufficient for our demonstration, but the approach described is valid for both discrete and continuous

data.

Node S_3 has a large mutual information with its parent node S_2, as represented by the direct link from S_2 to S_3. Similarly for the edge $S_3 \rightarrow D_3$. Arrows on each edge indicate the causal direction in the network.

Highly synergistic random variables from Quax et al. [88] were then appended to the system, as shown by the wide edges for $(S_1, S_2) \rightarrow S_4$ and $(S_2, S_3) \rightarrow S_6$, which we refer to as triplets. We constructed them by jointly optimising the conditional probabilities to maximise the so-called Whole-Minus-Sum objective function. That objective minimises the pairwise mutual information and maximises the combined mutual information of both inputs about the target. For the edge $(S_1, S_2) \rightarrow S_4$, this synergistic association ensures that the pairwise mutual information terms $MI(S_1, S_4)$ and $MI(S_2, S_4)$ are approximately zero. Considered jointly, the two nodes nonetheless carry a large mutual information about the target. A similar setup is seen with the triplet $(S_2, S_3) \rightarrow S_6$. However, in this case the mutual information $MI(S_2, S_3)$ is significant, as shown by the directed edge, thus making the synergistic effect weaker. Finally, OR-gates were used for the dynamics of diseases D_1 and D_2, where D_1 activates if either S_1 or S_4 activates, while D_2 activates if S_5 or S_6 activates. This combination of pairwise and synergistic edges results in a hypergraph structure.

Pairwise associations alone cannot reconstruct this network structure. We assume the ground truth is known and focus on how synergistic interactions influence model interventions. We do not address how such a BN might be reconstructed from data, or the presence of latent variables in this synthetic model. The only latent variables present are noise-generating functions. Figure 2.1(C) shows a matrix of mutual information values for each pair of variables, computed from a synthetic dataset generated by the model. Each shaded square represents an edge in the corresponding pairwise network. Some associations are missing, which incorrectly results in three disconnected pairwise networks (see Appendix Section A.7). Including synergistic associations is needed to recover the complete structure.

2.2.3. Model Interventions

Next we performed two types of interventions to gain insight into this complex system. The first is a mass influence on the whole population. The second acts on a single patient.

We simulated the population-level intervention by picking a sign variable and nudging its marginal distribution slightly. The marginal probability distribution reflects the prevalence of the sign in the population. By nudging we mean moving probability mass from one state to the other in the probability mass function (pmf) for a variable, slightly changing its marginal probabilities while keeping all conditional probabilities unchanged. In causal inference this is known as a soft intervention [89]. In our model, conditional probabilities reflect causal mechanisms (e.g., smoking leads to increased risk of lung cancer), so we only change marginal probabilities. This type of intervention is akin to a shift in society's normal behaviour, or to an attempt to manipulate population-level causes of the incidence of sickness through public health measures rather than to protect

high-risk individuals within a population [90]. For example, a nudge of 0.1 on node S may correspond to increasing $P(S=1)=0.3$ to $\widehat{P}(S=1)=0.4$, which changes the probability by $\left|P-\widehat{P}\right|=0.1$. The effect of the intervention is then computed using the change in log-odds ratio for each node in the network. In the binary setting, a nudge is not probabilistic and so is fully reproducible. However, when using multivariate or continuous variables there are many ways to shift the pmf, and some sampling techniques would be needed to analyse the effects of nudge interventions, as the results would vary each time.

Our second simulated intervention, on the other hand, pertains to clinical practice where we are concerned with a single patient rather than population-level scores. Here we may ask: "Given that this patient has this sign, what other sign is she likely to acquire?". Or, "Can the onset of a particular sign be prevented by a strategic intervention, given the other signs of a patient?". In this case it no longer makes sense to speak of a pmf of a sign: a given patient has a well-defined value for a sign, which in our binary formalism means a patient either has or does not have a particular sign. To this end, individual-level interventions were performed by generating one realisation (value for each sign and disease in the network), changing the state of one or more nodes, and recording the downstream changes. This is akin to hard interventions in do-calculus, as it forces a node to take a certain value [89].

2.2.4. Uncovering synergistic associations in medical data

We then searched for synergistic associations in empirical data using the *O-information* heuristic. The data set is the *EASYcare Twostep Older persons Screening data* [68], covering 587 older persons aged 60+ from GP practices in the Netherlands. It holds cross-sectional data on 37 signs and on the absence or presence of diseases, taken from a complete medical examination. Disease presence is binarised from the *Cumulative Illness Rating Scale-Geriatric* (CIRS-G): each of the 14 CIRS-G subscales scoring 2 or higher counts as a disease. Screening proceeds in two steps. General practitioners first flag older patients they judge to be potentially frail, and only those patients receive the full second-step assessment. The sample is therefore enriched for frailty and multimorbidity rather than representative of older adults in general. The counts reported below are frequencies within this screened group, not population prevalences. A permutation test was used to determine the significance of each synergistic triplet, with the Benjamini-Hochberg procedure controlling the false discovery rate at 0.15 [91].

2.3. Results

2.3.1. Interventions at the population level

First, we nudged the pmf of S_5 down by 0.2 (Figure 2.2(B)), changing its marginal probabilities while keeping all conditional probabilities unchanged. The prevalence of D_2 fell slightly, while all other variables were unaffected. The pairwise network anticipates this direct causal effect, since it infers a connection from S_5 to D_2. No other node is a descendant of S_5.

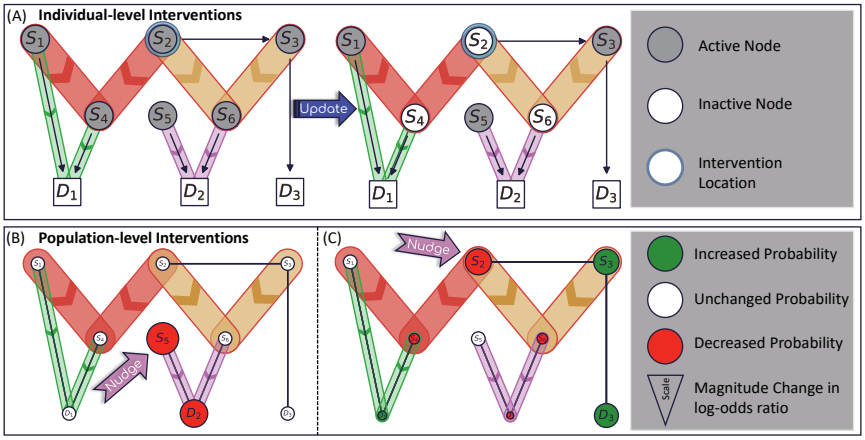


Figure 2.2: **(A)** An example realisation of the network. Coloured nodes are active ('on') in the 'patient', while white nodes are inactive. An individual-level intervention turns off one node, updating downstream nodes as they also become inactive. **(B)** The effect of a (population-level) nudge intervention on node S_5. This intervention updates downstream marginal probabilities from where the nudge intervention occurred. Larger nodes indicate a larger change in the log-odds ratio compared to the system state before interventions. Green nodes show increased prevalence (marginal probabilities), red nodes show decreases, and white nodes show no change. **(C)** The corresponding effect of nudging node S_2.

In a second intervention, we nudged the prevalence of S_2 by the same amount (Figure 2.2(C)). This time the effect spread. The pairwise links $(S_2 \rightarrow S_3)$ and $(S_3 \rightarrow D_3)$ accounted for the changes a pairwise analysis would predict. By construction, the synergistic associations also shifted the distributions of S_4 and S_6, which in turn changed D_1 and D_2. The pairwise network missed these changes to S_4, S_6, D_1, D_2.

2.3.2. Interventions at the individual level

Suppose a patient presents with signs $S_1, S_2, S_3, S_4, S_5, S_6$ (Figure 2.2(A)), where we have no reason to prioritise treating one sign over another. When considering only the direct pairwise associations, we incorrectly conclude that the active signs are not connected to each other. This implies that each sign must be treated individually, meaning four separate treatments for the patient. However, including the synergistic associations changes this. In the synthetic model, sign S_2 has a synergistic causal effect on signs (S_4, S_6) by construction, so this intervention will impact both signs, as making S_2 inactive also makes both (S_4, S_6) inactive. Ignoring synergistic interactions leads to overtreatment of the patient, as this involves intervening on every active sign node.

2.3.3. Synergistic associations in medical data

For computational efficiency, we first found triangles of low dyadic mutual information ($MI < 0.05$), as these are likely to be missed when performing a pairwise analysis. Of

these, 184 triplets were significantly synergistic by permutation test, with the Benjamini–Hochberg procedure controlling the false discovery rate at 0.15. At that threshold roughly 28 of the 184 are expected to be false discoveries, so individual triplets are best read as candidates rather than as established associations. The pre-filter also biases the search toward purely synergistic triplets and may exclude triplets with mixed redundant-synergistic structure. Chapter 3 relaxes this constraint.

One such example is weight loss (w), dyspnoea (d), and locomotor pain (pl), which all have weak dyadic associations but significant synergistic associations. A patient with only locomotor (joint) pain may be advised to lose weight to reduce the strain on their joints [92]. Similarly, a patient with dyspnoea may be advised to lose weight to alleviate its effects [93, 94]. However, a patient with both signs is further limited in the intensity and duration of physical activity. This combination adds to the patient's burden, making weight loss to reduce joint pain less likely to succeed [95]. The probability of weight loss (0.087) is almost unchanged given the presence or absence of a single sign:

$$P\left(w \middle| pl = 1\right) = 0.094,$$

$$P\left(w \middle| d = 1\right) = 0.084,$$

$$P\left(w \middle| pl = 0\right) = 0.077,$$

$$P\left(w \middle| d = 0\right) = 0.088.$$

However, conditioning on both signs changes the probabilities substantially:

$$P\left(w \middle| pl = 0, d = 0\right) = 0.048,$$

$$P\left(w \middle| pl = 0, d = 1\right) = 0.169,$$

$$P\left(w \middle| pl = 1, d = 0\right) = 0.126,$$

$$P\left(w \middle| pl = 1, d = 1\right) = 0.045.$$

These four values do not follow from the single-sign values above. Dyspnoea without locomotor pain carries roughly double the marginal probability (0.169), while the two signs together give the lowest value of the four (0.045). The synergy claim rests on this non-additivity across the four cells rather than on any single one of them. The four probabilities come from 587 people, roughly 50 of whom record weight loss, so the third decimal place should not be read as precise. Significance rests on the permutation test, not on these numbers.

Other synergistic triplets include *('Stiff restricted', urinary tract infection, CIRS-G Lower GI)* and *(Movement limitation, Hearing problem, CIRS-G Heart)*. A collection of synergistic triplets is given in Appendix A.9.

2.4. Discussion

Multimorbidity research is moving from individual treatment of diseases towards a complex-systems view, but it remains unclear how exactly this is to be achieved. Network models of diseases and disease clusters, which are based on dyadic associations, are becoming increasingly popular to study multimorbidity [53, 54]. However, such networks may miss important synergistic relationships in data, and higher-order associations should therefore be included. In a medical dataset, 184 triplets were significantly synergistic. These 184 are drawn from the roughly 20,000 triplets that 37 signs and 14 disease subscales can form, and from a search restricted in advance to triangles of low dyadic mutual information. The count therefore establishes that synergistic associations are present and recoverable rather than that they are common. Establishing their true prevalence requires more datasets, a direction pursued in Chapters 3–6 using independent cohorts. The synthetic demonstration is necessarily constrained by its topology: a single nine-node network cannot establish that pairwise analyses will miss synergistic structure in every possible configuration. Its value is as a proof of concept—showing that such failures *can* arise by construction—complemented by the empirical evidence from EASYcare and the independent cohorts analysed in later chapters.

Melis et al. [13] argue that methods modelling multimorbidity at the disease level may have limited value, as they do not accurately inform clinical decision making. With increasing disease burden, the presenting signs become less specific, and it is more difficult to identify a clear set of underlying diseases. Therefore, network models must uncover relationships among signs and diseases. Constructing such networks with synergistic associations uncovers interdependencies among signs and diseases, including those missed when only pairwise associations are considered.

There are two key points to take from the analysis of the synthetic model. Firstly, synergistic associations may cater for changes in the system that pairwise associations cannot. Statistical analyses should therefore go beyond pairwise associations and use methods able to recover synergistic structure in empirical data. Intervention effects cannot otherwise be fully determined.

Secondly, the synthetic model demonstrated that local interventions propagate through synergistic pathways in ways that pairwise representations cannot predict. Networks built from pairwise associations alone may miss carry-through effects mediated by synergistic interactions, leading to inaccurate expectations about global consequences and potentially targeting more or fewer nodes than needed. The implications of this structural principle for intervention planning—including the contrast between single-target and multi-target strategies—are developed in the General Discussion.

These approaches are not without limitations. Finding synergistic associations brings additional computational costs, especially when including higher-order groups, such

as quadruplets and quintuplets—a scalability bottleneck that Chapter 3 addresses with stochastic search algorithms. Capturing higher-order associations is constrained by the availability of data, which limits the complexity of the resulting hypergraph. Quantifying the amount of synergy also remains a theoretical problem [88, 96, 97]. Spurious findings should also be considered when constructing network models [36, 98], a concern that Chapter 3 formalises as the echo problem and mitigates through filtering strategies. Despite these limitations, it is important to be aware of such synergistic associations, especially when they uncover relations missed in the pairwise-based network.

To conclude, patients with multimorbidity have complex patterns of signs and diseases, each of which should simultaneously be included in networks constructed to study multimorbidity. Studying pairwise associations alone is not sufficient in such a complex setting. Synergistic associations between signs occur and must be included in future analyses. These arguments were supported by both results from the simulations in the synthetic model and analyses of real-world data.

Having established that higher-order associations both exist and matter, the remaining question is how to find them efficiently at scale.



Chapter 3

EFFICIENT SEARCH ALGORITHMS FOR IDENTIFYING SYNERGISTIC TRIPLETS IN HEALTH DATA

Based on:

Hourican, C., Li, J., Mishra, P.P., Lehtimäki, T., Mishra, B.H., Kähönen, M., Raitakari, O.T., Laaksonen, R., Keltikangas-Järvinen, L., Juonala, M., & Quax, R. (2024). Efficient Search Algorithms for Identifying Synergistic Associations in High-Dimensional Datasets. Entropy, 26(11), 968. *https://doi.org/10.3390/e26110968*

This chapter develops stochastic search algorithms that identify high-synergy variable sets without exhaustive computation, and formalises a systematic bias—the echo problem—in which strong pairwise associations inflate the apparent number of synergistic triplets. In Chapter 2, we demonstrated that synergistic associations are a prevalent feature of empirical sign-disease data, but that analysis was feasible only because the dataset contains a manageable number of variables. In metabolomic panels, biobank-scale surveys, and dense symptom inventories, however, the number of candidate triplets grows combinatorially, and exhaustive enumeration becomes impractical. A dataset of 400 variables, for instance, generates over ten million possible triplets. At the scale of modern biobanks, even triplet-level screening is computationally prohibitive. Left unaddressed, the echo problem can distort results when synergy is quantified in large datasets. We propose filtering strategies to control it. The search algorithms and echo-filtering heuristic are evaluated on two epidemiological datasets, the Young Finns Study and UK Biobank NMR metabolomics data, to demonstrate that the approach generalises beyond the clinical multimorbidity setting of the preceding chapter.

3.1. Introduction

Synergistic associations are information carried jointly by a group of variables but absent from any subset, as defined in Chapter 1 and demonstrated empirically in Chapter 2. They are complementary to the pairwise associations that dominate current analytical practice. The analysis of pairwise associations is commonplace, but synergistic associations have recently received growing attention and have been found in several different domains, such as neuroscience [74, 75, 100], cardiovascular disease [101], multimorbidity [48], ageing [102], and psychometrics [103], among others.

As noted in the preamble, evaluating all synergistic associations requires enumerating all variable subsets—a combinatorial barrier that Chapter 1 formalised and that limits applicability on large-scale datasets. Many related works confine their search to a smaller set of variables to remain feasible, but this risks leaving synergistic associations undiscovered.

This chapter proposes to identify the strongest synergistic associations by making use of global optimisation algorithms. We employed two stochastic search strategies as well as a direct network-based approach. The main idea is that the brute-force method to identify the strongest synergistic associations of order n (i.e., among n variables) would require a combinatorial number of calculations. Instead, the search algorithms were guided by lower-order features (i.e., statistical quantities computed from sets of variables of size less than n), an approach that is much less time-consuming. The hypothesis is that the search algorithms still reach the extreme synergistic tail of the brute-force distribution, that is, triplets ranked among the top associations identified by exhaustive search. What we report is how often runs reach that tail, not what fraction of the ranked top sets they recover. Focusing on only the top synergistic associations simplifies the search task compared to brute force. A total ordering is not sought, only a partial ordering over a small number of associations.

There is no universally accepted metric for computing synergistic associations. Several algorithms estimate (the lower bound of) the amount of synergistic information in a given set, while also considering discretisation, bias, and sample-size effects. Some are based on partial information decomposition (PID) [22, 40]. Others are computationally expensive, requiring various optimisations for computing synergistic random variables [88] or finding geodesics between sub-manifolds [104]. Even for more lightweight approaches, such as the O-information introduced in Chapters 1 and 2 [42], the time complexity for computing sets of size p on a dataset with k variables is $O(k^p)$; computing all possible sets has complexity of $O(2^k)$. This limits the applicability of these approaches on large-scale datasets. The approaches presented here are evaluated on two synergy measures—a bias-corrected O-information and a PID-based approach—but the search framework is general and not limited to any particular synergy measure.

3.1.1. Search algorithms in outline

A wide range of global optimisation algorithms exist for minimising an objective function over a search space. In this chapter, we explored three search strategies for finding syner-

This chapter is based on C. Hourican, J. Li, P. P. Mishra, T. Lehtimäki, B. H. Mishra, M. Kähönen, O. T. Raitakari, R. Laaksonen, L. Keltikangas-Järvinen, M. Juonala, and R. Quax. Efficient Search Algorithms for Identifying Synergistic Associations in High-Dimensional Datasets. *Entropy*, 26(11):968, 11 2024. ISSN 1099-4300. doi: 10.3390/e26110968

gistic associations. The first two are stochastic—simulated annealing (SA) and particle swarm optimisation (PSO)—and the third is a more direct network-based approach. For this, we constructed a network of order $n-1$ (e.g., pairwise) metrics (e.g., conditional entropy or mutual information) which could be used to identify order n (e.g., triplet) synergistic associations. The predictive $n-1$ feature is itself found through model selection, where a regression model is trained on a selection of $n-1$ features to predict synergistic information. We then define a clique-based heuristic algorithm that searches for triangles in this network which are likely to contain synergistic information. This approach allows us to quickly identify a set of potentially synergistic associations, although the strongest synergistic sets are not necessarily found.

3.1.2. The echo problem

We discovered a complication that arises when identifying synergistic associations in data which we refer to as the echo problem. That is, whenever a strong pairwise association is connected to a synergistic n-plet, a further n-plet will frequently be found, because the endpoints of the pairwise association are (almost) interchangeable. The implication is probabilistic rather than deterministic. Figure 3.4 gives a counterexample in which a strongly associated pair produces no echo. Since any dataset contains both pairwise and synergistic associations, this can lead to an explosion of identified n-plets, even if the underlying data generating mechanism contains only a few synergistic n-plets. We consider such n-plets, which are present only because of a connecting pairwise link, as spurious. In the literature, this appears to be a so-far underappreciated aspect of inferring higher-order interactions and information-theoretic hypergraphs [103]. Echoes are common in practice: in the UK Biobank NMR data, 62% of the top 100 synergistic triplets were rejected as echoes at a mutual information threshold of 0.6. Inflation on this scale is expected to alter the resulting network structure, changing which nodes appear central, the distribution of path lengths, and clustering. We did not quantify that distortion here. Significance testing is affected more directly, because an inflated number of sets forces a multiple-comparison correction that is often unfeasible. Therefore, we propose an additional regularisation step that is incorporated into the global optimisation.

The purpose of this chapter is two-fold: first, to efficiently identify the most synergistic associations in the data, and second, to filter out synergistic sets with overlapping associations. Along the way, we observe that pairwise feature selection can suppress genuine synergistic associations. Search performance was assessed against the brute-force results. We compared the most synergistic triplets recovered by each strategy to the distribution of all triplet scores in the data, and we compared running times and the number of calls to the synergy metric. It follows that variables should not be pre-filtered on pairwise criteria, but retained until higher-order structure has been assessed. The approaches are tested on two datasets: the Cardiovascular Risk in Young Finns Study (YFS) [28] and NMR data from the UK Biobank [27].

3.2. Background

3.2.1. Measuring synergy

Two complementary synergy metrics were selected to evaluate the search algorithms: bivariate PID and O-information. The first is the BROJA measure of Bertschinger et al. [105], estimated with the optimisation procedure of Makkeh et al. [106] in the `idtxl` package [107]. The second is the O-information, introduced in Chapter 2 [42]. The PID metric solves a convex optimisation problem to decompose the mutual information $I(X_1, X_2; Y)$ that sources X_1 and X_2 jointly carry about target Y into synergistic, redundant and unique components. The synergistic part is always non-negative. O-information, by contrast, is a direct calculation of the balance (arithmetic difference) between synergy and redundancy, so its value can be negative (synergy-dominated) or positive (redundancy-dominated). O-information has the benefit of being instantaneous to compute (no optimisation needed) but it is conservative, i.e., some interactions may be at least partly synergistic but the O-information measure fails to detect them [42] when redundant information is also present. The converse is not possible: if O-information infers a synergistic relationship then there must be significant synergy, i.e., there must be more synergy than redundancy. This metric and its variants have recently seen a surge of interest, especially in neuroscience [74, 75, 108–112]. In this chapter we assess only whether the search algorithms are a good approximation of a brute-force approach in ranking synergistic associations. We do not compare synergy values or rankings across the two metrics, since they are not expected to agree. The O-information values reported here use the bias-corrected estimator of Gehlen et al. [113], which adjusts for the finite-sample upward bias in plug-in entropy estimates. The correction does not remove the sparsity of the state space. With 336 participants in the UKB NMR data, a triplet discretised by Sturges' rule spans on the order of 10^3 joint states, well above the number of observations. These O-information values are therefore read as a ranking rather than as calibrated magnitudes, while the YFS analysis, with 1684 participants, is far less exposed. The search algorithms are judged against brute-force results computed on the same data with the same estimator, so any residual bias applies to both sides of the comparison. Both metrics are treated within a minimisation framework: for O-information, more negative values indicate stronger synergy, while for PID-based synergy we negate the score so that the same search logic applies. All update rules in the SA and PSO algorithms follow this convention.

3.2.2. Simulated Annealing and Particle Swarm Optimisation

SA and PSO are metaheuristic algorithms for global optimisation. SA begins at a high "temperature" that gradually "cools", exploring the solution space broadly at first and then narrowing onto promising regions. Early in the search it probabilistically accepts poorer solutions to escape local optima, and this acceptance probability falls as the temperature drops. PSO instead initialises a population of candidate solutions, or "particles", that move through the search space according to their own best-known positions and the swarm's best-known position [114]. Although PSO is primarily used in continuous spaces,

variants have been developed for discrete spaces and for set-based frameworks that address combinatorial optimisation problems [115–117]. Here the global minimum is the most synergistic n-plet (subset) in a dataset for a given set size n. SA can become trapped in local minima, so multiple runs are often necessary. Because its particles share information about their individual successes, PSO is parallelisable, scalable, and less prone to premature convergence in complex, multimodal landscapes. SA was applied in [118] to search for synergistic information using O-information. Our extension guides the search toward synergistic sets by incorporating pairwise metrics into the proposal of new sets at each step. To our knowledge, no such search strategy has been applied to other synergy metrics.

3.3. Methods

The methodological pipeline had two stages. First, a regression model identified which pairwise metrics best predicted triplet synergy. Second, those metrics were used to guide three search strategies: SA, PSO, and a clique-based heuristic. The same procedure extends in principle to higher orders. Only triplets are evaluated in this chapter. The objective was to assess whether combinations of pairwise metrics were predictive of the presence of synergistic triplets and, if so, to transform them into pairwise-based metrics for biasing the search direction of the global search algorithms.

3.3.1. Dataset preparation

We tested the effectiveness of the search algorithms on two biomedical datasets with different focus areas: the NMR data from the UK Biobank [27] and the Cardiovascular Risk data in the Young Finns Study (YFS) [28].

- The **UK Biobank** is a large-scale biomedical database and research resource containing in-depth genetic and health information from half a million UK participants. NMR (Nuclear Magnetic Resonance) spectroscopy data in the UK Biobank refer to metabolic profiling of blood samples. NMR spectroscopy provides detailed information on lipoproteins, metabolites, and other molecules in the blood.
- **YFS** is an epidemiological cohort study that has followed the risk factors for cardiovascular disease from childhood onwards. The study aims to understand how childhood lifestyle, biological factors, and social environment influence cardiovascular health in adulthood. It contains clinical assessments, psychological evaluations, lifestyle questionnaires, and biological samples, tracking participants from childhood to adulthood.

For computing synergy values based on O-information we used the continuous variables from the UKB NMR dataset and discretised them by quantile binning, with the number of bins set by Sturges' rule. Discretisation was done so that meaningful entropy-based features could be computed, and quantile binning was used to reduce any biases arising from the number of bins used. Sturges' rule gives ten bins for the 336 UKB NMR

participants, so a triplet's joint distribution spreads 336 observations over 10^3 = 1,000 cells. Most of those cells are empty. Entropy estimates remain heavily biased at this density, even with the bias-corrected estimator of Gehlen et al. [113]. The O-information values reported here are therefore a scoring function for ranking triplets within a dataset rather than calibrated estimates of higher-order information.

The UK Biobank NMR extract used here consists of 336 participants with detailed metabolomic data, a subsample of the half a million participants described above. Variables with more than 20% missing entries were removed, leaving 1,426 variables. Missing values in the remainder were imputed with a K-Nearest Neighbours Regressor, using the `IterativeImputer` from scikit-learn. Imputation ran with a convergence threshold of 0.01 and a maximum of 1,000 iterations, stopping as soon as the average change between successive imputations fell below the threshold. For computing synergy values based on the PID framework we used four states per variable in this dataset, because the computational cost scales with the number of states per variable. The UK Biobank arm is therefore a computational test case for the search strategies rather than a source of population-level estimates. The stability of the brute-force ranking under resampling was not assessed.

In the YFS, data from 1,684 participants were used and continuous variables were discretised in the same manner. No imputation was performed. Only participants with complete data on the variables entering a triplet contribute to its score. When searching for O-information synergy, 377 variables were used from the 2007 wave. Existing discrete variables such as gender and depression scores were also included so that the search algorithms could be assessed on data with non-uniform entropy distributions. When searching for PID-based synergy, we used 684 variables from the 2001, 2007 and 2011 waves, all of which could be discretised into exactly five states. Pooling waves means a triplet can combine measurements taken up to ten years apart. Such sets describe dependence across the life course rather than contemporaneous association.

3.3.2. Using lower-order information-theoretic metrics to predict synergistic associations

We computed several variants of pairwise metrics with aggregation functions such as minimum, maximum and average, as well as all triplet synergies (brute force) in both the UKB NMR and YFS datasets. These brute-force values provide both the regression target and the ground truth against which the search algorithms are evaluated. Their $O(k^p)$ cost is what the search algorithms avoid. Predicting a triplet's synergy from the pairwise metrics is a multi-regression problem, with predictors such as the minimum pairwise mutual information, the sum of joint entropies, and the spread of mutual information scores. We used the Light Gradient Boosting Machine (LightGBM) and performed a grid search of hyperparameters using five-fold cross-validation to minimise the mean squared error.

Recursive feature elimination with cross-validation (RFECV) was used to perform a feature selection on our gradient boosting model. This process systematically discarded

the least important features, splitting the data into five folds to iteratively train the model and evaluate its performance using the root mean squared error (RMSE). After feature selection, the `train_test_split` method from the scikit-learn library [119] was used to create a single split of 70% training data and 30% test data. Each combination of the features retained by RFECV was then evaluated on this split, with a grid search over hyperparameters to minimise RMSE. The model using the best-found hyperparameters for each combination was fitted to the training data and scored on the held-out test data. This procedure identified the most effective feature subsets for our search strategies.

3.3.3. Simulated Annealing for synergistic associations

SA is a probabilistic algorithm for approximating the global optimum of a function by gradually cooling the search process. The algorithm begins by initialising a particle at a random position $X^0 = \{x_j, x_k, \ldots, x_m\}$ within the search space of all sets of the desired cardinality. For example, when searching for synergistic triplets, each set's cardinality is three, with each x_j representing a unique variable in the dataset. This initial position is evaluated using a synergy scoring function, $\Omega(X^0)$.

In each iteration, a new candidate position $\hat{X}^t$ is proposed by modifying one or more variables in X^t. The number of variables changed is drawn from a temperature-dependent distribution, defined below. The proposed set is then evaluated. If $\Omega(\hat{X}^t) < \Omega(X^t)$, indicating a more synergistic set, $\hat{X}^t$ replaces X^t as the current set. If the candidate set is less optimal, it may still be accepted with probability $\exp\left(-\frac{\Omega(\hat{X}^t)-\Omega(X^t)}{T}\right)$, where T is the algorithm's temperature and decreases over time. This cooling schedule lowers the acceptance rate of suboptimal sets as T approaches zero, allowing broad exploration early on to avoid local minima and refinement toward an optimal solution as T cools.

Our SA algorithm extends the approach of [118] by drawing the variables in proposed sets from non-uniform distributions. Each run begins from a random set, which is scored to give an initial synergy value. Two choices remain: how many variables to change, and which ones.

At each iteration, the number of components to modify follows a geometric distribution whose parameter depends on the current temperature. The probability parameter p is defined by $\left(\frac{\text{current temperature}}{\text{initial temperature}}\right)^{0.1}$, which keeps p within $(0, 1]$. The number of variables to change is sampled from this distribution. Because p falls as the system cools, the expected number of variables swapped per step grows. The search therefore narrows through the Metropolis acceptance criterion rather than through the size of the proposed move.

Once the number of variables to swap is set, the algorithm identifies which specific variables to replace. If one variable is to be replaced, it is selected probabilistically based on weights derived from the inverse mutual information between each variable in the current set and candidate variables. If two variables are to be replaced, the pairwise inverse mutual information is used similarly to determine selection probabilities. Specifically, the inverse mutual information values between each pair of current and candidate

variables are calculated, normalized into a probability distribution, and used to sample the variable(s) for replacement.

For example, if two variables are to be replaced based on the inverse mutual information metric, the algorithm calculates the inverse mutual information between each variable in the current set and all candidate variables. Variables that exhibit lower mutual information (higher inverse mutual information) are more likely to be replaced, thereby enhancing synergy by favoring replacements that reduce redundancy. This selection method guides the search toward more synergistic sets by reducing information overlap.

The replacement weights were chosen from the pairwise metrics that our regression models ranked most predictive of triplet synergy. Their effect on search performance is reported in Table 3.1.

3.3.4. Particle Swarm Optimisation for synergistic associations

The PSO algorithm uses a collection of P particles to explore the search space for synergistic associations. Each particle starts at a random position, where a position represents a specific set of variables used to calculate synergy. Particles track their best positions and scores (personal bests) while also accessing the global best position and score identified by any particle in the swarm. The swarm iteratively updates particle positions based on their current velocity (V_i), their historical best-known positions (P_{local_i}), and the swarm-wide best-known position (P_{global}).

At each iteration, particle i considers a set of three variables. This set, denoted X_i^t, is scored based on the synergy of the variables it contains. Both metrics are treated within the minimisation framework used throughout, so a lower score indicates a more synergistic set. If the score at X_i^t improves on the particle's personal best, the particle updates its record. If that score also improves on the global best, the global best position is updated for the swarm:

$$\text{If } \Omega(X_i^t) < \Omega(X_{\text{local}_i}), \text{ then } X_{\text{local}_i} = X_i^t$$

$$\text{If } \Omega(X_{\text{local}_i}) < \Omega(X_{\text{global}}), \text{ then } X_{\text{global}} = X_{\text{local}_i}$$

Particles move probabilistically at each step, similar to the SA approach. A normal distribution sample determines how many components to change in the current position. The further the sampled value is from the mean, the more elements are replaced in X_i^t. When within one standard deviation, a single variable in X_i^t is replaced. This replacement is selected based on a blended probability distribution that considers mutual information with other candidate variables and nudges particles toward both the global best and their local best positions. This blended approach encourages particles to explore promising regions influenced by both personal success and swarm success.

Probability Distribution Update

The initial probability distribution $P = [p_1, p_2, \ldots, p_n]$ is derived from pairwise metrics such as mutual information. This distribution is then adjusted to incorporate information

from the current local and global bests, forming an updated distribution P'. The nudge distributions Δ_{local} and Δ_{global} place their mass on the variables occurring in the particle's local best position and in the swarm's global best position. The nudge factor sets how much weight is taken from P. Parameters α and β modulate the influence of the local and global bests, and their values are fixed by the nudging strategy in use, described below. Blended weights are calculated by combining the original distribution with these "nudge" distributions:

$$\text{blended_weights} = (1 - \text{nudge_factor}) \cdot P + \alpha \cdot \Delta_{\text{local}} + \beta \cdot \Delta_{\text{global}} \tag{3.1}$$

The updated probability distribution P' is obtained by normalising the blended weights so that they sum to one:

$$P' = \frac{\text{blended_weights}}{\sum_{i=1}^{n} \text{blended_weights}_i}$$

This probability update was developed from our regression analysis. That analysis identified which pairwise metrics predict triplet synergy. The distribution used to propose new positions is therefore biased toward that pairwise structure. By integrating these metrics, PSO's probability distributions and adaptive nudging parameters are refined, allowing particles to be directed more effectively toward promising search areas and enhancing both efficiency and accuracy in the synergy search.

Particle Movement Strategies

Various movement strategies were implemented using different pairwise metrics to compute candidate probabilities before applying nudges. These strategies were:

- uniform probabilities
- mutual information between the removed node and candidate nodes
- inverse mutual information between the removed node and candidate nodes
- mean (inverse) mutual information between the remaining nodes and candidate nodes
- mean mutual information between the remaining nodes and candidate nodes
- a combined strategy using the product of entropy and mutual information

Blended probabilities are computed as described in Equation (3.1). They balance the current movement direction (influenced by mutual information) with a tendency to explore new areas based on both local and global best positions.

Four nudging strategies were also explored to balance exploration and exploitation. No nudging acted as a baseline, where particles move independently of local and global bests. Constant nudging provided a fixed nudge factor that directed particles toward local

and global bests. With exponential decay nudging, nudge strengths decreased over time according to:

$$\text{LocalNudge} = \text{GlobalNudge} = \text{NudgeFactor} \cdot e^{-\frac{t}{\text{max_steps}}}$$

In an adaptive nudging strategy, nudge strengths instead adjust dynamically, guiding particles toward regions with high synergy based on current performance and proximity to the best scores. Nudge factors α and β are dynamically adjusted based on particle performance and proximity to the best scores, using sigmoid functions:

$$\text{Local Nudge} = \sigma\left(\frac{t}{\text{max_steps}} + \alpha \cdot \|\text{cost}_i - \text{global_best_score}\|\right)$$

$$\text{Global Nudge} = \sigma\left(\frac{t}{\text{max_steps}} + \beta \cdot \|\text{cost}_i - \text{particle_best_score}_i\|\right)$$

where $\sigma(x) = \frac{1}{1+e^{-k \cdot x}}$ is the sigmoid function. Here α, β and k are scaling factors of the adaptive schedule rather than the nudge coefficients of Equation (3.1). The resulting Local Nudge and Global Nudge are the values those coefficients take.

3.3.5. A clique-based heuristic to identify synergistic interactions

The third and final approach we have developed is a clique-based heuristic algorithm. Its aim is to identify synergistic interactions directly, without stochastic search or brute-force calculation. In this iterative approach, we construct a pairwise network by adding edges based on an aggregated pairwise information-theoretic measure. The heuristic rests on a single assumption: if high synergy is associated with low pairwise conditional entropies, then each pair of variables in a synergistic triplet should also have a low conditional entropy. We therefore add edges one at a time, in descending order of the sorted values for the pairwise metric correlated with synergy. Such an association is aggregate and need not hold pair by pair. In the Results we compare candidate pairwise metrics against the ground truth synergy scores. For example, when computing triplets, we add edges until a specified number of three-cliques (i.e., triangles) are present in the network. The algorithm keeps track of the cliques formed and halts once that number is reached, set to 10,000 cliques in the runs reported below. These cliques are the candidate synergistic n-plets (sets). The result is an ordered list of cliques extracted from the graph, each representing a potentially synergistic set. A discussion of the algorithm's computational complexity is given in Appendix B.

3.3.6. Reducing the number of synergistic sets to analyse

The number of sets X of p nodes with $\Omega(X) < 0$ is typically too large to test or interpret directly. Two objectives guide the reduction. First, every node in a retained set should contribute to the synergistic information. Second, that information should not be redundant with respect to sets already accepted.

To determine the statistical significance of a synergistic set, a permutation test or a bootstrapping procedure is often employed [103]. Under bootstrapping, a set is categorised as significantly synergistic when its confidence interval for the O-information excludes zero and its p-value falls below a predetermined threshold. However, when a large number of echoes are present in the data, resulting in an inflated number of synergistic triplets, a large, and often unfeasible, p-value adjustment may be needed to correct for (potential) false discovery rates when making multiple comparisons. By first removing echoes it becomes feasible to test the statistical significance of synergistic sets.

Once synergy scores had been computed for all sets of order p, we sorted them in order of decreasing synergy and initialised an empty set of accepted synergistic sets. We then entered a loop that continued until there were no remaining sets to process. In each iteration, the most synergistic set is selected. For each set in the sorted data frame, if the set has all but one element overlapping with the most synergistic set, the synergy of the merged set (union of the two sets) is computed. If the merged set's synergy is less than that of the overlapping set, the overlapping set is marked for rejection. At the end of each iteration, the most synergistic set is added to the accepted sets if it has not been rejected, and all sets marked for rejection are removed from the data frame. This process repeats, updating the data frame by removing accepted and rejected sets, until no more sets remain to be processed. The final result is a list of accepted synergistic sets.

We argue against removing variables prior to computation merely because they are highly correlated. Although highly correlated variables often exhibit similar pairwise associations with other variables, this pattern does not necessarily extend to higher-order associations. Later, we present examples of strongly correlated variables with distinct synergies, demonstrating how removing one variable can result in the omission of important synergies from the analysis.

3.3.7. Improvement factor for qualitative biological insights

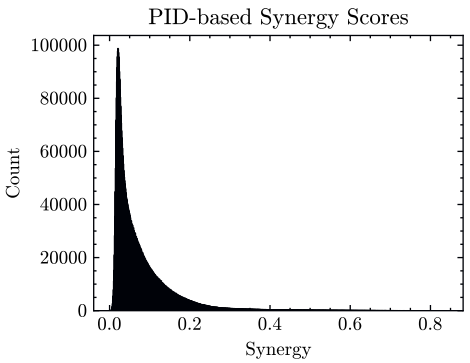
In addition to identifying the top triplets with the lowest bias-corrected O-information, we use the improvement factor to assess the biological relevance of certain biomarker interactions. The improvement factor is calculated as the ratio of the joint mutual information (MI) of two variables about a third variable to the sum of their individual pairwise MI values. When the improvement factor exceeds 1, it suggests that the combination of two variables provides additional, unique information about the third beyond what each variable contributes independently. Higher improvement factors, therefore, indicate stronger synergies, where interactions between variables reveal insights into complex, higher-order relationships not captured by pairwise associations alone.

In this chapter, the top ten triplets with the largest improvement factors were analysed alongside the top ten triplets with the lowest O-information to identify higher-order interactions relevant to lipid metabolism. By examining triplets with high improvement factors, we gain insight into interactions that might be relevant to lipid homeostasis and metabolic processes, beyond what pairwise correlations alone reveal.

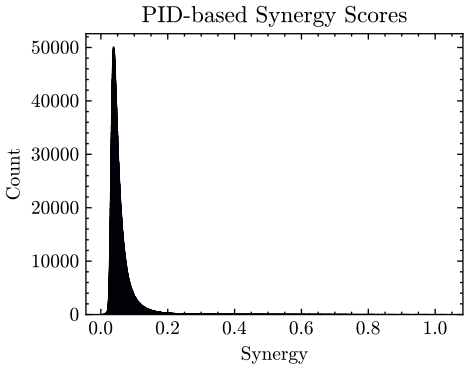
3.4. Results

3.4.1. The prevalence of synergy in data

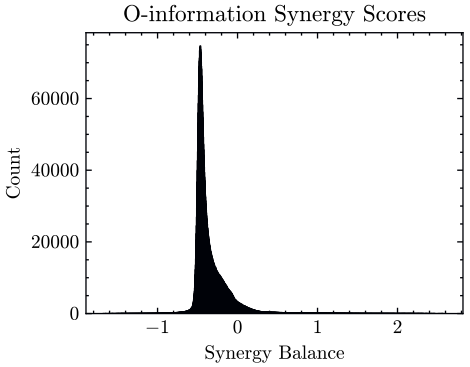
Strong synergy was rare in both datasets and under both metrics. Figure 3.1 illustrates the distribution of PID-based synergy and O-information scores in the UKB NMR data. In Figure 3.1a, most synergy scores were concentrated near the lower end of the scale, with frequency declining rapidly as the score increased. The distribution was therefore highly skewed with a long right tail, and only 5.8% of triplets had a synergy score above 0.2. This skewed distribution pattern suggests that most interactions within the dataset exhibit low synergy, while a few interactions achieve significantly higher synergy, potentially representing unique cases of high collaborative effects. The O-information scores in Figure 3.1c showed the same pattern: most triplets had negative scores, indicating possible synergy, but only a small subset of these were strongly synergistic. The mean score was -0.35, while 9% of triplets had an O-information score below -0.5 and only 0.4% had a strongly synergistic score below -0.7. Similar trends were seen in the YFS data, as shown in Figure 3.1d and Figure 3.1b.



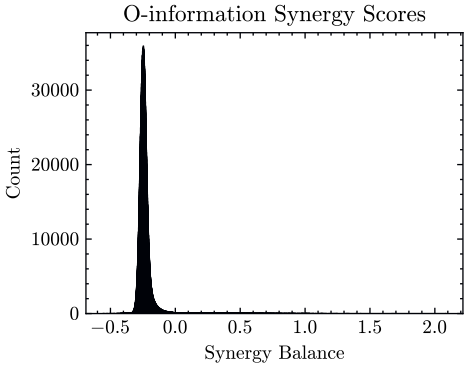
(a) UKB NMR PID-based Synergy Scores using data discretised into four states via quantile binning.



(b) YFS PID-based Synergy Scores using data discretised into five states via quantile binning.



(c) UKB NMR O-information Synergy–redundancy balance Scores using data discretised into 12 states via quantile binning.



(d) YFS NMR O-information Synergy–redundancy balance Scores using data discretised into 12 states via quantile binning.

Figure 3.1: Distribution of synergy scores across the YFS dataset and the NMR variables in the UKB data, using two different synergistic metrics and different data binning. The PID-based approach in (a,b) shows a notable skew towards lower values. The data exhibits a heavy right tail, indicating that higher synergy scores are less common. Similarly in (c,d), the majority of triplets have a very low synergy score (near zero). In the UKB NMR data only 0.4% have a strongly synergistic score below −0.7, while the minimum synergy score in the YFS data is −0.56.

3.4.2. Predicting synergy from lower-order features

For predicting PID-based synergy, RFECV retained the following variables in the YFS dataset: Mutual Information Sum, Mutual Information Spread, Entropy Spread, Spearman Correlation, Kendall Correlation, Chi-Squared, Minimum Mutual Information, Maximum Mutual Information, Variation of Information, Normalised Variation of Information, Information Overlap, Normalised Information Overlap, Jensen–Shannon Divergence for three variables, and Co-Inertia. The UKB dataset retained the same variables plus Average

Entropy.

The selection for O-information retained fewer variables. For the UKB dataset, RFECV kept Mutual Information Sum, Mutual Information Spread, Minimum Mutual Information, Maximum Mutual Information, Custom Dual Entropy, Variation of Information, Normalised Variation of Information, and Information Overlap. The YFS dataset retained these along with Kendall's Tau between x and z, Normalised Information Overlap, and Jensen–Shannon Divergence. Custom Dual Entropy was calculated as the average of the sum of the entropies of each variable minus the entropies of their pairwise combinations.

We therefore used combinations of these variables in our regression analysis to identify the pairwise features best suited to our search algorithms. RMSE fell as the number of features increased (Figure 3.2). This indicated that models incorporating a greater number of pairwise features generally exhibited improved predictive accuracy. Notably, certain combinations significantly outperform others with a large variability in RMSE scores for each feature count. The models with the best predictive power are observed with the inclusion of approximately 9 to 12 features, where the RMSE scores reach their minimum values.

The best model achieved RMSEs of 0.015 and 0.009 for predicting O-information and PID-based synergy scores respectively in the UKB NMR data. Appendix Tables B.1, B.2, B.3, and B.4 show the scores when predicting PID-based synergy from combinations of six or fewer pairwise metrics. This showed that combinations of pairwise metrics could indicate the strength of synergy within a set of variables. However, computing a vast array of pairwise features to predict synergy values does not immediately benefit our goal of quickly identifying synergy in datasets as gathering the training data also suffers from the combinatorial explosion facing synergy estimators. Therefore, our objective is to identify a robust set of features that can be used in search algorithms as proxies to guide our search efficiently.

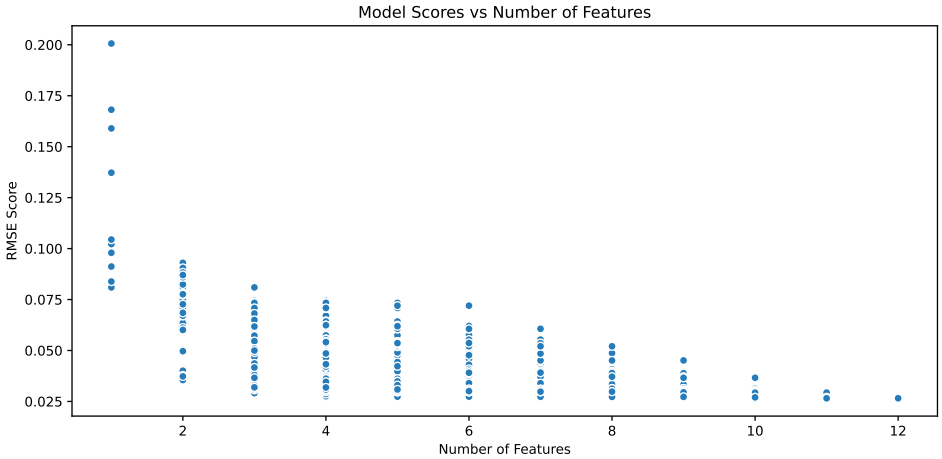


Figure 3.2: RMSE scores of the LightGBM model for predicting O-information scores based on different combinations of pairwise features. This plot illustrates the relationship between the number of features and the model's prediction accuracy for O-information on the UKB NMR data. As expected, increasing the number of features generally leads to improved model accuracy, resulting in lower RMSE scores. A similar trend was found for both synergy metrics and both datasets.

For PID-based synergy in the UKB NMR data, the best model restricted to at most six features used Mutual Information Sum, Mutual Information Spread, Minimum Mutual Information, Maximum Mutual Information, Variation of Information and Normalised Variation of Information, reaching an RMSE of 0.0169. If only two features were used, the best pair was Mutual Information Spread and Information Overlap, with an RMSE of 0.0541. Among individual variables, Mutual Information Sum was the most effective, at an RMSE of 0.0657. Mutual Information Sum and Mutual Information Spread together reached 0.0556, and Minimum Mutual Information alone reached 0.0674. As the top performing combinations rely on variants of mutual information and conditional entropy, these are the pairwise measures that are included in our subsequent search strategies.

3.4.3. Reducing the number of synergistic sets to analyse: remove triplets after computing synergy

Echo filtering removed most of the top-ranked synergistic triplets. The heuristic of Section 3.3.6 rejects an overlapping set when the merged set is less synergistic than the overlapping set itself. Here we screened the n most synergistic triplets in the UKB NMR dataset, scored by O-information, with a cheaper mutual information variant of that rule. A triplet sharing two of three variables with a higher-ranked triplet was rejected when the mutual information between the non-overlapping variables exceeded a threshold τ. Figure 3.3 shows the proportion of triplets retained across a grid of thresholds, with the top n ranging from 100 to 10,000 and τ from 0 to 1 in 20 steps. At $\tau = 0.6$, 62% of the top

100 triplets and 80% of the top 1,100 triplets were rejected as echoes. That proportion reflects the stringency of the criterion rather than the prevalence of spurious synergy, since a triplet is labelled an echo precisely when the mutual information exceeds τ. Filtering is nonetheless advisable before significance testing or biological interpretation, because a large number of echoes forces a multiple-comparison correction that is often unfeasible. A high mutual information score does not guarantee that two variables are interchangeable, as the next subsection shows, so the criterion will also reject sets that are not echoes. Echo filtering was not applied in the clinical applications of Chapters 4–6. How many of the associations reported there would survive it remains an open question.

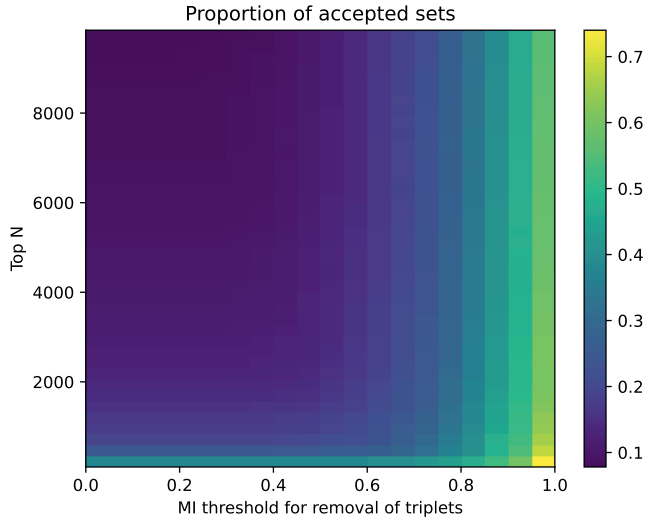


Figure 3.3: Prevalence of echoes in UKB NMR. If we select the top n synergistic sets in the data (y-axis) and then set an MI threshold for similarity (x-axis), the plot shows the proportion of synergistic sets that would remain after filtering. Most applications would lie in the bottom right of this figure.

3.4.4. Reducing the number of synergistic sets to analyse: remove variables before computing synergy

To reduce the computational burden of studying complex systems with a large number of variables, researchers often use feature selection methods to first reduce the number of variables to analyse. Methods such as Correlation-based Feature Selection, Mutual Information-based Feature Selection and Forward/Backward Selection examine pairs of variables to decide which to remove. Pairwise screening can drop a variable that is synergistic only in combination with others. We therefore argue that the feature set should be reduced after synergistic associations are computed, not before.

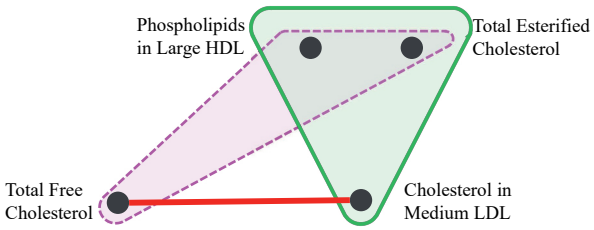


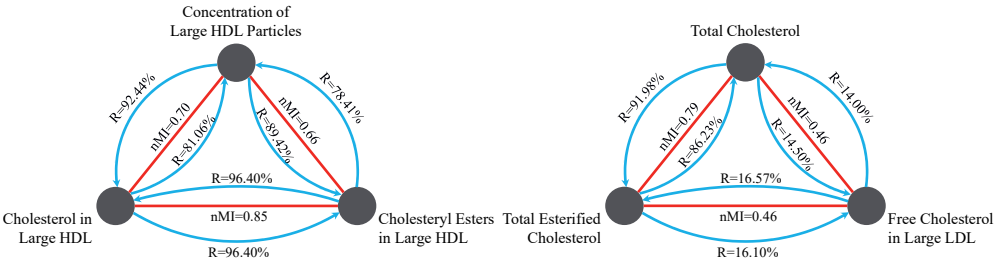
Figure 3.4: Example of a non-echo. *Total Free Cholesterol* and *Cholesterol in Medium LDL* have a strong MI score (red line). *Cholesterol in Medium LDL* forms a high synergy score with two other variables (green triangle). However, the triplet containing *Total Free Cholesterol* instead of *Cholesterol in Medium LDL* does not form a highly synergistic triplet (dashed purple triangle). This shows that if *Total Free Cholesterol* was used in the search for synergy, but *Cholesterol in Medium LDL* was not, then this synergistic association would be missed.

To illustrate this point, Figure 3.4 shows a set of four variables: *Cholesterol in Medium LDL*, *Total Free Cholesterol*, *Total Esterified Cholesterol*, and *Phospholipids in Large HDL*. *Cholesterol in Medium LDL* and *Total Free Cholesterol* have a high mutual information score of 1.2, placing them in the 97.6th percentile of all MI scores. However, the variables exhibit different synergies when combined with other variables. The triplet (*Cholesterol in Medium LDL*, *Total Esterified Cholesterol*, *Phospholipids in Large HDL*) is strongly synergistic, with a bias-corrected O-information score of -1.22, while the triplet (*Total Free Cholesterol*, *Total Esterified Cholesterol*, *Phospholipids in Large HDL*) scores only -0.16.

We repeated this check across the UKB NMR data. For each highly correlated pair, we took every highly synergistic triplet containing one member of the pair and substituted that member with the other. Let N be the number of such triplets and k the number that remain highly synergistic after substitution. The retention rate is then $R = (k/N) \times 100\%$, plotted on the curved blue lines in Figure 3.5. A low R means that two variables with a strong pairwise association carry very different synergies.

Correlated variables do not always share their synergies. Notably, Figure 3.5a depicts a scenario where three variables have strong pairwise associations and generally maintain their synergies. In contrast, Figure 3.5b demonstrates a case where variables, despite having strong pairwise associations, exhibit very different synergies, as evidenced by the low R scores. This analysis underscores the importance of retaining all variables to capture the full spectrum of higher-order associations in the system, rather than using univariate methods to reduce the size of the system. By retaining all relevant variables until the synergistic interactions are fully explored, we can uncover extra synergies that

would otherwise remain hidden. This approach ensures a more comprehensive and accurate understanding of the underlying complexities in the system.



(a) Set of triplets with mostly the same synergies

(b) Set of triplets with differing synergies

Figure 3.5: Examples of highly correlated sets of variables in the UKB NMR data. For each highly synergistic triplet (defined as having less than twice the mean O-information synergy score) containing each variable, we substitute an element from the triplet with a highly correlated variable from this set and record the retention rate R, the percentage of triplets that remain highly synergistic, shown by the percentages on the curved blue lines. The arrow on the curved blue lines indicates that the variable (one of the highly correlated variables) in the highly synergistic triplets is replaced with the other variable in the highly correlated pair. Normalised mutual information (nMI) scores are depicted by the red straight lines. Figure (a) illustrates a case where three variables exhibit strong pairwise associations and share similar synergies, while Figure (b) depicts variables with strong pairwise associations but very different synergies, as indicated by the low values of R.

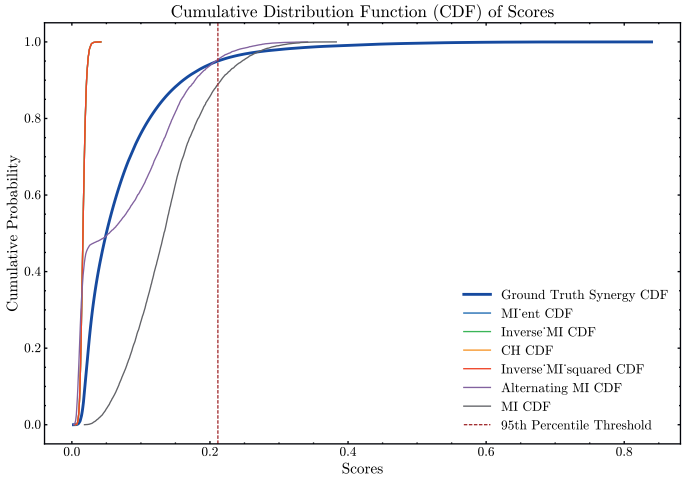
3.4.5. Clique identification in pairwise network

We compared the performance of different pairwise metrics in the clique search strategy using the UKB NMR and YFS datasets. Figures 3.6 and 3.7 show the cumulative distribution functions (CDFs) of PID and negated O-information synergy scores for each strategy. The ground truth synergy scores are highlighted with a thicker blue line, and percentile thresholds are marked with dashed lines. Curves that are initially below and to the right of the solid blue line indicate better performance in identifying highly synergistic variable sets. The performance of different pairwise metrics varies with the synergy metric. PID-based synergy metrics perform similarly across datasets (Figures 3.6a and 3.6b), as do O-information-based synergies (Figures 3.7a and 3.7b).

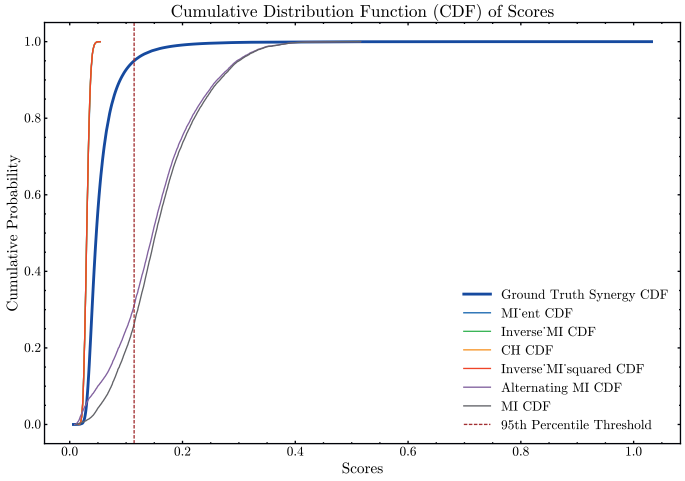
For the UKB NMR dataset PID synergy scores, the strategies *MI ent* (computed as $H_{XYZ} \odot e^{I^2_{XYZ}}$), *Inverse MI*, *CH*, and *Inverse MI squared* displayed identical performance metrics. The mean score was approximately 0.0164, the median 0.0161, the standard deviation 0.0039, and the interquartile range (IQR) 0.0051. These four scores return near-identical statistics for both datasets and both synergy metrics, so they are best read as one strategy rather than as four. None of their scores surpassed the 50th percentile of the ground truth, indicating limited effectiveness in identifying high-synergy sets. The Alternating MI strategy (motivated by the *MI spread* in the regression analysis) exhibited a higher mean score of approximately 0.0768 and a median score of 0.0541, with a higher standard deviation of 0.0723 and an IQR of approximately 0.1226. This strategy performed better, with 50.63% of its scores above the 50th percentile of the ground truth, though its

effectiveness decreased at higher percentiles. The *MI* strategy outperformed all others, with a mean score of approximately 0.1381, a median score of 0.1334, a standard deviation of 0.0577, and an IQR of 0.0738. It placed 95.86% of its scores above the 50th percentile and 11.08% above the 95th percentile of the ground truth, well ahead of every other strategy.

Similarly, for the YFS dataset, *MI ent*, *Inverse MI, CH*, and *Inverse MI squared* again produced almost no scores above the ground truth percentiles. The Alternating MI strategy performed well, with 90% of its scores above the 50th percentile of the ground truth and 68.86% above the 95th percentile. The *MI* strategy performed better still, with a mean score of 0.1622, a median score of 0.1526, a standard deviation of 0.0736, and an IQR of 0.0928. It placed 96.40% of its scores above the 50th percentile and 74% above the 95th percentile of the ground truth. The exact scores can be seen in Appendix Tables B.7, B.8, B.9, and B.10.

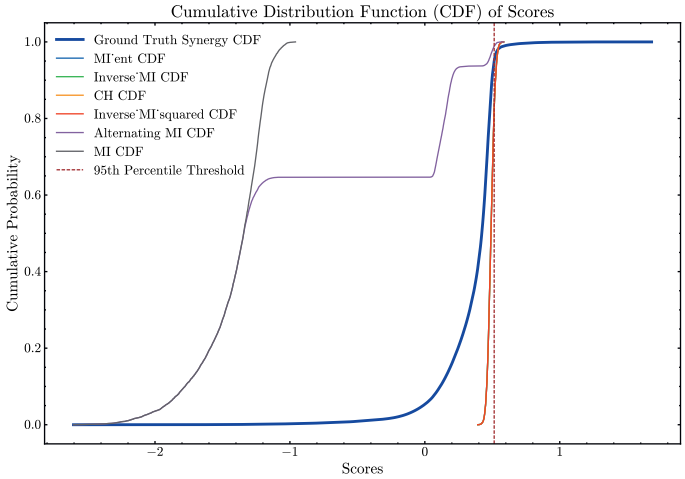


(a) CDF of PID scores for different pairwise metrics with the clique search strategy on UKB NMR.

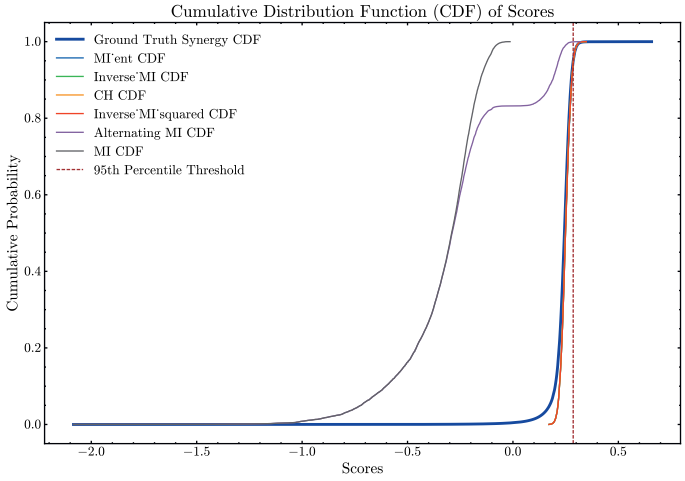


(b) CDF of PID scores for different pairwise metrics with the clique search strategy on YFS.

Figure 3.6: CDF of PID scores for different pairwise metrics with the clique search strategy on UKB NMR and YFS datasets.



(a) CDF of negated O-information scores for different pairwise metrics with the clique search strategy on UKB NMR.



(b) CDF of negated O-information scores for different pairwise metrics with the clique search strategy on YFS.

Figure 3.7: CDF of negated O-information scores for different pairwise metrics with the clique search strategy on UKB NMR and YFS datasets. These plots compare the performance of each strategy against the ground truth synergy scores, with the ground truth CDF highlighted by a thicker blue line for distinction. Percentile thresholds are indicated by dashed lines. Better performance is represented by curves that are initially below and to the right of the solid blue line.

The analysis using the O-information metric across both the UKB NMR and YFS datasets revealed substantial performance differences among the scoring strategies, as shown in Appendix Tables B.11, B.12, B.13, and B.14. For the UKB NMR dataset, the strategies *MI ent*, *Inverse MI*, *CH* and *Inverse MI squared* performed similarly, with 99%

of their scores exceeding the 50th percentile of the ground truth. Their effectiveness diminished at higher percentiles, with only around 17.34% of their scores surpassing the 95th percentile. In the YFS data these strategies had 59.43% of their scores above the 50th percentile and about 5.52% above the 95th percentile. The *Alternating MI* strategy had only 6.30% and 1.37% of its scores above the 50th percentile in the UKB NMR and YFS data respectively, while none of the *MI* scores exceeded the 50th percentile. This difference between the datasets is likely due to the YFS data having differing numbers of states per variable, while the UKB NMR data have the same number of states across variables. The contrast between the two synergy measures is sharper. Under PID, *MI* was the strongest guiding metric, with 95.86% and 96.40% of its scores above the 50th percentile in the UKB NMR and YFS data. Under the O-information it was the weakest, and the other four strategies showed the mirror-image pattern. The O-information measures the balance between synergy and redundancy, so a triangle assembled from high-MI edges tends to be redundancy-dominated. The PID synergy atom is non-negative and subtracts no redundancy, so the same triangle can still score highly. The search framework is metric-agnostic. The pairwise feature that guides it is not, and should be re-selected for each synergy measure. After the pairwise metric is computed (brute-force) for the dataset, the algorithm has an average runtime of 240 seconds when searching for 10,000 triangles, using the same computer as described in the simulated annealing section.

3.4.6. Simulated Annealing for synergistic associations

Our approach expands upon the methodology presented in [118] by incorporating non-uniform probability distributions for selecting new node sets. Unlike the approach in [118], where replacement nodes are chosen randomly - a method we describe as 'no weight' and associate with the geometric cooling strategy - we introduce three alternative variants. These variants are designed based on the probability distribution of potential replacement nodes, defined as follows:

1. Using a probability distribution $1 - \frac{MI}{\max(MI)}$.
2. Using mutual information directly.
3. Using $H_{XYZ} \odot e^{I^2_{XYZ}}$.

In the third variant, $\odot$ denotes element-wise multiplication between matrices, H_{XYZ} is a vector containing the entropy of individual variables, and $e^{I^2_{XYZ}}$ is the exponentiated matrix of pairwise mutual information values. These features were derived from our pairwise prediction models. The metric combines the two quantities so that a set scores highly when entropies are large and mutual information is comparatively low, and neither H nor MI dominates the score.

Table 3.1 reports the best scores found by each configuration on the discretised UKB NMR data, all under the geometric cooling schedule. *Inverse MI*, *no weights*, *MI* and *Inverse MI squared* were near-identical in central tendency, with means between -1.4304

and -1.4469 and medians between -1.4554 and -1.4803. They differed in consistency: *Inverse MI* and *no weights* had the narrowest interquartile ranges, at 0.1398 and 0.1391. *MI ent* was worse on every measure, with a mean of -1.3556, a median of -1.3671 and an interquartile range of 0.2304. The synergy threshold counts how often a run terminated on one of the 107 most synergistic triplets in the data, ranked by brute-force O-information. *Inverse MI* and *no weights* reached it in 68 and 66 of 100 runs, *MI ent* in only 42. Figure 3.8 shows the same ordering in the full distributions of best scores. Statistical tests further confirmed that MI ent geometric was significantly different from the other configurations, which performed similarly to each other, as seen in Figure 3.8.

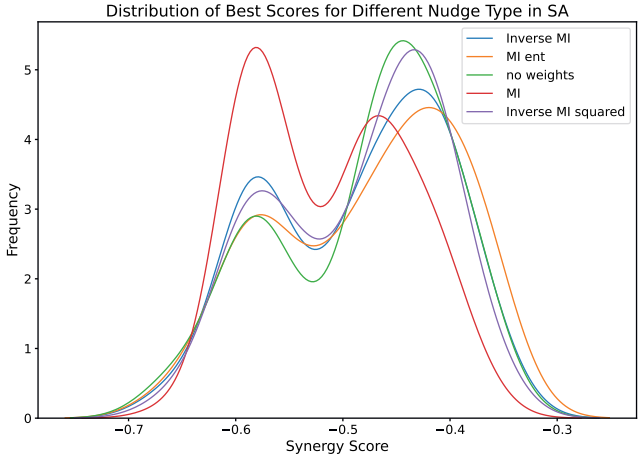
The cooling schedule also mattered. Under Boltzmann cooling, an alternative temperature schedule to the geometric one used above, the algorithm found highly synergistic triplets less consistently. With the *Inverse MI* weighting it often stalled on suboptimal solutions, as shown in Figure B.4(a). When it did reach a highly synergistic triplet, it converged cleanly, as shown in Figure B.4(b,c). This behaviour suggests that while the SA algorithm can perform well under certain conditions, it often struggles with escaping local optima and requires further refinement to enhance its reliability and consistency.

The performance of various SA configurations in identifying the most synergistic sets on the UKB NMR data using the O-information metric is summarised in Table 3.1. Similar results for the YFS data are shown in Table B.15. All configurations exhibited similar mean and median best scores. Notably, the inverse MI and no weights configurations demonstrated narrower interquartile ranges (IQR), indicating more consistent results compared to other approaches. Additionally, a synergy threshold was set to evaluate how frequently each SA run identified a highly synergistic set among the global best 107 sets. While all configurations consistently reached this threshold, the MI ent strategy did so less frequently. These results suggest that the SA algorithms can effectively estimate the most synergistic triplets without exhaustive computation or extensive random sampling.

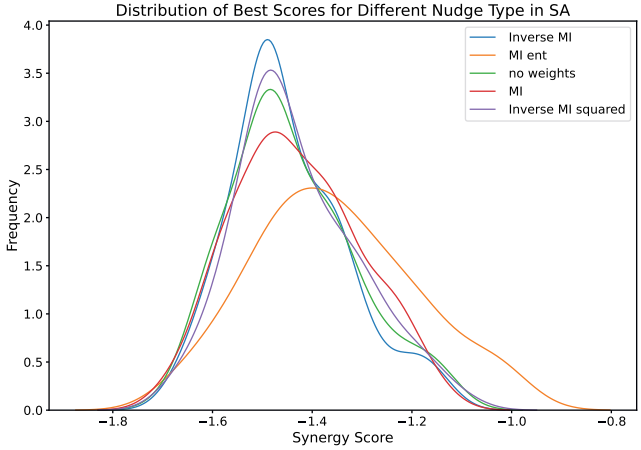
Configuration	Mean	Median	Std Dev	IQR	Below Threshold
Inverse MI	-1.4469	-1.4803	0.1149	0.1398	68
MI ent	-1.3556	-1.3671	0.1612	0.2304	42
no weights	-1.4403	-1.4803	0.1238	0.1391	66
MI	-1.4315	-1.4554	0.1243	0.1569	62
Inverse MI squared	-1.4304	-1.4584	0.1221	0.1504	65

Table 3.1: Summary Statistics for the best synergistic sets found by different SA configurations on the UKB NMR data using O-information. All approaches have similar mean and median best scores. The interquartile range (IQR) for the inverse MI and the no weights setups are narrower than other approaches, indicating more consistent results. We also set a synergy threshold to determine how many times each SA run found a synergistic set in the global best 100. All approaches consistently reached this threshold, although the MI ent strategy did so less frequently than the others. Reaching the threshold indicates that the SA algorithms can accurately estimate the most synergistic triplet in the data without needing to brute-force calculate or randomly sample a large number of triplets.

The training trajectories of the SA algorithm, shown in Figure B.4, varied considerably



(a) Distribution of best synergy scores found in the YFS data for 100 runs of the SA algorithm.



(b) Distribution of best synergy scores found in the UKB NMR data for 100 runs of the SA algorithm.

Figure 3.8: Distribution of best synergy scores found in the YFS and UKB NMR data for 100 runs of the SA algorithm using different strategies to guide particle movement. Synergy was computed via the O-information. The SA algorithm used a geometric cooling strategy with random initialisations for each run. Lower values (to the left) indicate triplets with more synergistic information were found. Each particle movement strategy converged multiple times to triplets in the overall top 100 synergistic triplets in each dataset.

across runs. The optimisation landscape appeared to be relatively flat, so some runs quickly found optimal solutions while others became trapped in local minima. Running SA multiple times is therefore necessary to ensure that highly synergistic sets are found. A particle in the SA search may accept a worse solution, and returning to the better solution appears not always to be possible because of this difficult landscape. To address this challenge, employing a PSO algorithm could be advantageous, as it allows the swarm of particles to collaboratively explore the search space, potentially enabling quicker escapes from flat regions and always nudging particles back toward their local optima, further improving the overall search efficiency.

The SA algorithm was executed 100 times for each parameter setting, with each run comprising 1,000 samples. Running times ranged from approximately 30 to 57 minutes per complete set of runs. The simulations were run on a laptop with a 12th-generation Intel Core mobile processor, 10 physical cores, a maximum clock frequency of 1.7 GHz and 31.64 GB of RAM. Multiprocessing would reduce these times further when the algorithm is deployed on larger datasets.

3.4.7. Particle Swarm Optimisation

The PSO algorithm was evaluated across different parameter settings for nudge and weight types. Figures B.5 and B.6 show how the global best score evolved over time for each configuration, where the same weight types were used as in the SA algorithm. Some configurations converged faster and more consistently than the no-nudge baseline. However, performance did not appear to differ substantially between nudge types.

Figure B.3 presents the initial and final scores for each particle under one parameter setting. This figure illustrates that the PSO algorithm successfully guides particles towards better solutions, as evidenced by the significant improvement in scores from start to end. Unlike the SA approach, where particles often get stuck in local minima, the PSO algorithm effectively explores the search space, ensuring more particles reach optimal or near-optimal solutions.

3.4.8. Synergistic information patterns in lipid metabolism

To further investigate the identified complex interactions among NMR metabolomics biomarkers, we performed univariate and bivariate mutual information analyses on the top ten biomarker triplets with the lowest bias-corrected O-information. (Please see the Appendix Table B.5 and Figure B.2 for the top ten triplets in terms of the largest improvement factors in the UK Biobank NMR dataset.) These triplets spanned thirteen unique biomarkers, described in Table B.6 and Figure 3.10. We grouped and coloured these biomarkers by their roles in cholesterol transport, lipoprotein particle concentration, phospholipids, and apolipoproteins. Each triplet combined biomarkers from one or more of these functional groups, providing a structured basis for interpreting biologically meaningful synergies. This four-way grouping is distinct from the nine-category annotation of the full NMR panel used in Figure 3.9.

We recall that the improvement factor quantifies the extent to which the joint MI

of two variables about the third exceeds the sum of their individual pairwise MI values. The distribution of improvement factors across strongly synergistic triplets, taken to be those with O-information below −0.4, is shown in Figure B.1. Most factors were between 3 and 5, with a peak around 4 (mean 4.13, standard deviation 0.69). This indicates that, on average, the joint interaction between two variables provides roughly four times more information about the third variable compared to pairwise associations. In other words, if we focused only on pairwise associations, we would miss about 75% of the explanation (reduced conditional entropy) of each variable in these triplets. A long tail extends towards values up to 8, marking a few triplets with exceptionally strong synergies. Such triplets may involve complex biological interactions that pairwise MI does not capture at all. A pairwise screen alone would rank them as unremarkable. This highlights the added value of our analysis.

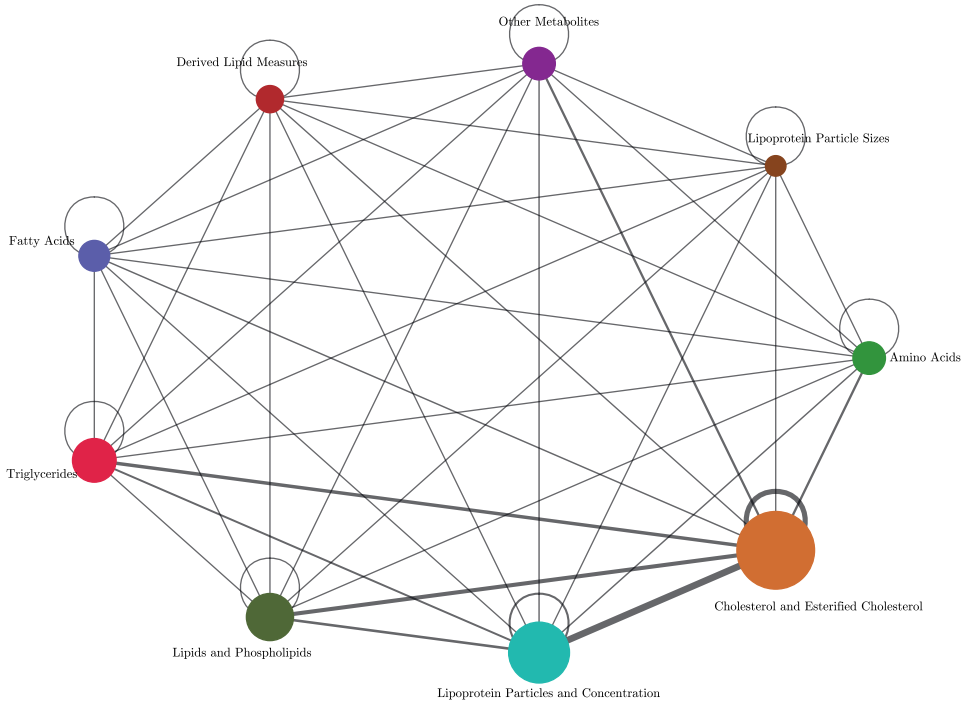


Figure 3.9: Network visualisation of synergistic interactions between metabolite categories. Each node represents a category of metabolites, and the size of each node reflects the number of variables in the UKB NMR data assigned to that category. Edge thickness represents the number of synergistic triplets between pairs of categories, with thicker edges indicating more synergistic associations between those variable categories. Node colours correspond to distinct categories: Amino Acids, Lipoprotein Particle Sizes, Other Metabolites, Derived Lipid Measures, Fatty Acids, Triglycerides, Lipids and Phospholipids, Lipoprotein Particles and Concentration, and Cholesterol and Esterified Cholesterol. Self-loops, indicating within-category synergy, are positioned outside the nodes to enhance clarity.

Cross-referencing these results with the variable categories, we find that triplets involving biomarkers from different functional categories frequently exhibit strong synergies across categories, as shown in the network of Figure 3.9. An edge is drawn between two categories whenever a synergistic triplet contains at least one variable from each, weighted by the number of times this occurs. All edges are present in the network, indicating that synergy is present between all variable categories. This indicates that triplets comprising biomarkers from distinct functional categories exhibit strong synergies between categories. The network therefore establishes that cross-category synergy is pervasive, not that it is stronger than synergy within a category.

However, it is interesting to note that the ten most synergistic triplets shown in Figure 3.10 draw their members from only six of the nine categories. The strongest synergies

are therefore concentrated in part of the panel rather than spread evenly across it. Turning to individual triplets, row 10 of Figure 3.10 shows that the triplet of HDL Cholesterol, Total Phospholipids, and Phospholipids in VLDL demonstrates moderate pairwise mutual information (MI) values, yet a substantial higher-order MI, highlighting a significant synergy between these variables. Similarly, the triplet including the Average Diameter of HDL, Apolipoprotein A1, and Total Lipids in Small HDL exhibited low pairwise MI values but revealed a high improvement factor when including Average Diameter, indicating interactions potentially related to HDL size, cholesterol efflux capacity, and functionality.

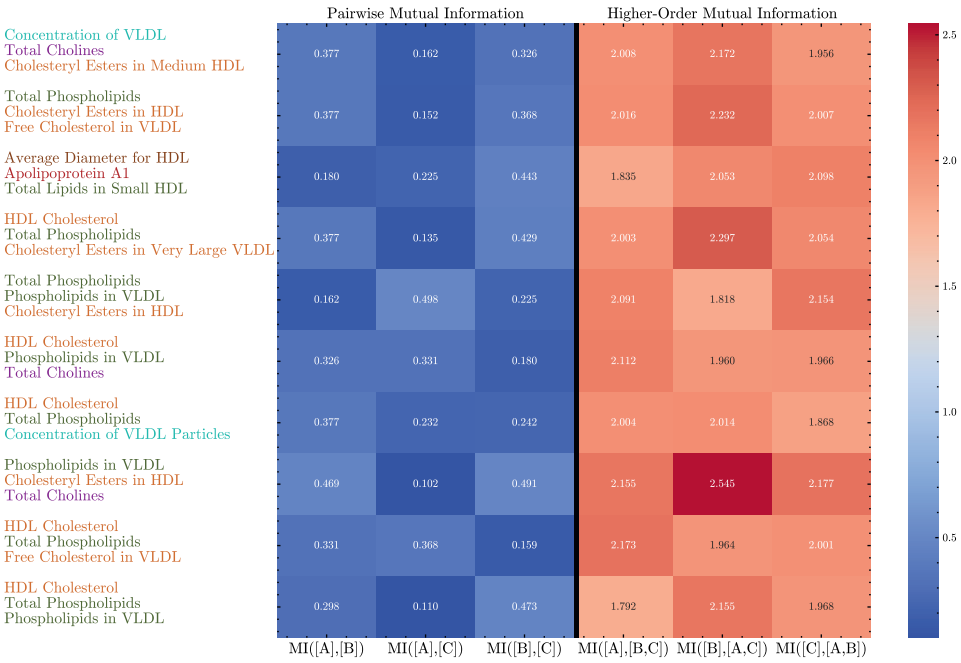


Figure 3.10: Heatmap of pairwise and higher-order mutual information (MI) for lipid metabolism triplets, with biological groupings. The heatmap displays the pairwise MI (columns 1–3) and the higher-order MI (columns 4–6) for each triplet, where variables are coloured based on their biological cluster. The strong synergies (higher-order MI) between variables from different clusters, such as HDL Cholesterol and Phospholipids in VLDL, highlight cross-functional interactions in lipid metabolism.

Plausible mechanistic interpretations

To further our qualitative analysis, we propose speculative interpretations of selected synergistic triplets, based on current knowledge of lipid metabolism. These interpretations probe the biological plausibility of our findings. They are not definitive mechanistic conclusions.

The triplet of HDL Cholesterol, Total Phospholipids, and Phospholipids in VLDL is the simplest case. Its synergy may reflect coordinated regulation of HDL and VLDL particles,

in which phospholipids are exchanged between the two classes to maintain lipid balance. Such exchange would affect lipid transport and cholesterol homeostasis.

A strong synergy was also observed in the triplet containing the Average Diameter of HDL, Apolipoprotein A1, and Total Lipids in Small HDL. The pairwise MI between Apolipoprotein A1 and Total Lipids in Small HDL was very low (MI = 0.102). Adding Average Diameter as a third variable raised the higher-order MI substantially (MI = 2.545). Two mechanisms could plausibly generate this pattern. We consider each in turn.

The first is structural. HDL particles are heterogeneous in size and composition, with ApoA1 serving as the primary structural protein [120]. ApoA1 forms a scaffold-like structure that encapsulates lipids in HDL particles, so varying the number of ApoA1 molecules per particle would affect both particle size and lipid capacity. ApoA1 would then act as a modulator: the amount available sets how HDL size relates to lipid content. In other words, more ApoA1 allows larger particles that can accommodate more lipids.

The second runs in the opposite direction. Lipid content also influences the size of HDL particles, so even with ample ApoA1 in the environment this second process can explain the synergy. ApoA1 has two conformational states, large and small, and the amount and type of associated lipid can drive the protein between them as it adapts to encapsulate different lipid quantities.

Many further biological processes act in parallel and may affect the information synergy that we eventually measure. Neither account is therefore confirmatory. That a coherent mechanistic story can be told for this triplet nonetheless suggests that the measured synergy is biologically meaningful. Where no mechanistic explanation, or no sufficiently detailed one, is yet available, our results may instead give direction to further investigation of such higher-order effects. This example also shows how variables from two different functional groups, lipoprotein particle size and apolipoprotein function, can act together to create complex, non-linear mechanisms.

One notable triplet – Total Lipoprotein Particle Concentration, Medium VLDL Particle Concentration, and Cholesteryl Esters in Medium HDL – demonstrated moderate pairwise MI values but an improvement factor of 6.2. One plausible explanation is that an increase in total lipoprotein particles could indicate elevated lipid flux, potentially triggering increased VLDL production by the liver. This rise in VLDL particles would provide more acceptors for cholesteryl esters transferred from HDL via Cholesteryl Ester Transfer Protein (CETP) activity. So far, the process is linear. Simultaneously, however, the heightened lipid flux might also stimulate reverse cholesterol transport, which increases cholesteryl ester content in medium HDL. Transferring those esters from HDL to VLDL would then tend to decrease HDL cholesteryl ester content. This happens during a process of lipolysis (making lipids smaller): as VLDL particles undergo lipolysis, they interact with HDL particles, exchanging lipids and apolipoproteins [121].

The second explanation is a coding artefact. The ‘Medium VLDL’ concentration is not monotonic over the course of this process, first increasing and then decreasing again, so a low value is ambiguous. If ‘Medium VLDL’ is low, the VLDL could be predominantly large or small. These are different biological states, and in the dataset they are measured

by variables that were not part of this triplet. Such interactions may become clear when analysing quadruplets, quintuplets, and larger sets.

The metabolic route is non-linear in the interplay of lipoprotein production, CETP-mediated lipid transfer, and reverse cholesterol transport. The coding route is non-linear in the variable itself. Either could explain the high higher-order MI despite low pairwise MI values. These speculative interpretations illustrate the kind of hypothesis-generating analysis that synergy search enables, but they remain untested.

3.5. Discussion

Guided stochastic search recovers the top-ranked synergistic triplets in both datasets at a small fraction of the cost of exhaustive enumeration. Many triplets that appear synergistic are echoes of strong pairwise associations. At a mutual information threshold of 0.6, 62% of the top 100 UK Biobank NMR triplets were rejected as echoes. Together these results address the two aims of the chapter: finding the most synergistic associations efficiently, and filtering out sets whose apparent synergy is an artefact of overlapping pairwise associations. The study of higher-order dependencies in multivariate data has grown quickly. Traditional heuristics for computing synergy do not scale, which is what makes efficient search-space navigation necessary.

Our methodology employs the stochastic strategies SA and PSO, alongside a direct network-based clique algorithm, to identify synergistic triplets. These methods alleviate the computational burden of exhaustive search. They perform well on datasets with a large number of variables. Incorporating lower-order statistical features into the search recovered the top synergistic associations about as reliably as the unguided random replacement used in [118]. The inverse MI weighting matched that baseline on both best score and run-to-run consistency, and only the MI ent weighting performed worse (Table 3.1).

Our results demonstrated that while SA could identify highly synergistic triplets, it often became trapped in local minima, requiring multiple runs to ensure comprehensive exploration. Conversely, the collaborative nature of PSO reduced how often particles remained trapped, which suggests more robust results. That comparison with SA is qualitative. The two algorithms were not evaluated on a common set of summary statistics. The clique-based heuristic algorithm, though faster, serves as a preliminary assessment tool rather than a definitive method for synergy identification.

A further contribution of this chapter is the identification of the echo problem. When strong pairwise associations connect to synergistic triplets, they inflate the apparent number of synergistic sets. Specifically, if two variables share a strong pairwise association, they can often be interchanged across synergistic triplets, leading to an explosion of identified triplets, many of which convey the same information. Because synergistic sets are commonly summarised as networks or hypergraphs, this inflation would be expected to propagate to node centrality, path-length distributions, and clustering, which can mislead downstream analysis. We did not quantify these effects here. We propose a naive regularisation step incorporated into the global optimisation process to filter out

these spurious associations. Removing sets whose apparent synergy is attributable to a strong pairwise association improves the reliability and interpretability of the results. The threshold that sets the stringency of this filter is chosen by the analyst, not by the procedure.

We also caution against removing variables before computing synergy. Highly correlated variables often carry different synergies with other variables, so univariate screening can discard the variables that participate in higher-order structure. The examples in Figure 3.5 showed synergies that surface only when all variables are retained. In practice this argues for screening for higher-order structure before, rather than after, any pairwise feature reduction.

The biological results of this chapter are a plausibility check on the search procedure rather than a contribution to lipid physiology. The synergistic triplets recovered by these algorithms span multiple lipoprotein classes, consistent with known regulatory links in lipid metabolism. Improvement factors for strongly synergistic triplets averaged 4.13, so a pairwise reading of these triplets misses about 75% of the information they carry. The precise mechanisms remain speculative.

Extending the approach to higher-order sets and alternative synergy measures is a natural next step; the General Discussion revisits this direction as part of a broader programme for dynamic higher-order inference. Further exploration of heuristic seeding for search algorithms and network community detection could enhance the efficiency and robustness of these methods.

3.5.1. Limitations

Several limitations should be acknowledged. First, the discretisation of continuous variables using Sturges' rule—a heuristic designed for histogram display rather than entropy estimation—may affect the magnitude of synergy scores. No sensitivity analysis with respect to the number of bins was performed, and alternative approaches such as Bayesian binning or kernel density estimation may yield different results.

Second, statistical power in the UKB NMR analysis is limited. At twelve states per variable a triplet's joint distribution spans 1,728 cells, estimated here from 336 participants, so the great majority are empty. The bias-corrected estimator of Gehlen et al. [113] adjusts for the systematic component of this error but not for the imprecision. O-information values from this dataset are therefore best read as rankings rather than as stable estimates. The YFS analysis, with 1,684 participants at the same resolution, is better powered.

Third, missing values in the UKB NMR data were imputed using K-nearest-neighbours regression, which induces pairwise correlations that do not exist in the original data. Since synergy captures precisely the information that cannot be explained by pairwise relationships, imputation through pairwise structure may systematically underestimate synergy. No complete-case sensitivity analysis was conducted.

Fourth, the echo-filtering threshold is a free parameter for which no principled selection criterion is provided. The heatmap in Figure 3.3 illustrates the trade-off between stringency and retention, but the optimal threshold will depend on the dataset and the

downstream application.

Fifth, although the search framework generalises in principle to sets of any order, only triplets were evaluated in this chapter. The performance of SA and PSO on quadruplets or larger sets remains untested. Finally, all evaluation was internal: algorithms were compared against brute-force rankings on the same datasets, without a synthetic benchmark containing planted synergies where true-positive and false-positive rates could be measured.

3.6. Conclusion

Identifying top-ranked synergistic triplets in large datasets is feasible using stochastic search algorithms. A brute-force ranking of the 1426 variables retained in the UKB NMR data requires scoring roughly 4.8×10^8 triplets. A complete set of 100 SA runs of 1,000 samples each used 10^5 evaluations and took 30 to 57 minutes on a single laptop. Both SA and PSO recovered rare, highly synergistic triplets, and in our runs PSO left fewer particles trapped in local minima than SA. The two were not compared on a common summary statistic. The clique heuristic is faster still, averaging 240 seconds per search for 10,000 triangles once the pairwise metrics are computed, but it gives an initial assessment rather than a trustworthy ranking. A search-algorithm follow-up remains necessary. Two cautions apply in practice. Echoes must be filtered out before a synergy network is interpreted, and variables must be retained until synergy has been assessed. What this filtering step implies for higher-order discovery workflows is taken up in the General Discussion.

In conclusion, the stochastic search strategies and echo-filtering heuristic presented here provide a practical toolkit for identifying higher-order dependencies in large multivariate datasets. Retaining all variables until synergistic interactions have been assessed guards against the premature elimination of variables that participate in distinct higher-order structures.



Chapter 4

DECOMPOSING THE INFORMATIONAL ROLES OF SYMPTOMS, SIGNS, AND DISEASES IN FUNCTIONAL LIMITATION

Based on:
Peeters, G., Hourican, C., Lees, M., Quax, R., Melis, R., Kok, A., van Schoor, N.M., & Olde Rikkert, M. (2026). Symptoms provide predictive information on daily functioning beyond signs and diseases in middle-aged and older adults, particularly those with multimorbidity. Age and Ageing, 55(2), afag046. *https://doi.org/10.1093/ageing/afag046*

The preceding chapter demonstrated two things. First, stochastic search algorithms can identify high-synergy variable sets without exhaustive enumeration. Second, there is an echo problem whereby strong pairwise associations inflate the apparent number of synergistic triplets. Chapter 2 established that synergistic associations exist in health data. With the computational tools now in place, the present chapter asks whether they matter for a concrete outcome, daily functioning, and which clinical domain carries the information. The data come from the Longitudinal Aging Study Amsterdam, and the domains are symptoms, signs, and diseases. In persons with multimorbidity, disease labels may carry little unique or additional information once their shared content is accounted for. Chapter 5 takes up the same observation at the level of individual symptoms and signs, and together these two perspectives motivate a shift toward symptom-level assessment. We examine whether its value is greatest precisely where disease labels are most redundant.

4.1. Introduction

The prevalence of multimorbidity (defined in Chapter 1) increases with age from approximately 10–30% at the age of 40 to 50–85% at the age of 75 [49, 123], bringing declines in functioning and quality of life, and rising care costs [124–126]. As detailed in Chapters 1 and 2, the single-disease approach is inadequate for persons with multimorbidity. Under that approach clinical guidelines treat each condition independently and clinicians monitor signs per diagnosis [15], and parallel treatment can lead to high treatment burden, overtreatment, and fragmented care [17, 127]. A more holistic approach is needed, one that considers symptoms, signs, and diseases—the three domains defined in Chapter 2—as interacting components of a single clinical picture rather than independent entities. Focusing on symptoms may help to adopt such an approach. Such an approach aligns better with the patient's experience.

In the context of multimorbidity, a single symptom may be caused by multiple diseases or by interactions between these diseases. For example, thyroid disorder is a common comorbidity in heart failure, and both diseases, independently and jointly, can cause fatigue [128]. Symptoms may also trigger new symptoms independent of underlying diseases, resulting in atypical disease presentations. For example, in frail older adults, COVID-19 often presents with a fall as a consequence of reduced attention [129]. Atypical disease presentations are more common in (frail) older adults [130] and challenge the common conception that (changes in) symptoms are mere presentations of new or exacerbating diseases. This suggests that symptoms may behave partly independently of disease activity, carrying information that diagnostic disease labels capture insufficiently. The observation that symptoms trigger other symptoms illustrates the need to look at symptoms collectively rather than individually, and to account for potential interactions between symptoms [48].

A symptom-oriented approach, complementary to the disease-oriented approach, rests on an empirical premise. The hypothesis tested here is that symptoms and signs provide predictive information on functional limitations beyond that of diseases in people with multimorbidity. To test it, we decompose the information that symptoms, signs, and diseases carry about functional limitations using Partial Information Decomposition. That framework, introduced formally in Chapter 1, distinguishes unique, redundant, and synergistic contributions. Random Forest prediction models provide a non-parametric validation.

The analysis uses four aggregate variables: symptom count, sign count, disease count, and functional limitations. A three-source decomposition is tractable and interpretable at this level of granularity, and domain-level aggregation tests directly whether entire clinical domains carry distinct informational roles relative to a patient-centred outcome. Chapters 5 and 6 subsequently disaggregate these domains to the level of individual symptoms and signs. This hypothesis is expected to hold specifically under multimorbidity. The chapter therefore also examines the general community-dwelling middle-aged and older population, to check whether the pattern is condition-dependent. The pooled

This chapter is based on G. Peeters, C. Hourican, M. Lees, R. Quax, R. Melis, A. Kok, N. M. van Schoor, and M. Olde Rikkert. Symptoms provide predictive information on daily functioning beyond signs and diseases in middle-aged and older adults, particularly those with multimorbidity. *Age and Ageing*, 55(2):afag046, 3 2026. ISSN 1468-2834. doi: 10.1093/ageing/afag046. G. Peeters and C. Hourican contributed equally to this work.

results support the hypothesis: in the multimorbidity subgroup, symptoms contribute unique information (34.1% of the total) that neither signs nor diseases do, and adding them to a model of signs and diseases raises agreement from κ=0.287 to κ=0.459.

4.2. Methods

4.2.1. Study design and participants

Data were used from participants in the Longitudinal Aging Study Amsterdam (LASA), an ongoing population-based cohort study [26, 131]. The original cohort of 3,107 community-dwelling older adults aged 55–85 years was recruited in 1992/93 with triennial follow-up. The sample was replenished with participants aged 55–65 years in 2002/03 (n = 1,002) and 2012/13 (n = 1,023). Data were collected via face-to-face structured interviews and self-completed questionnaires covering physical, cognitive, emotional, and social domains. The study was approved by the Medical Ethics Committee of the VU University Medical Center Amsterdam (file numbers: 92/138, 2002/141, 2012/361, 2016.301). All participants signed informed consent.

For the current analyses, we used data from the 1995/96 through 2021/22 waves. Participants with more than 50% missing values across all relevant variables in a certain wave were excluded from the analyses for that wave. Missing values were imputed using iterative random forest procedures (Multiple Imputation by Chained Equations, 40 iterations, maximum depth 5, minimum samples per leaf 10, 100 estimators) [119, 132]. After imputation, the analytic dataset comprised 14,717 observations from 4,025 participants across all waves. Chapter 5 applies the same exclusion rule to a partly different set of LASA variables and reports 14,716 observations from the same 4,025 participants.

4.2.2. Symptoms, signs, and diseases

In this chapter, 'diseases' refers to the diagnostic labels of abnormal conditions that affect the structure or function of part or all of the body and usually are associated with specific signs and symptoms [133]; 'signs' refers to objectively measured manifestations of diseases; and 'symptoms' refers to subjectively experienced manifestations of diseases. Selection of diseases, symptoms, and signs was based on (i) availability of data for the whole sample (ruling out symptoms measured in subgroups with a certain disease diagnosis); and (ii) consistent measurement across waves to be able to include as many waves as possible. Six symptoms and five signs were measured and dichotomised using standardised questionnaires and cut-points as described in Appendix C (Expanded methods). Symptoms included depressive symptoms, anxiety symptoms, sleep problems, pain, hearing complaints, and memory complaints. Signs included mean arterial pressure, grip strength, BMI, waist-hip ratio, and cognitive functioning. Eight chronic diseases were measured as self-reported diagnosed diseases, namely chronic non-specific lung disease, cardiac disease, peripheral arterial disease, diabetes mellitus, stroke, osteoarthritis, rheumatoid arthritis, and cancer. The numbers of present symptoms (0–6), signs (0–5),

and diseases (0–8) were summed.

4.2.3. Functional limitations

Functional limitations were assessed using six self-reported items on limitations in daily activities, namely stair use, dressing, cutting toenails, walking outside, sitting/rising from a chair, and transportation. These items were derived from the 6-item Capacity for Self-Care indicator and the 9-item Organisation for Economic Co-operation and Development (OECD) indicator, with item selection based on the items with the highest item-rest correlation [134]. Each item was rated on a five-point ordinal scale from 0 ("Yes, without difficulty") to 4 ("No, I cannot"). Responses were summed to create a total functional limitations score (range 0–24, with higher scores indicating more limitations).

4.2.4. Statistical analyses

To quantify how symptoms, signs and diseases individually and jointly relate to functional limitations, we required a multivariate correlation measure that could capture non-linear relationships among categorical data, and a method to decompose these correlations. Therefore, we used mutual information, where the decomposition was done by Partial Information Decomposition (PID) [22] using the Minimum Mutual Information (MMI) redundancy measure [135]. Mutual information quantifies the statistical dependence of the outcome variable (functional limitations) on the explanatory variable (e.g. symptoms). Mutual information is a non-parametric method that deals with non-linear and complex relationships between categorical variables. PID decomposes total mutual information, expressed as 'bits' of information, into its constituent parts, i.e. unique, redundant, synergistic and lost information.

These parts are interpreted as follows. *Unique information* reflects the predictive contribution of a single explanatory variable independently of the others. For example, if a symptom provides zero unique information, this indicates that its predictive value is either entirely shared with other variables or only emerges when combined with them. *Redundant information* represents the shared predictive insights about the outcome variable that multiple explanatory variables offer (e.g. the overlapping information conveyed by both signs and symptoms). *Synergistic information* captures the added value that arises when explanatory variables are considered together. It quantifies how a combination of, say, signs and symptoms can predict functional limitations more accurately than the sum of their individual contributions. *Lost information* is not a separate atom but a derived summary. It is the information forfeited if a domain is excluded, and equals that domain's unique information plus its synergy lower bound. It is therefore not the sum of the unique information and the pure synergy reported for that domain in Table 4.1.

Two properties of the MMI decomposition bear on the interpretation of the results below. First, MMI takes redundancy to be the smallest mutual information that any single source shares with the target, so at most one domain can carry non-zero unique information, whatever the data. A zero must therefore be read as a property of the estimator as

much as of the data. Second, no sensitivity analysis to the choice of redundancy measure was performed in this chapter, and the atoms reported below are point estimates without confidence intervals.

To ensure a balanced distribution in each of the variables while preserving interpretability and ordinal structure, the continuous functional limitations score (0–24) was categorised as 0, 1, 2–5 and 6–24 (based on approximate tertiles for scores > 0). Sign and disease counts were categorised into four discrete groups: 0, 1, 2, and ≥ 3. The symptom count was categorised as 0–1, 2, 3, and ≥ 4.

To check whether our results depended on one methodological approach, and to examine the added predictive value of symptoms to signs and diseases, we additionally ran Random Forest (RF) models with 100 trees to predict functional limitations. Rather than using sum scores, we included each symptom, sign and disease as a separate explanatory variable to make optimal use of the available information. The functional limitations score was included as a discrete variable (range 0–24). The RF models therefore draw on a finer information set than the three domain counts entering the PID. The two analyses are complementary rather than strictly corroborating. Agreement between them shows that the added value of symptoms does not rest on one modelling choice.

Models were run in two series, adding all symptoms, all signs and all diseases step by step. The first started from diseases and added signs, then symptoms. The second started from symptoms and added signs, then diseases. Model performance was quantified as the agreement between observed and predicted functional limitations using quadratic-weighted Cohen's κ, reported as a mean with a 95% confidence interval.

Both the PID and RF models were run on the total sample and separately in the multimorbidity subgroup, which was our primary interest. The total sample served as a reference for interpreting those results. Data from waves 1995/96 to 2021/22 were merged and treated as one "cross-sectional" dataset to increase the number of observations available for the PID and RF models. The models did not account for the dependence between observations from the same person, which may inflate or skew the results. Participants contributed up to nine observations to the pooled dataset, so the same person could appear in both the training and the evaluation part of a Random Forest split. The pooled κ values are best read as an upper bound rather than as an unbiased estimate of out-of-sample agreement. Sensitivity analyses examined the influence of ignoring this dependence by repeating the RF models using wave 1995/96 only and wave 2021/22 only, the first and last waves. Each participant contributed at most one observation per wave. These single-wave analyses therefore serve as the participant-level check.

4.3. Results

Across all waves, the 4,025 participants were on average 73 ± 8 years of age (range 57–102). Of these, 46% were women and 59% had multimorbidity. Participants had a median (IQR) count of 2 (2–3) symptoms, 1 (1–2) signs and 1 (1–2) diseases. The median (continuous) number of functional limitations was 1 (0–4).

In the total sample, total mutual information from symptoms, signs, and diseases

was 0.252 bits, capturing 13.3% of the entropy in functional limitations (Table 4.1). This was almost entirely interaction-driven. None of the three domains contributed unique information, and pure synergy accounted for most of the explained entropy (signs 9.3%, symptoms 34.7%, diseases 34.7%). Direct mutual information with the target was modest (symptoms 7.1%, diseases 6.7%, signs 1.6%), underscoring the dominance of cross-domain interactions. Removing any single domain led to similar information loss for symptoms (35.4%) and diseases (34.7%), with a smaller loss for signs (9.3%). Sensitivity analyses repeating the PID models using waves 1995/96 and 2021/22 data only showed roughly similar patterns (Appendix C, Table A), with information explained by symptoms, signs and diseases again predominantly interaction-driven.

In participants with multimorbidity, total information was lower: 0.169 bits, or 8.5% of entropy, against 0.252 bits and 13.3% in the total sample (Table 4.1). Symptoms now provided 34.1% unique information (0.058 bits, or 2.9% of the entropy in functional limitations), and omitting them forfeited 64.8% of explanatory power (0.109 bits). Diseases and signs provided no unique information, and their removal produced far smaller losses (15.8% and 14.9%, respectively). Pure symptom and disease synergy halved to 15.8%, whereas synergy from signs increased to 14.9%, indicating that signs contributed chiefly through interactions. Direct mutual information likewise declined (symptoms 5.9%, diseases 2.0%, signs 1.1%), showing that functional limitations in participants with multimorbidity are driven more by symptom burden, with diseases and signs adding little beyond symptoms. The null result for diseases should be read with care. Selecting on disease burden restricts the range of the disease count within this subgroup, and no within-subgroup distribution or entropy is reported for the three domains. Informational saturation and range restriction therefore cannot be separated here. Sensitivity analyses using single waves showed mostly similar results: in both waves 1995/96 and 2021/22, pure synergy was higher for all domains than in the pooled analysis, suggesting that a significant proportion of the information is contributed through interactions. However, in wave 2021/22, symptoms did not provide unique information in the multimorbidity subgroup (Appendix C, Table A). One possible explanation is that the 2021/22 subsample is smaller and drawn from an older, more selected cohort in which the symptom–function relationship may have shifted. Another is that the pooled analysis, by combining waves, captures a broader range of disease–symptom configurations that enhances symptom distinctiveness. We note that pooling observations across waves from the same participants inflates the effective sample size (14,717 observations from 4,025 participants). The single-wave analyses therefore serve as the more conservative test, and the non-replication in 2021/22 should temper the strength of the unique-information claim.

As in the PID models, the RF models showed that adding symptoms to diseases produced substantial improvements in predictive performance (Table 4.2). In the total sample, agreement improved minimally after adding signs (κ=0.354) and substantially after adding symptoms (κ=0.511), a 65% improvement over the disease-only model (κ=0.310). In the multimorbidity group, agreement improved by 121% after adding signs (κ=0.287) and symptoms (κ=0.459) to the disease-only model (κ=0.208). Conversely, the gain in

agreement was much smaller when diseases and signs were added to symptom-only models (Table 4.2). Sensitivity analyses repeating the RF models using waves 1995/96 and 2021/22 data only showed similar results, but with wider confidence intervals: agreement improved more when symptoms were added to diseases or signs than when diseases or signs were added to symptoms (Appendix C, Table B).

4.4. Discussion

Taken together, the information-theoretic and predictive analyses present a coherent picture. In the total sample, signs, symptoms and diseases show considerable overlap in the information they provide on functional limitations (i.e. high redundancy), while also contributing synergistic information. Notably, symptoms offer a substantial predictive advantage beyond what diseases alone can explain. Similar patterns emerged in the multimorbidity subgroup, with the key distinction that symptoms contributed more unique information in this group. These findings support the view that, in individuals with multimorbidity, symptoms convey information beyond what is captured by disease diagnoses alone.

Findings from both models indicate that symptoms capture unique predictive information for daily functioning not contained in disease diagnoses, while signs provide minimal additional benefit when added to diseases. This suggests that symptoms reflect distinct aspects of daily functioning not captured by signs or diseases. Total mutual information was higher in the total sample (13.3% of entropy) than in the multimorbidity subgroup (8.5%), yet the relative contribution of symptoms was more pronounced in the multimorbidity subgroup, supporting the idea that symptoms may behave partly independently of disease activity in people with multimorbidity, an observation Chapter 5 examines at the level of individual variables. In people with a single disease, the link between symptoms and diseases is more direct. Under multimorbidity, a single symptom can be caused by multiple diseases, and symptoms can trigger new symptoms. For example, both joint stiffness and poor eyesight can lead to a fall, which can lead to pain. That pain is not directly linked to the eye disease causing the poor eyesight or to the osteoarthritis causing the stiffness. Symptom information therefore complements disease-based assessment under multimorbidity rather than being redundant with it.

Beyond unique information, the finding that symptoms, signs and diseases provide synergistic information implies that their combination tells us more about daily functioning than the sum of the information provided by each variable. This synergistic pattern aligns with deficit-based definitions of frailty [136], in which frailty emerges from the accumulation and *interaction* of multiple deficits rather than from any single deficit.

The current findings may be intuitive to clinicians working in geriatrics and rehabilitation. Direct evidence for the pattern is nonetheless scarce in the literature. Clinical reasoning based on differential diagnosis often focuses on single-disease causality, which may underestimate the role of cross-domain interactions in geriatric care. Our findings align with the few studies that decomposed the shared and unique contributions of symptoms, signs and diseases. Previous analyses in the same cohort showed

that disease-specific symptoms provide more detailed information about the impact of chronic diseases on mobility than diseases alone [137]. A study among 1,073 people living with HIV showed that adding signs and symptoms to multimorbidity patterns changed the strength of associations with patient-reported outcomes [138]. The robustness of these findings requires validation in other cohorts and settings, particularly clinical ones. Chapters 5 and 6 provide partial validation, using different variable sets and a different clinical domain. Cross-cohort replication within the same multimorbidity–functioning framework remains an open step.

Key strengths of this chapter include the large, well-characterised LASA cohort of middle-aged and older adults, its long-term follow-up, and the dual use of information theory and predictive RF models. Because the waves were pooled, that follow-up supplied repeated observations rather than support for longitudinal inference. Attrition across the 1995/96 to 2021/22 window was not examined. Loss through death, institutionalisation and refusal is unlikely to be independent of symptom burden or of functioning, so the later waves represent a surviving and healthier subset.

Four limitations qualify these findings. First, in a study as large as LASA it is impossible to measure all potentially relevant symptoms, signs and diseases consistently. The items that were measured were selected for their high prevalence and their relevance to an ageing community-dwelling population. Second, severity of symptoms and signs was not taken into account. This information was available for some but not all items, so all items were treated equally by dichotomising symptoms and signs as present or absent. Incorporating severity would plausibly result in stronger associations with functional limitations. Third, symptoms and diseases were self-reported, and that reliance is unevenly distributed across the three domains. The six symptoms and the functional-limitations outcome were both reported by the participant, whereas the five signs were objectively measured. Shared reporting style may therefore inflate the symptom–function dependence relative to the sign–function dependence, by an amount these data cannot quantify. Repeating the decomposition against an objectively measured functional outcome would bound that concern. Fourth, PID required discretisation, which may mask subtle effects. Whether a continuous-variable decomposition would alter the main conclusion was not tested here.

In conclusion, these findings support the hypothesis that symptoms provide information beyond that of diseases, particularly in people with multimorbidity. Whether a symptom-oriented approach adds value beyond disease-based assessment in routine multimorbidity care is taken up in the General Discussion, alongside the variable-level and network-level findings of Chapters 5 and 6.

Table 4.1: Variability in functional limitations explained by symptoms, signs and diseases

Domain	Lost Information	Unique Information	MI with Functional Limitations	Pure Synergy	Synergy Bounds (%)
Total sample: entropy = 1.889, total MI = 0.252					
Signs	9.31% (0.023 bits)	0	1.60% (0.030 bits)	9.31% (0.023 bits)	9.31–17.57% (0.023–0.044 bits)
Symptoms	35.39% (0.089 bits)	0	7.10% (0.134 bits)	34.67% (0.087 bits)	35.39–42.93% (0.089–0.108 bits)
Diseases	34.67% (0.087 bits)	0	6.70% (0.127 bits)	34.67% (0.087 bits)	34.67–42.93% (0.087–0.108 bits)
Multimorbidity subgroup: entropy = 1.991, total MI = 0.169					
Signs	14.90% (0.025 bits)	0	1.13% (0.022 bits)	14.90% (0.025 bits)	14.90–29.77% (0.025–0.050 bits)
Symptoms	64.76% (0.109 bits)	34.13% (0.058 bits)	5.87% (0.117 bits)	15.77% (0.027 bits)	30.63–30.63% (0.052–0.052 bits)
Diseases	15.77% (0.027 bits)	0	1.98% (0.039 bits)	15.77% (0.027 bits)	15.77–30.63% (0.027–0.052 bits)

Mutual information (MI) quantifies how much of the variability in functional limitations is captured by information on symptoms, signs and diseases. Total mutual information captures the combined predictive power of all three domains. In the total sample this is 0.252 bits, equalling 13.3% of the total variability in functional limitations (entropy = 1.889 bits = 100%). In the multimorbidity subgroup it is 0.169 bits, equalling 8.5% of the variability (entropy = 1.991 bits = 100%). Lost information and unique information are defined in the Methods. A zero in the Unique Information column means that the domain's predictive value about functional limitations is either shared with the other domains or emerges only in combination with them. Pure synergy is the information gained only when that domain is paired with at least one other domain. The synergy lower bound is the guaranteed minimal extra information from combinations involving that domain—the information forfeited if it never interacts with another domain. It is computed as the lost information minus the unique information. The synergy upper bound is the maximal extra signal that could come from combining that domain with one or both of the other domains. It is the ceiling on the interaction-driven information that domain might contribute. Percentages in the MI with Functional Limitations column are relative to the entropy of functional limitations (1.889 bits in the total sample, 1.991 bits in the multimorbidity subgroup). All other percentages are relative to the total mutual information of the three domains (0.252 and 0.169 bits). Higher percentages indicate a larger contribution.

Table 4.2: Additional predictive value of symptoms, diseases and signs beyond that of the other domains.

	Diseases only Mean Cohen's κ (CI)	**Diseases + Signs** Mean Cohen's κ (CI)	**Diseases + Signs + Symptoms** Mean Cohen's κ (CI)
Total sample	0.310 (0.283–0.337)	0.354 (0.325–0.381)	0.511 (0.487–0.536)
Multimorbidity subgroup	0.208 (0.172–0.230)	0.287 (0.259–0.315)	0.459 (0.422–0.489)
	Symptoms only Mean Cohen's κ (CI)	**Symptoms + Signs** Mean Cohen's κ (CI)	**Symptoms + Signs + Diseases** Mean Cohen's κ (CI)
Total sample	0.440 (0.416–0.458)	0.465 (0.438–0.488)	0.511 (0.487–0.536)
Multimorbidity subgroup	0.386 (0.358–0.416)	0.413 (0.385–0.445)	0.459 (0.422–0.489)
	Signs only Mean Cohen's κ (CI)	**Signs + Symptoms** Mean Cohen's κ (CI)	**Signs + Symptoms + Diseases** Mean Cohen's κ (CI)
Total sample	0.084 (0.068–0.101)	0.466 (0.441–0.491)	0.511 (0.487–0.536)
Multimorbidity subgroup	0.106 (0.086–0.128)	0.414 (0.387–0.445)	0.459 (0.422–0.489)

The mean Cohen's κ and its 95% confidence intervals (CI) express the agreement between observed and predicted functional limitations. Higher agreement indicates more accurate prediction.



Chapter 5

FROM TRIPLET CATALOGUES TO INTERPRETABLE ARCHETYPES OF SYNERGISTIC DEPENDENCE

Based on:
Hourican, C., Peeters, G., Melis, R.J.F., Kok, A., van Schoor, N.M., Wezeman, S., Lees, M., Olde Rikkert, M.G.M., & Quax, R. (2026). Interpreting higher-order dependence in multimorbidity using cohort data: A partial information decomposition approach. PLOS Computational Biology, 22(6), e1014386. *https://doi.org/10.1371/journal.pcbi.1014386*

The preceding chapter demonstrated that symptoms carry unique and synergistic information about daily functioning, and that in persons with multimorbidity, disease labels alone become informationally insufficient. That analysis was deliberately constrained: three broad domain scores, a single target, one population split. The goal was to isolate a domain-level question before moving to individual variables. At that resolution, source–target combinations are few. Each result can therefore be inspected on its own terms. The challenge changes, however, once we scale the analysis up. Multimorbidity data contain many individual variables that can be paired and evaluated against many targets. Moreover, higher-order findings may differ across subgroups defined by age, sex, or disease burden. This scaling creates an interpretation bottleneck. PID outputs grow far faster than they can be read one triplet at a time. This chapter addresses that bottleneck by introducing the BUST framework (Breadth–Uniformity–Synergy–Total). This four-dimensional summary characterises each variable pair's synergistic profile across targets and strata. It supports cohort-scale comparison and reveals which feature constellations carry joint information that no single variable provides alone. The chapter retains the journal ordering of the published article, so the Methods follow the Results and Discussion.

5.1. Introduction

Clinical features seldom act alone. In high-dimensional clinical data, multiple measurements can carry information about targets that is not present in any single variable—a property called *synergy* (Chapter 1). Detecting such interactions is difficult for several reasons: the number of candidate feature sets grows combinatorially; multivariate information estimates can be biased in finite samples; and higher-order effects are challenging to interpret because many analytical tools either blur pair-level specificity or produce triplet-level results that are hard to compare across targets and subgroups [42, 140]. An interpretation-first approach is needed if higher-order dependence is to inform clinical reasoning and study design. Chapter 4 showed that symptoms carried unique and synergistic information about daily functioning in persons with multimorbidity. This chapter asks whether that finding scales from three broad domain scores to individual-variable resolution, and how the resulting higher-order landscape can be made clinically legible.

As the preceding chapters argued (Chapters 1–4), conventional multimorbidity approaches characterise co-occurrence and additive burden. Disease counts, composite indices, and pairwise networks offer limited traction on genuinely irreducible joint influences. The key gap is not the absence of a synergy measure. PID and the O-information (Chapters 1 and 2) provided the information-theoretic foundation. Chapter 3 developed the computational infrastructure for scalable detection. What is missing is an interpretation layer for comparing PID outputs across targets and subgroups. Two barriers have limited cohort-level application. First, detection is fragile: multivariate estimators exhibit small-sample bias [42] and naïve pipelines can misclassify redundancy as synergy or vice versa. Second, interpretation is underdeveloped. Triplet-level results are hard to compare across targets or subgroups. Existing summaries remain either global—losing pair-level specificity—or hyper-local, making generalisation difficult. The computational barrier was partly addressed in Chapter 3; this chapter confronts the second barrier head-on.

The information-theoretic notion of synergy adopted here is distinct from the epidemiological concept of interaction (as discussed in Chapters 1 and 2). Two variables exhibit synergy for a target when observing them jointly reduces uncertainty about that target more than would be expected from their individual contributions alone. The property is model-free and need not align with a particular regression interaction term or link function. This distinction matters clinically, because symptom burden often outperforms disease counts for predicting function, and certain constellations behave "worse than the sum of their parts" [141–144]. What remains unclear is how to identify, *a priori*, which combinations of clinical features are likely to interact non-additively—and how to summarise those interactions in a way that supports rather than overwhelms clinical interpretation.

We address these limitations with a bias-aware, interpretation-first workflow. The workflow couples permutation-screened PID estimation with a pair-centric summary of how synergy is organised across targets and populations. To make the results clinically legible, we summarise each pair's behaviour using a Breadth–Uniformity–Synergy–Total (BUST) representation. *Breadth* quantifies how widely a pair's synergy is distributed

This chapter is based on C. Hourican, G. Peeters, R. J. F. Melis, A. Kok, N. M. van Schoor, S. Wezeman, M. Lees, M. G. M. Olde Rikkert, and R. Quax. Interpreting higher-order dependence in multimorbidity using cohort data: A partial information decomposition approach. *PLOS Computational Biology*, 22(6):e1014386, 2026. doi: 10.1371/journal.pcbi.1014386

across targets (broad vs. narrow), whereas *Uniformity* quantifies how consistently these effects appear across age, multimorbidity and sex strata (universal vs. specific). *Synergy* and *Total information* capture the magnitude and context of these effects. The BUST map collapses many triplets into a smaller set of interpretable archetypes—for example, broad-uniform pairs whose synergy is consistent across the population, narrow-uniform pairs with focused but reliable effects, and subgroup-specific pairs whose synergy varies across demographic or multimorbidity strata. This addresses sub-question 4 of this thesis: how PID-based higher-order dependence can be summarised so that it remains clinically legible rather than becoming a catalogue of triplets.

We apply the workflow to data from the Longitudinal Aging Study Amsterdam (LASA; the same cohort studied in Chapter 4, now at individual-variable resolution) [145]. Because multimorbidity in older adults is often experienced as symptom burden and functional limitation rather than as a list of diagnoses, we focus on routinely assessed symptoms and clinical signs, on behaviours and lifestyle factors, and on self-rated health, cognitive and sensory measures. This broadens the variable set beyond the strict symptoms/signs/diseases trichotomy of Chapter 4. Chronic disease information is used only to define multimorbidity strata and is not itself treated as a source or target variable in the PID analysis. Apart from the stratification variables (age, sex and multimorbidity), we treat all health-related variables symmetrically as both sources and targets: each variable can appear either in a source pair or as a target in a source–source–target triplet. We repeat all analyses within age, sex and multimorbidity strata to characterise non-uniformity across subgroups. This design preserves clinical relevance because the measured variables span domains central to independent living and routinely assessed geriatric evaluation, while the pair-centric summaries mirror how clinicians reason about co-present problems [141].

The chapter pursues four aims: (i) quantify the prevalence and magnitude of triplet-level synergy in a large ageing cohort; (ii) identify which variable pairs inform broadly versus narrowly across targets and assess how consistently these patterns hold across demographic and multimorbidity strata; (iii) test whether synergistic information is diffuse across many pairs or concentrated in a few; and (iv) evaluate how synergy-based rankings relate to more traditional dependence measures such as correlation and total mutual information (TMI). The contribution is not a new information-theoretic measure, but a bias-aware PID workflow coupled with the BUST representation that renders higher-order dependence interpretable at the level of individual variable pairs. The results show that synergy is selective rather than ubiquitous, that synergy-ranked pairs are distinct from TMI-ranked pairs, that multidomain cliques involving subjective health, pain, cognition and grip strength concentrate non-additive configurations, and that the multimorbidity-specific reorganisation of information reported in Chapter 4 reappears at the individual-variable level, where a subset of pairs shows synergy profiles that vary chiefly across multimorbidity strata.

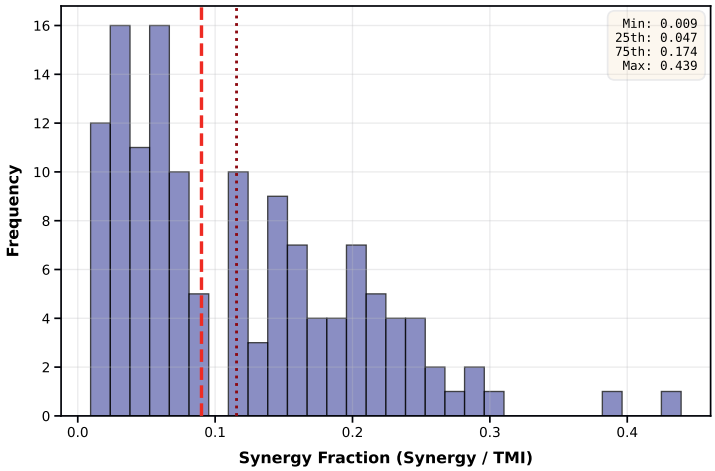


Figure 5.1: **Synergistic information constitutes a small but heterogeneous fraction of total mutual information (TMI).** Distribution of synergy fraction (synergy strength / TMI) across 135 source pairs. The right-skewed distribution indicates that a minority of relationships exhibit disproportionately strong synergistic contributions. The vertical dotted lines show the median (long dashes) and mean (short dashes).

5.2. Results

5.2.1. Prevalence and magnitude of synergistic information

Before decomposing shared information into synergistic, unique and redundant components, we first examined how often pairs of variables jointly carried any information about a third variable. With $V = 17$ analysed variables treated symmetrically as potential targets, this yields $V \times \binom{V-1}{2} = 17 \times 120 = 2{,}040$ distinct variable-pair–target triplets per dataset (i.e., all pairs among the remaining variables for each target). Across all age, sex and multimorbidity strata, 86–97% of these two-predictor–one-target triplets showed statistically significant joint mutual information under permutation testing with false-discovery-rate (FDR) control. In the full sample, 2,005 of 2,040 tested triplets (98.3%) had significant joint mutual information (FDR thresholds 0.042–0.048; Appendix Table D.1). These results indicated that joint dependence between variables was widespread. We then examined how much of this shared information was truly *synergistic*, as opposed to being explained by redundant or unique contributions.

Across the 135 source pairs with non-zero bias-corrected synergy for at least one target (Methods Section 5.5.5), synergy constituted a small but widely varying fraction of TMI (Figure 5.1). The distribution of synergy fractions was right-skewed. The median was 0.090 and the interquartile range 0.047–0.174, so for most pairs fewer than 10% of shared information bits arose from joint effects. Large synergistic contributions are therefore selective rather than ubiquitous, motivating methods that explicitly identify these rare but informative interactions.

To assess the stability of these estimates, we computed bootstrap confidence intervals (1,000 replicates). Across the top-10 BUST pairs, 95% CIs for mean synergy were narrow (median width 0.005 bits) and all excluded zero (Figure D.12). At the individual-target level, 73–93% of targets per pair had synergy CIs that excluded zero, indicating that the synergistic signal is not driven by a small subset of targets. The resampling unit was the person-wave record, not the participant. These intervals are therefore narrower than a bootstrap over the 4,025 participants would produce.

5.2.2. Simulation validation

Three simulation experiments confirmed that the quadratic extrapolation (QE) bias-corrected pipeline recovers known PID atoms and BUST scores under conditions matched to the empirical data (Methods Section 5.5.8; Section D.7). In the cardinality sweep (128 conditions, 50 replicates each), estimated PID atoms matched ground truths with mean absolute bias below 0.01 bits and RMSE below 0.05 bits for $K \leq 4$. Coverage of the 95% confidence intervals ranged from 90% to 98%. At $K = 5$ with $N = 500$ a small positive synergy bias (≈ 0.035 bits) appeared, but it diminished at larger sample sizes. Under LASA-matched asymmetric cardinality profiles, all atoms were recovered with RMSE < 0.02 bits at $N = 2{,}636$. Regimes with zero true synergy produced negligible estimated synergy (bias < 0.001 bits at $N \geq 1{,}000$). BUST scores were also faithfully recovered, with RMSE < 0.13 bits for synergy strength and < 0.008 bits for total information, and the relative ranking of pairs by all three BUST components was preserved. Breadth exhibited a conservative downward bias for broad-synergy pairs, meaning pairs classified as "broad" in the empirical BUST map genuinely possess broad profiles.

5.2.3. Breadth and uniformity of synergy across targets and strata

After estimating synergy for each source pair across all targets and strata, we summarised each pair on the BUST map (Figure 5.2). Breadth (B) measures how widely a pair's synergy is distributed across targets, whereas Uniformity (U) measures how consistently that synergy appears across age, sex, and multimorbidity strata. Both scores are median-centred, so $B = 0$ and $U = 0$ describe a typical pair in this dataset. Pairs with positive B exhibit broader-than-typical synergy profiles (spread across more targets), whereas negative B indicates synergy concentrated on a smaller subset of targets. Similarly, positive U indicates more uniform-than-typical synergy across strata, whereas negative U indicates subpopulation-specificity (stronger variation across strata). The two axes therefore define four archetypes: broad–uniform (BU), narrow–uniform (NU), broad–specific (BS) and narrow–specific (NS).

The highest composite BUST score (min–max normalised to $[0,1]$ across pairs) was that of perceived health relative to peers with physical activity (composite = 1.000). Perceived health relative to peers compares a respondent's own health rating with their estimate of their peers' health. This pair combined strong synergy (0.20 bits) with positive Uniformity ($U = 0.27$) and slightly below-median Breadth ($B = -0.17$). Its synergy is consistent across strata but concentrated on a subset of targets. The composite multiplies

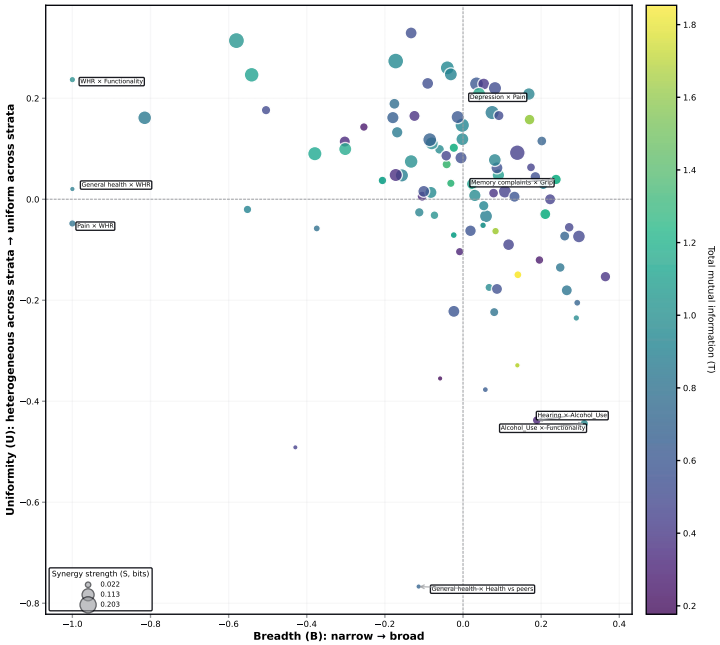
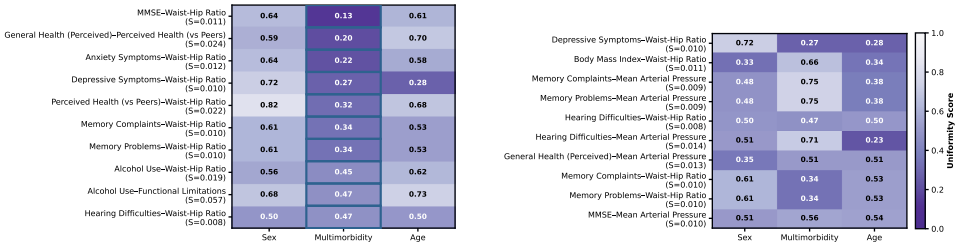


Figure 5.2: **BUST map summarises synergistic source pairs by breadth and uniformity.** Each point represents one source pair ($n = 135$). The x-axis is Breadth (B), measuring how widely a pair's synergy is distributed across targets. The y-axis is Uniformity (U), measuring how consistently that synergy appears across age, sex, and multimorbidity strata (see Methods). Both axes are median-centred, so $B = 0$ and $U = 0$ correspond to a typical pair. Dashed lines at zero define four archetypes: broad–uniform (BU), narrow–uniform (NU), broad–specific (BS), and narrow–specific (NS). Point size encodes synergy strength (bits) and colour encodes TMI (bits). The two most extreme pairs are annotated in each quadrant.

four components, each rescaled to $[0,1]$, so a below-median score on one axis can be offset by high scores on the other three (Methods Section 5.5.7). Next came alcohol use with grip strength (BU, composite = 0.945) and pain symptoms with physical activity (NU, composite = 0.915, synergy strength also 0.20 bits). The remaining seven were sleep quality and depression, each with perceived health relative to peers, sleep quality with physical activity, cognition with perceived general health, perceived general health with grip strength, BMI with alcohol use, and sleep quality with alcohol use. These interactions form an interpretable pattern in which subjective comparative health, sleep, pain, mood, activity, body composition, cognition, and physical capacity jointly encode non-additive information about multiple targets, including functional limitations.

Composite BUST scores provide a compact entry point by ranking source pairs that jointly combine (i) non-trivial synergy strength, (ii) above-typical Breadth across targets, and (iii) above-typical Uniformity across strata (Methods Section 5.5.6). Importantly, BUST does not reduce each pair to a single number: it retains each pair's Breadth and



(a) Pairs with the lowest multimorbidity-specific uniformity component.

(b) Pairs with the lowest median uniformity across sex, multimorbidity, and age.

Figure 5.3: **Low-uniformity pairs reveal which stratification drives context-dependent synergy.** Heatmaps show the entropy-based uniformity components for each source pair across sex, multimorbidity status (present vs. absent), and age group. Lower values indicate greater between-stratum variability in the pair's synergy profile for that stratification, whereas higher values indicate more stable behaviour across strata. Row labels report each source pair and its synergy strength S (bits).

5

stratum-specific Uniformity components, which enables targeted follow-up views once the full set of BUST scores has been computed. We therefore complement the composite ranking with an explicit analysis of low-uniformity pairs, which isolates interactions whose synergy profiles vary most across age group, sex, or multimorbidity status.

Context-dependent synergies and drivers of non-uniformity. Uniformity in the BUST representation is computed separately for each stratification variable (sex, multimorbidity status, and age group) and then aggregated for the 2D map. To make the "specific" (low-uniformity) region of the map interpretable in practice, we ranked source pairs by their *stratum-specific* uniformity components and visualised the lowest-uniformity examples (Figure 5.3). Each row shows a source pair, and each column reports the entropy-based uniformity score for that stratification, where lower values indicate greater between-stratum variation in the pair's synergy profile.

Figure 5.3a focuses on multimorbidity status (present vs. absent) by showing the ten pairs with the lowest multimorbidity uniformity component. For these pairs, the multimorbidity column is systematically lower than the corresponding sex- and age-based uniformity components, indicating that their context dependence is driven primarily by differences between the multimorbidity strata rather than by sex or age.

Figure 5.3b provides a complementary, global view by selecting the ten pairs with the lowest *median* uniformity across sex, multimorbidity, and age. In this view, low overall uniformity arises for different reasons: for some pairs the lowest component is age-based, for others it is sex-based, and for others it is multimorbidity-based. This decomposition shows that "non-uniform" synergy is not a single phenomenon: the same overall label can reflect instability along different subgroup dimensions, which guides which stratified analyses are most relevant for follow-up.

5.2.4. System-level organisation of synergy

We then examined how these pairwise synergies were organised at the system level. Using composite BUST scores as edge weights, we constructed a weighted synergy network covering 135 of the 136 pairs among the 17 variables. Despite this near-complete connectivity, Ward hierarchical clustering of the associated dissimilarity matrix (1 minus the composite BUST weight) revealed a structured organisation of four groups (Figure 5.4a). Two of the four groups were single variables. Waist–hip ratio (WHR) formed a metabolic branch and blood pressure (mean arterial pressure, MAP) a cardiovascular one. A larger subjective-health and functioning cluster grouped anxiety symptoms, perceived health relative to peers, pain symptoms, physical activity, sleep quality and functional limitations. The remaining variables formed a cognition/comorbidity cluster linking alcohol use, BMI, depression, general health, hearing difficulties, cognition, memory complaints, memory problems and grip strength. These groupings are broadly consistent with established geriatric syndrome domains (e.g., malnutrition, cardiovascular ageing, psychosocial functioning, dementia, sarcopenia). The correspondence was assigned post hoc rather than tested against an independent domain labelling. Read with that caveat, synergistic information appears to recover clinically meaningful higher-order structure.

To make the multivariate organisation visually explicit, we plotted a thresholded BUST network in which nodes represent variables and edges represent composite BUST scores (Figure 5.4b). For clarity, we retained only edges above the 75th percentile of the composite distribution (34 of the 135 edges) and removed nodes with no such edges. This representation reveals a compact backbone centred on sleep quality, pain symptoms, perceived health relative to peers, physical activity, depression, grip strength and alcohol use. These variables carry a disproportionate share of the strongest synergistic edges and act as an integrative core linking subjective health, behaviour, mood and physical capacity. Metabolic and cardiovascular markers are present in the dendrogram but contribute few strong BUST edges. They are peripheral to this core rather than part of it.

Stable synergistic constellations

Finally, we identified small groups of variables whose pairwise synergies remain consistently strong across multiple thresholds. Starting from the weighted BUST network, we created four unweighted graphs by retaining edges with composite BUST scores above the 67th, 75th, 80th and 85th percentiles. A clique was defined as stable if it appeared in at least three of these four thresholded graphs, highlighting local constellations whose pairwise synergies remain strong across a range of thresholds.

Eighteen stable maximal cliques were found, comprising sixteen triplets and two quadruplets. One recurrent multidomain constellation linked general health, pain symptoms, grip strength and cognition. This clique exhibited comparatively strong and widespread synergies across multiple targets. Joint-state analyses within this clique revealed several configurations that deviated markedly from additive expectations for targets including sleep, physical activity, hearing, memory complaints and functional limitations, and rule-augmented models frequently improved model fit (Table D.5 in the Chapter 5 appendix). Constellations differed in how many joint configurations improved

fit relative to additive expectations. Some were dominated by a single configuration, such as the alcohol use–grip strength edge. Others involved several, such as general health–pain–grip strength–cognition.

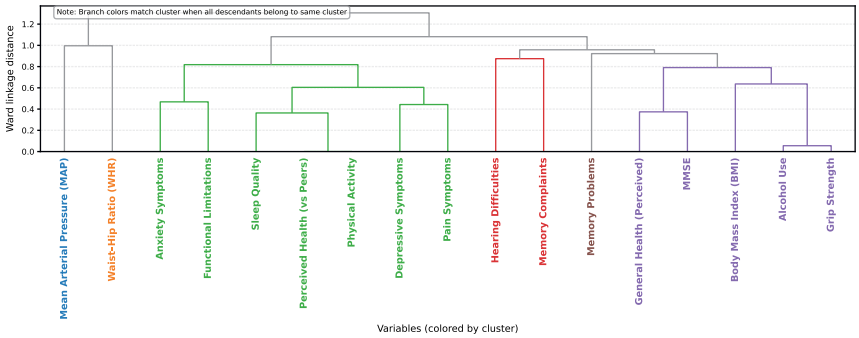
A second stable quadruplet—alcohol use, general health, pain symptoms and grip strength—shared three variables with the first. Its configuration structure was more focused. The alcohol use–grip strength pair was among the strongest BUST edges overall and had one of the highest synergy strengths. For each target examined, a single joint state drove the largest deviations from the additive model. That one configuration was nonetheless enough to improve fit for several targets, including depression, general health, perceived health relative to peers and memory problems (Table D.4 in the Chapter 5 appendix). This behaviour is consistent with a sharp, high-impact interaction nested within an otherwise broader synergy pattern.

Among triplets, the clique comprising perceived health relative to peers, physical activity and sleep quality had the highest mean composite BUST score and was present at every threshold. The sleep–physical activity edge within this set was one of the strongest composite BUST edges in the network and featured prominently in the BUST map. Joint-state analyses indicated that this behavioural clique occupies an intermediate position between the two quadruplets, with multiple abnormal configurations and model-fit improvements concentrated in pain, grip strength and functional targets (Table D.6 in the Chapter 5 appendix).

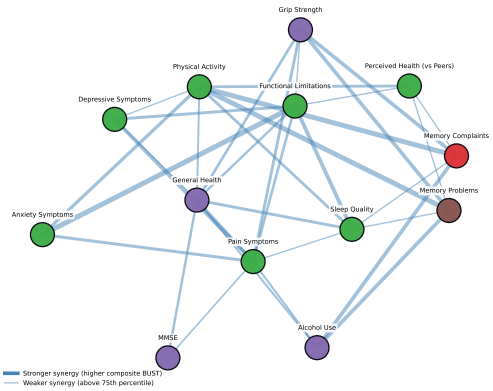
5.2.5. Relationship between synergy and total mutual information

We asked whether the pairs carrying the most synergy were also the pairs carrying the most information in total. TMI is the sum of the bias-corrected PID atoms rather than an external association measure, so this comparison is between two summaries of the same decomposition. If strong synergy simply reflected high overall information sharing, ranking pairs by TMI would recover the same source pairs as ranking by synergy. We therefore constructed two composite scores that summarised each pair's breadth and cross-stratum consistency but differed only in their magnitude component: one used synergy strength, the other TMI (Methods Section 5.5.6).

Synergy strength and TMI were only weakly correlated across the 135 source pairs (Pearson $r = 0.31$, $p < 0.001$). The top-10 pairs under the synergy-based composite showed *zero overlap* with the top-10 pairs under the TMI-based composite (Figure 5.5). Several high-TMI pairs, such as combinations of memory complaints and memory problems, had modest synergy fractions, whereas strongly synergistic pairs—such as alcohol use–grip strength, sleep–physical activity and perceived health relative to peers–physical activity—occupied mid-range TMI values. This confirmed that synergy-based rankings were not recoverable from TMI alone and underscored the added value of PID relative to correlation or TMI screening. Several TMI-composite top pairs were tied, consistent with near-redundant variables yielding similar total-information profiles.



(a) Dendrogram: hierarchical clustering of variables by BUST patterns.



(b) Weighted BUST network (edges above the 75th percentile of the composite distribution; node colours follow dendrogram clusters).

Figure 5.4: **Hierarchical and network views of synergistic structure under the BUST representation. (Top)** The dendrogram reveals four dominant clusters, corresponding to metabolic markers, cardiovascular measures, subjective health and functional status, and cognition/comorbidity variables. Coloured branches indicate clusters identified below a threshold of 0.7× the maximum linkage distance, while grey branches represent inter-cluster connections. Despite the high density of pairwise synergistic associations, a clear hierarchical organisation emerges, with metabolic and cardiovascular variables acting as bridges between functional domains. Physical activity and perceived health relative to peers merge at a Ward linkage close to zero, because the composite BUST weight of that pair is among the highest in the network. **(Bottom)** The same composite weights are visualised as a network, with nodes coloured by their hierarchical cluster assignment. For visual clarity, only edges above the 75th percentile of the composite weight distribution are shown, and isolated nodes are omitted. Edge thickness is scaled nonlinearly with the composite weight to emphasise the strongest synergistic relationships. Absence of an edge or node does not imply independence. It indicates that the pair's shared information is dominated by unique or redundant components rather than synergy, or that the pair falls below the plotting threshold. The resulting topology highlights a coherent core involving sleep quality, pain symptoms, perceived health, physical activity, depression, grip strength, and alcohol use.

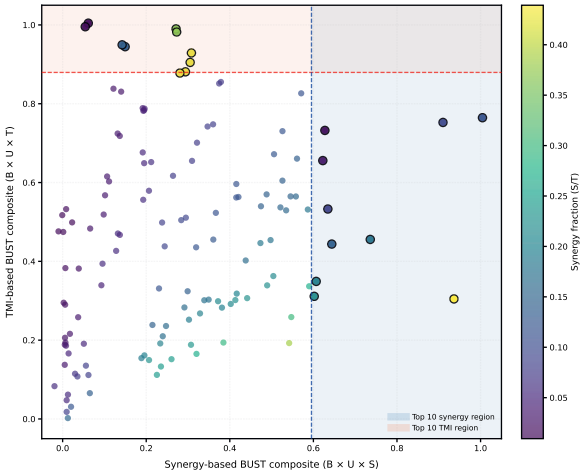


Figure 5.5: **Synergy- and TMI-weighted composites rank source pairs differently.** Each point is one source pair ($n = 135$). The x-axis shows the BUS composite, which uses Breadth (B), Uniformity (U) and synergy strength as the magnitude term. The y-axis shows the BUT composite, which keeps the same B and U terms and substitutes TMI. Scores are rescaled to $[0,1]$ for comparability. Dashed lines mark the tenth-largest score on each axis. No pair ranks in the top 10 under both composites, so the shaded overlap region is empty. Colour encodes the synergy fraction (synergy/TMI). Composites weighted by synergy therefore prioritise pairs that are not dominant in total shared information.

5.3. Discussion

In this chapter, we introduced an interpretation-first pipeline that couples bias-corrected partial information decomposition with pair-centred summarisation via a Breadth–Uniformity–Synergy–Total (BUST) representation. Applied to multimorbidity-related data from LASA, the pipeline mapped how synergistic information is distributed across symptoms, signs and other health-related variables, and across a broad set of health and functional targets. Five findings stand out. First, synergy accounts for a small but varying fraction of total shared information, with a median synergy fraction of 0.090. Second, the pairs ranked highest by synergy are entirely distinct from those ranked highest by total mutual information (TMI). Third, synergistic pairs occupy distinct regions of the BUST map, separating broad and uniform synergies from narrow or context-dependent (specific) ones. Fourth, robust multi-way synergies concentrate in small, cross-domain cliques. Fifth, the synergy network is dense yet hierarchically organised, with some variables acting as integration hubs and others as structural bridges. The BUST representation aligns with symptom-oriented views of multimorbidity, in which functional status emerges from interacting problems across domains rather than from diagnoses considered in isolation. What it adds is a measure of which combinations carry information that neither member supplies alone, computed for every pair and every target rather than for a few pre-specified constellations. The General Discussion develops the

implications for multimorbidity classification and clinical assessment (Chapter 7).

5.3.1. PID reveals information architecture beyond mutual information

The complete non-overlap between the top-10 synergy-based and top-10 TMI-based composite rankings demonstrates that bivariate PID provides non-redundant information relative to conventional mutual information analysis. High TMI pairs were not simply high-synergy pairs with additional redundant content; instead, they belonged to a distinct set of relationships whose shared information was dominated by redundant and unique components. Conversely, strongly synergistic pairs occupied mid-range TMI values and were not identifiable by total information content alone.

5.3.2. BUST archetypes recover expected constellations and highlight new candidates

The BUST map places individual source pairs into archetypes and links synergistic patterns to clinical narratives (Figure 5.2). Several of the strongest and most clinically interpretable pairs—those with plausible joint-state narratives in routinely assessed domains—fall in the broad–uniform (BU) and narrow–uniform (NU) quadrants. Their synergy is expressed consistently across age, sex and multimorbidity strata. Examples include sleep with physical activity, sleep with perceived health relative to peers, alcohol use with grip strength, and depression with perceived health relative to peers. These pairs combine positive Uniformity with either broad or narrower Breadth, and jointly encode information about many functional outcomes. Their positions on the BUST map are consistent with evidence that poor sleep, low physical activity, depression, problematic alcohol use and perceived poor health cluster within individuals and are collectively associated with declines in physical and mental health. For instance, depression and excess weight reinforce each other bidirectionally in longitudinal cohorts [146, 147], and such mood–weight constellations often coexist with disturbed sleep and low activity in older adults. Perceived health relative to peers is particularly notable. Self-rated health predicts retirement due to disability and mortality beyond objective diagnoses [148, 149]. Its repeated appearance in high-scoring BU and NU pairs is consistent with a cross-domain integrative role that the General Discussion examines as part of an informational-roles taxonomy (Chapter 7).

Pairs in the broad–specific (BS) quadrant show synergy that spans multiple targets but varies across strata. These include sleep with alcohol use, pain with alcohol use and sleep with hearing difficulties. Two of the three involve alcohol use. Their placement is consistent with evidence that alcohol-related effects on metabolic and behavioural functioning differ substantially across intake levels and population subgroups. Heavy drinking is more strongly linked with adiposity and cardiometabolic risk, whereas light or moderate consumption shows weaker or mixed associations [150, 151]. Similar variation is reflected in population studies reporting associations between alcohol intake, abdominal obesity and health-related behaviours [152]. The BS quadrant therefore captures pairs whose joint behaviour is broad but modulated by demographic or multimorbidity

context.

Pairs in the narrow–specific (NS) quadrant exemplify concentrated but population-dependent synergies. These include cognition with perceived health relative to peers and several combinations involving BMI, waist–hip ratio (WHR), or blood pressure. Here synergy is consistent with subgroup-conditional coupling between subjective appraisal and objective markers. Combinations of self-rated health with cognitive or physical capacity may carry joint information that is not captured by additive contributions alone, and the strength of this coupling can vary across strata. Similarly, configurations combining functional limitations with poorer perceived health reflect closely related constructs, but the presence of synergy suggests that their joint states can still encode non-additive information about other targets, in a way that differs across subgroups.

Low-uniformity pairs add a clinically relevant layer to this picture. A pair can behave synergistically in the pooled sample and still behave differently between subgroups. A single population-wide summary may then be inappropriate for screening, risk stratification or hypothesis generation. Decomposing Uniformity into its sex-, age- and multimorbidity-specific components helps distinguish interactions that are broadly stable from those whose joint implications differ across age groups or between participants with and without multimorbidity. One reading is that multimorbidity status changes the context in which symptoms, behaviours and physiological markers co-occur, so that the same two-variable constellation can then encode different joint information depending on baseline health complexity. Reporting the aggregate BUST position is therefore not enough. Analyses should also name which stratification dimension drives any observed context dependence; the General Discussion situates this point within a broader argument for systematic higher-order reporting standards (Chapter 7).

The heterogeneity of BUST archetypes among the strongest pairs also invites interpretation. Several high-ranking narrow–uniform pairs, such as perceived health relative to peers–physical activity, pain–physical activity and depression–perceived health relative to peers, combine relatively focused Breadth with consistent cross-stratum behaviour. These cases can often be understood through clinical mechanisms. For example, the synergy between pain and physical activity may reflect different coping patterns: individuals who remain active despite pain tend to exhibit more adaptive coping and better functional trajectories, whereas those who reduce activity may follow a more passive pattern associated with poorer outcomes. BUST does not distinguish these causal pathways, but it flags that the joint state of pain and activity is more informative about downstream targets than either measure alone. Likewise, narrow–specific constellations such as cognition with perceived general health, or cognition with perceived health relative to peers, may arise because cognitive performance aligns more directly with an individual's own perceived health than with more cognitively demanding comparative assessments. These examples illustrate how BUST-derived synergy patterns can map onto clinically meaningful behavioural and cognitive processes, helping to generate mechanistic hypotheses about why certain constellations exhibit strong or population-dependent synergy.

Overall, these quadrant-level patterns serve two purposes. First, they provide face

validity: BUST recovers clinically familiar constellations—links among sleep, perceived health, mood, physical activity and alcohol use—in positions that match established knowledge. Second, the map highlights pairs whose interactions are less well characterised in routine multimorbidity analyses, including pain–alcohol use, anxiety–physical activity and sleep–alcohol use. These combinations are not typically examined together despite plausible mechanistic connections. Their location on the BUST map suggests synergistic behaviour that is strong yet not readily apparent from correlation or co-occurrence alone. Such pairs are candidates for mechanistic or interventional study. Some pairs act as population-wide constellations, others as domain-focused interactions. That contrast carries design implications for synergy-informed clinical assessment that the General Discussion develops into an adaptive assessment framework (Chapter 7).

5.3.3. Synergistic clusters mirror established geriatric domains

Our synergy-based clustering also aligned closely with established geriatric domains. The dendrogram separated a metabolic marker (WHR) and a cardiovascular marker (blood pressure) from a cluster of symptoms and behaviours (sleep quality, physical activity, pain symptoms, anxiety symptoms, perceived health relative to peers, and functional limitations), and from a broader cognition/comorbidity cluster (depression, perceived general health, hearing difficulties, cognition, memory complaints and problems, BMI, alcohol use and grip strength) (Figure 5.4a). This structure is consistent with epidemiological work showing that central obesity and elevated blood pressure are core components of cardiometabolic multimorbidity that predict cardiovascular events and later cognitive decline, bridging metabolic, cardiovascular and neurological pathways [153–156].

The close grouping of sleep disturbance, pain, mood and reduced activity gives variable-level resolution to the synergy–frailty connection established at the domain level in Chapter 4. Co-occurring insomnia, chronic pain and depression are strongly associated with frailty and reduced physical function [157], and the BUST dendrogram now shows which variable pairings underlie the pattern.

These clusters arise purely from the organisation of synergistic information and therefore complement, rather than replicate, additive or correlation-based structures. The cognition/comorbidity cluster, which links depression, hearing loss, subjective memory concerns, poorer perceived general health, and weaker grip strength, also mirrors current understanding of age-related cognitive vulnerability. Hearing impairment and depression have both been associated with increased dementia risk [158], and subjective cognitive complaints are common among patients attending audiology clinics, where multiple dementia risk factors can co-occur [159]. Grip strength and BMI further anchor this cluster in frailty: lower muscle strength and adverse body-composition profiles predict poorer physical and cognitive outcomes in later life [160–162]. Interpreted clinically, the high-weight backbone of synergistic relationships links subjective health, mood, behaviour and physical capacity—domains often jointly implicated in frailty and functional vulnerability—whereas cardiometabolic markers appear to contribute more selectively within this dataset. The emergence of these clinically recognisable constellations from

purely synergistic information suggests that the BUST representation captures meaningful higher-order structure in the multimorbidity landscape.

5.3.4. Clinical implications and future directions

From a clinical perspective, our findings suggest that synergy-aware analyses can complement existing multimorbidity tools rather than replace them. Uniform pairs behaved as integrative signals: their joint states carried information consistently across demographic and multimorbidity strata. Some sat in the broad–uniform quadrant, with synergy spread across many targets, such as sleep–physical activity and alcohol use–grip strength (composite = 0.945). Others were narrow–uniform, concentrating synergy on a subset of targets, including perceived health relative to peers–physical activity (B = -0.17) and depression–perceived health relative to peers. Both kinds are candidate probes for routine assessment, a use the General Discussion develops into a concrete adaptive assessment pathway (Chapter 7). Context-dependent pairs such as sleep–alcohol use instead mark constellations that may warrant closer scrutiny in particular subgroups.

Synergy can be read as a value-of-information signal—how much predictive content a second measurement adds once the first is known. Three translation pathways follow. First, *assessment prioritisation*: in time-constrained settings, BUST-ranked pairs such as perceived health relative to peers–physical activity could guide which cross-domain questions to ask together. Second, *context-sensitive follow-up*: a high-synergy joint state—for example, a patient reporting both poor sleep and low physical activity—could prompt targeted screening for outcomes that neither symptom predicts well in isolation, such as pain, grip strength and functional limitations. Third, *subgroup-aware interpretation*: low-uniformity pairs such as depression–WHR, which showed context-dependent behaviour across multimorbidity strata, may warrant being weighted differently in patients with established multimorbidity than in otherwise healthy individuals. These pathways remain illustrative. The General Discussion (Chapter 7) develops them into a fuller value-of-information framework for synergy-guided assessment.

Because the analyses in this chapter are observational and based on a single cohort, the synergies identified here are best treated as hypothesis-generating rather than ready-made decision rules. Applying similar pipelines to other longitudinal studies is a natural next step; Chapter 6 provides a partial precedent by replicating synergy rankings across two independent cohorts with high concordance (ρ = 0.78). The broader potential for synergy-guided maps to inform combination-intervention design and prognostic modelling is taken up in the General Discussion (Chapter 7).

5.3.5. Study limitations

We pooled observations across LASA waves without explicitly modelling within-subject dependence. Waves are spaced approximately three years apart, and many of the variables we analyse (symptoms, signs, wellbeing indicators and functional limitations) change on much shorter time-scales, so repeated measurements are not near-duplicate records and can reflect genuinely different health constellations. Our focus is population-

level dependency structure rather than individual trajectories, and we therefore treat each record as an independent sample from a relatively stable underlying mechanism. Nonetheless, repeated participation introduces within-person correlation that violates the exchangeability assumption underlying the permutation test. Target-label permutations that swap labels both across and within individuals generate a null distribution whose width may not reflect the true sampling variability. Because the FDR correction inherits any inflation in the underlying p-values, the 98.3% pass rate for joint mutual information should be interpreted as an upper bound. A single-wave analysis or cluster-aware permutation scheme would provide a more conservative screening gate, but was not performed in the present study; the qualitative patterns reported here—the divergence of synergy- and TMI-based rankings, the composition of stable cliques, and the BUST archetype structure—rest on relative rankings rather than absolute significance thresholds, and are unlikely to be driven by pseudo-replication. The pair-selection step is the exception. The 135 source pairs plotted on the BUST map are those with statistically significant synergy for at least one target, and both Breadth and Uniformity are median-centred over exactly that set. Map membership and axis origins therefore inherit the screening gate.

Additional limitations relate to study design and measurement structure. The three-year interval between waves constrains temporal resolution, and the available variables may not capture all clinically relevant domains that contribute to synergistic behaviour. Moreover, our analyses rely on discretised versions of predominantly continuous variables, which can inflate or deflate higher-order structure depending on the choice of cutpoints.

Continuous PID formulations exist that could, in principle, avoid discretisation altogether. Gaussian PID [135, 163] provides closed-form decompositions when all variables are jointly Gaussian, and non-parametric mutual-information estimators based on k-nearest-neighbour statistics [39] could serve as building blocks for continuous PID in more general settings. However, our dataset contains an inherent mix of variable types: some predictors are binary (memory complaints, anxiety) or ordinal with few levels (perceived health, depression), while others are originally continuous (grip strength, BMI, blood pressure). Gaussian PID requires all variables to be continuous and jointly Gaussian, which binary and coarsely ordinal measures violate. Mixed discrete–continuous PID remains an open problem in information theory. Formulations defined for sources and targets of arbitrary type have been proposed, but none has yet been validated at the cardinality profile and per-stratum sample sizes of our data. Our discretisation strategy, while it inevitably discards some fine-grained variation in continuous variables, offers three practical advantages: it handles mixed variable types within a single estimation framework; it produces clinically interpretable categories anchored to established thresholds, and it is fully compatible with the well-characterised $I_{\min}$ redundancy lattice used throughout our analyses. Moreover, the post-hoc sparse-state merging procedure (Section 5.5.5) ensures adequate cell counts for reliable estimation. A binning sensitivity analysis (Appendix D.8) confirms that the main qualitative findings—including the iden-

tity of high-ranking BUST edges and the separation between synergy- and TMI-based rankings—are robust to alternative discretisation choices.

Generalisability is also limited. LASA is a community-based cohort of older Dutch adults, and the synergy patterns observed here may differ in younger populations, in clinical or hospital settings, or in cohorts with different disease burdens or measurement practices. The BUST representation and PID pipeline are directly transferable to such contexts, but the specific locations and magnitudes of synergies should be interpreted as cohort-specific until replicated elsewhere. At the same time, LASA is a population-based cohort sampled from municipal registers across three Dutch regions (Amsterdam, Zwolle and Oss), which supports generalisability of the main patterns to community-dwelling older adults in the Netherlands, even though replication in other cohorts and clinical settings remains necessary.

A further limitation is the restriction to bivariate PID, which the framework itself does not impose. PID generalises naturally beyond two sources. The number of partial information atoms is governed by the Dedekind numbers and grows super-exponentially with the number of sources: from 4 atoms for two sources, to 18 for three, 166 for four, and 7,579 for five [22, 164]. Enumerating and estimating these atoms remains computationally feasible for up to roughly five sources. The binding constraint is the joint state space required for reliable estimation. With discrete variables whose levels range from 2 to 7, the number of cells in the joint contingency table grows exponentially. Five three-level sources and one three-level target produce $3^6 = 729$ joint states, which yields fewer than four observations per cell in an age stratum ($N \approx 2{,}600$) and about twenty per cell even in the pooled sample ($N = 14{,}716$). Both are far below the minimum requirement for stable information-theoretic estimation. Even with four sources the situation is marginal ($3^5 = 243$ cells, ~11 observations per cell). Our focus on bivariate PID (two sources, one target) was therefore a design choice, balancing the detection of irreducibly joint effects against the power available at the sample sizes and variable cardinalities of the LASA cohort. Additionally, a bivariate atom describes what one pair jointly conveys about one target, and it need not coincide with the corresponding component of a decomposition of the full seventeen-variable system. BUST scores are therefore best read as a relative ranking and a screening device rather than as each pair's share of system-level synergy. Future applications to substantially larger cohorts or to variables with fewer discrete levels could extend the analysis to three- or four-source PID and thereby capture higher-order synergies that bivariate decomposition cannot resolve.

Despite these limitations, bootstrap resampling (1,000 replicates) indicated that the synergy estimates of the top-ranked pairs are stable, with narrow 95% confidence intervals that exclude zero for all top-10 BUST pairs (Appendix D.9). Two qualifications apply. The bootstrap resamples person-waves, not participants, so it inherits the within-person dependence described above and gives narrower intervals than a participant-level cluster bootstrap would. The ten pairs were also selected on the same data used to form the intervals, so their uniform exclusion of zero shows the stability of selected extremes rather than unbiased significance. Strategies for addressing within-subject dependence

and temporal resolution—including cluster-aware resampling and designs with finer granularity—are discussed in the General Discussion (Chapter 7) as part of a broader research programme for dynamic higher-order inference.

5.4. Summary of findings

This chapter introduces a bias-aware, interpretation-focused workflow for detecting and summarising synergistic relationships in cohort data. By combining permutation-screened bivariate PID with the Breadth–Uniformity–Synergy–Total (BUST) representation, network visualisation and clique analysis, we map higher-order dependence in a form that remains interpretable at the level of individual variable pairs. The pipeline is implemented in open, reusable code. It requires nothing beyond a standard epidemiological design, and therefore transfers directly to other multimorbidity and ageing studies.

Applied to LASA, the BUST representation reveals that synergistic information is present but concentrated: most pairs show small synergy fractions, whereas a minority exhibit clearly "more than the sum of the parts" behaviour. Synergistic pairs occupy distinct regions of the BUST map, some broad and relatively uniform across strata, others narrow or population-specific across age, sex and multimorbidity strata. The ten pairs ranked highest by synergy and the ten ranked highest by TMI have no member in common, and synergy strength correlates only weakly with TMI (Pearson $r = 0.31$). Synergy therefore identifies information architectures that TMI-based screening cannot recover alone.

Variables such as sleep, BMI, depression, grip strength and perceived health relative to peers participate disproportionately in high-scoring generalist synergies and act as integrative hubs, while combinations involving alcohol use and physical capacity form more focused but robust specialist synergies. These patterns suggest that a small set of routinely collected measures, including brief questions on comparative health and mood alongside simple functional tests, can yield joint information about daily functioning and related targets that is not available from any single variable alone.

Although our analyses are observational and confined to one cohort, the workflow provides a practical template for future work. Synergy-guided variable maps could be used to refine screening batteries, to prioritise combinations of risk factors for mechanistic study, and to identify candidate constellations for multi-target interventions—for example, jointly addressing sleep and activity, mood and weight, or alcohol use and physical performance. Chapter 6 asks whether decompositions of this kind survive the move to symptom networks and to independent cohorts. The broader implications of these findings—for multimorbidity classification, symptom-network methodology, and clinical assessment—are synthesised in the General Discussion (Chapter 7).

5.5. Methods

5.5.1. Overview of the workflow

We developed an interpretation-first pipeline to detect, debias, and summarise synergistic information at scale. The pipeline proceeds in three steps. First, for each target we screened source–source–target triplets using permutation tests with false-discovery-rate control. Second, we estimated bivariate PID atoms with small-sample bias correction and redundancy-definition sensitivity. Third, we compressed triplet-level results to pair-level profiles via a Breadth–Uniformity–Synergy–Total (BUST) map, which quantifies how widely a pair's synergy extends across targets, how consistent it is across strata, and how strong and information-rich it is overall. Data curation, discretisation, and stratification follow clinical guidance and cohort documentation.

5.5.2. Information-theoretic framework

The following definitions recapitulate the notation established in Chapter 1. Mutual information (MI) quantifies the reduction in uncertainty about a target after observing a source. For discrete variables,

$$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)} \;=\; H(Y) - H(Y \mid X),$$

where $H(\cdot)$ denotes Shannon entropy (log base 2), measured in bits. For two predictors $\{X_1, X_2\}$, MI does not separate overlap from complementarity. Partial information decomposition (PID) partitions the joint information about Y into four non-negative components [22]:

$$I(\{X_1, X_2\}; Y) = I_{\text{red}} + I_{\text{uniq}}(X_1) + I_{\text{uniq}}(X_2) + I_{\text{syn}},$$

interpreted respectively as information shared by both sources (redundancy), information available only from X_1 or only from X_2 (unique), and information that emerges only when both sources are observed together (synergy).

Defining the PID atoms. The decomposition requires a definition of redundancy. All other atoms then follow by subtraction. We adopted the $I_{\min}$ measure of Williams and Beer [22], which defines redundancy as the outcome-wise minimum of the specific information each source provides:

$$I_{\text{red}}(X_1, X_2; Y) = \sum_{y} p(y)\, \min\big\{ I_{\text{spec}}(y; X_1),\; I_{\text{spec}}(y; X_2) \big\},$$

where the *specific information* of outcome y given source X is

$$I_{\text{spec}}(y; X) = \sum_{x} p(x \mid y) \log \frac{p(y \mid x)}{p(y)} \quad (\geq 0).$$

Intuitively, I_{red} captures, for each target outcome separately, the lesser of the two sources' predictive contributions, then averages over outcomes. The remaining atoms are obtained by subtraction:

$$I_{\text{uniq}}(X_i) = I(X_i; Y) - I_{\text{red}}, \qquad i \in \{1, 2\}, \tag{5.1}$$

$$I_{\text{syn}} = I(\{X_1, X_2\}; Y) - I(X_1; Y) - I(X_2; Y) + I_{\text{red}}. \tag{5.2}$$

Alternative redundancy measures. $I_{\min}$ is a closed-form computation over probability tables and is therefore fast, but it can overestimate redundancy in certain joint distributions [105]. The BROJA measure [105, 106] takes a different approach: it defines unique information as the minimum conditional mutual information $I(X_1; Y \mid X_2)$ over all joint distributions $q(x_1, x_2, y)$ that preserve the observed marginals $p(x_1, y)$ and $p(x_2, y)$, and derives redundancy by subtraction. Because BROJA requires solving a convex optimisation problem over the polytope of consistent distributions [165], it is substantially slower than $I_{\min}$, particularly for variables with many levels. We therefore used $I_{\min}$ for the primary analyses. Sensitivity analyses with BROJA and I_{mmi} are reported in Appendix D.3.1. All three measures produced qualitatively similar results. BROJA gave systematically lower synergy magnitudes but preserved the ranking of pairs (Spearman $\rho = 0.989$ against $I_{\min}$), so the redundancy definition moves the reported magnitudes rather than the rankings. Computations were performed using the `dit` Python library [166].

We used I_{syn} to identify irreducibly joint contributions beyond main effects. Throughout, we reported synergy both in absolute bits and as a percentage of TMI, defined as the sum of PID atoms.

Assumptions and scope of the PID framework. PID requires a joint probability mass function $p(x_1, x_2, y)$ but imposes no parametric assumptions beyond this: it does not presuppose linearity, a particular link function, or a specific noise model. The framework is therefore applicable whenever a discrete (or discretised) joint distribution can be estimated from observed frequencies. Source correlation does not violate any assumption of PID, though it affects which atoms carry information. Our simulation study (Section 5.5.8) confirms accurate recovery under both independent and correlated sources. Continuous PID formulations exist (e.g., Gaussian PID [135, 163]), but no mature framework currently handles the mix of binary, ordinal, and continuous variables present in our dataset; see the Study Limitations section for a fuller discussion. Our discretisation strategy handled all variable types within a single estimation framework while producing clinically interpretable categories. Adequate sample sizes are needed to populate the joint states of each source–source–target triplet. The Data and Preprocessing section describes the discretisation and the sparse-state merging that address this requirement, and Section 5.5.5 describes the bias correction.

5.5.3. Synergistic information

Synergistic information is the piece of predictive power for a third variable (target) that appears only when two source variables are observed together. If either source is considered in isolation, this component of the signal vanishes. It represents information that is carried jointly by both sources and cannot be recovered from either alone. Consequently, any model that assumes the total effect of two predictors can be written as a sum (additive) or a product (multiplicative/independent-action) will miss that irreducible joint, or synergistic, part.

Mutual information (MI) quantifies how much uncertainty about an outcome is reduced by observing a feature, without assuming a specific model or scale. Partial Information Decomposition (PID) separates the joint information from two features into redundant, unique, and synergistic parts, allowing us to test whether combinations matter beyond their individual contributions.

For example, if PID indicates that 18% of the information predicting an outcome is synergistic, then an additive or multiplicative model cannot recover that 18% from main effects alone; at minimum, an explicit interaction term ($X_1 X_2$) or a more flexible model is required to capture the observed pattern.

5.5.4. Screening candidate source–target triplets

To focus computation on informative groups, we screened triplets $(X_a, X_b \to Y_j)$ by testing $I(\{X_a, X_b\}; Y_j) > 0$ using 500 target-label permutations per triplet while preserving the joint distribution of (X_a, X_b). We controlled the false discovery rate across triplets using the Benjamini–Hochberg procedure at $q = 0.05$ and retained dependence-positive triplets for PID.

5.5.5. PID estimation with bias correction

Plug-in estimates of information-theoretic quantities are positively biased in finite samples. For multivariate measures such as O-information, this bias can distort the balance between redundancy and synergy [113]; for PID atoms specifically, Koçillari et al. [167] showed that the bias of synergy grows *quadratically* with the number of discrete variable levels, whereas biases in unique and redundant information grow more slowly. Correcting for finite-sample bias is therefore especially important for synergy estimation in variables with multiple categories.

The expectation of a plug-in information estimator admits a series expansion in inverse sample size [167]:

$$\mathbb{E}[\hat{I}] = I_\infty + \frac{a_1}{N} + \frac{a_2}{N^2} + \mathcal{O}(N^{-3}),$$

where I_∞ is the population quantity of interest and a_1, a_2 are unknown, distribution-dependent constants. Quadratic extrapolation (QE) exploits this expansion by evaluating the estimator at multiple sample sizes and fitting a degree-2 polynomial in $1/N$; the intercept recovers I_∞, removing both the first- and second-order bias terms. A linear

(two-point) extrapolation would leave the a_2/N^2 component uncorrected—a material concern given the quadratic scaling of synergy bias with variable cardinality.

Concretely, for each PID atom we compute the plug-in estimate on three nested samples: the full dataset (N observations), a random subsample of $N/2$, and a random subsample of $N/4$. These three evaluation points at $1/N$, $2/N$ and $4/N$ define a degree-2 polynomial whose intercept (at $1/N \to 0$) is the QE estimate.

To remove any residual estimator-specific bias, we combine QE with a target-shuffle null: we repeat the same three-point QE procedure on data in which the target variable has been randomly permuted (while preserving the joint distribution of the two sources), and subtract the shuffled QE estimate from the original:

$$\mathrm{PID}_{\mathrm{QE+SH}} = \mathrm{QE}\big(\mathrm{PID}_{\mathrm{orig}}\big) - \mathrm{QE}(\mathrm{PID}_{\mathrm{shuffled}})\,.$$

By shuffling only the target while keeping the sources fixed, the joint dependency structure among sources is preserved. This ensures that the null distribution reflects the absence of genuine source–target information without destroying the joint information shared between sources. We average over K=30 independent target shuffles per subsample fraction to stabilise the null estimate.

We truncate negative atoms to zero for reporting. This preserves the non-negativity of reported synergies but breaks exact additivity of PID atoms, and it can inflate the prevalence of very small positive synergies near zero.

5.5.6. Breadth–Uniformity–Synergy–Total (BUST) classification

For each unordered source pair (X_a, X_b) we asked two related questions: (i) is its *synergy* spread across many targets (broad) or concentrated on a few (narrow)? (ii) are those synergistic effects similar across population strata (uniform) or do they vary (specific)? We answered the first by quantifying how the pair's synergy is *distributed* across targets, and the second by quantifying how *even* that distribution is across strata. Both quantities are built from normalised synergy-based probability distributions and summarised using bounded deviation-from-uniformity measures.

Notation. Let $\mathscr{J}$ denote the set of all targets. For each unordered pair (X_a, X_b) and target $j \in \mathscr{J}$, let $s_j \geq 0$ be the bias-corrected PID synergy (Section 5.5.5), with $s_j = 0$ when no positive synergy is detected. The pair's total synergy strength is

$$S = \sum_{j \in \mathscr{J}} s_j,$$

and the associated *synergy shares* across targets are

$$w_j = \frac{s_j}{S}, \qquad j \in \mathscr{J}, \quad \sum_j w_j = 1,$$

defined whenever $S > 0$. For stratified analyses, let $s_j^{(s)} \geq 0$ and $\mathrm{TMI}_j^{(s)} \geq 0$ denote the synergy and TMI (sum of all PID atoms) for target j in stratum s of stratification type t (sex, multimorbidity, or age group).

Breadth (B): synergy spread across targets. We quantify how broadly synergy is distributed over targets using the normalised Shannon entropy of the synergy shares $\{w_j\}_{j\in\mathscr{J}}$. For each pair with $S > 0$ we define

$$H_B = -\frac{\sum_{j\in\mathscr{J}} w_j \log w_j}{\log|\mathscr{J}|} \in [0,1],$$

with $H_B = 0$ when all synergy is concentrated on a single target and $H_B = 1$ when synergy is perfectly uniform across all targets. We then compute the median breadth across all pairs, $H_{B,\mathrm{med}}$, and map H_B to a median-centred breadth score

$$B = 2\left(\frac{H_B}{H_{B,\mathrm{med}}} - 1\right),$$

clipped to $[-1,1]$ for stability. Thus $B \approx +1$ denotes broadly distributed synergy (generalist behaviour), $B \approx -1$ denotes highly concentrated synergy (specialist behaviour), and $B \approx 0$ corresponds to the median breadth observed in the data. Pairs with $S = 0$ do not receive a B score and are excluded from subsequent BUST plots.

As robustness checks, we also compute alternative breadth measures by replacing H_B with (i) 1 minus a normalised Gini coefficient of the synergy shares, and (ii) 1 minus a normalised Kolmogorov–Smirnov (KS) distance between the empirical cumulative distribution of w_j and the uniform distribution. These alternative B scores are reported in the Supplement and are highly correlated with the entropy-based B.

Uniformity (U): cross-stratum stability of synergy. To characterise whether a pair's synergy is uniform across strata or context-specific, we first form synergy-based distributions across strata for each target and stratification type. For target j and stratum s of type t we define the *synergy fraction*

$$\pi_j^{(s,t)} = \begin{cases} s_j^{(s)}/\mathrm{TMI}_j^{(s)}, & \mathrm{TMI}_j^{(s)} > 0, \\ 0, & \mathrm{TMI}_j^{(s)} = 0, \end{cases}$$

so that $\pi_j^{(s,t)} \in [0,1]$ whenever defined. For a given j and t, we collect the fractions $\{\pi_j^{(s,t)} : s \in \mathscr{S}_t\}$ across strata, add a small stabilising constant ε to each value to avoid zero cells, and normalise to obtain a probability vector

$$u_j^{(s,t)} = \frac{\pi_j^{(s,t)} + \varepsilon}{\sum_{r\in\mathscr{S}_t}(\pi_j^{(r,t)} + \varepsilon)}, \qquad s \in \mathscr{S}_t,$$

whenever at least two strata contribute non-zero information. The same ε is used for every pair, target and stratum. For each target–type combination we then compute the normalised entropy

$$H_{U,j}^{(t)} = -\frac{\sum_{s\in\mathscr{S}_t} u_j^{(s,t)} \log u_j^{(s,t)}}{\log|\mathscr{S}_t|} \in [0,1],$$

so that $H_{U,j}^{(t)} \approx 1$ when synergy fractions are similar across strata of type t (uniform), and $H_{U,j}^{(t)} \approx 0$ when synergy is concentrated in a subset of strata (specific to e.g. a particular age group or sex). Normalisation removes the overall level of synergy, so $H_{U,j}^{(t)}$ describes the shape of the across-stratum profile rather than its magnitude. A target whose synergy fraction is near zero in every stratum therefore yields a near-uniform vector and a high $H_{U,j}^{(t)}$. Uniformity measures cross-stratum consistency, not importance.

Because some targets contribute little synergy overall and may be noisy, we aggregate these target-level uniformities using weights proportional to their full-population synergy s_j. For each stratification type t we define

$$H_U^{(t)} = \frac{\sum_j s_j H_{U,j}^{(t)}}{\sum_j s_j},$$

summing over targets with well-defined $H_{U,j}^{(t)}$ and $s_j > 0$. The overall uniformity index for a pair is then obtained by averaging across available stratification types:

$$H_U = \frac{1}{|\mathscr{T}^*|} \sum_{t\in\mathscr{T}^*} H_U^{(t)},$$

where $\mathscr{T}^*$ is the subset of types (sex, multimorbidity, age group) with at least two informative strata. As for breadth, we median-centre this raw uniformity index using its empirical median $H_{U,\text{med}}$ across pairs and define

$$U = 2\left(\frac{H_U}{H_{U,\text{med}}} - 1\right),$$

clipped to $[-1,1]$. Large positive U values indicate synergy fractions that are stable across strata (uniform, or "universal"). Large negative values indicate synergy concentrated in specific demographic or multimorbidity strata (context-specific).

Coefficient-of-variation-based metrics become unstable when means are small. All U scores are therefore built from bounded probability distributions $\{u_j^{(s,t)}\}$ with explicit ε-stabilisation, which avoids division by near-zero denominators. In sensitivity analyses we recompute uniformity using (i) 1 minus a normalised Gini coefficient and (ii) 1 minus a normalised KS distance for the $\{u_j^{(s,t)}\}$, rather than entropy. These alternative U measures are highly correlated with the entropy-based U and do not materially change the BUST map.

Synergy strength (S) and total information (T). For each pair, the *synergy strength* is defined as the total amount of positive synergy across all targets,

$$S = \sum_{j \in \mathcal{J}} s_j,$$

measured in bits. The *total information* (T) is defined by summing TMI_j over all targets with non-zero synergy in the full population,

$$T = \sum_{j \in \mathcal{J}} \mathrm{TMI}_j,$$

thereby capturing the overall information content carried by the sources about the outcome space, irrespective of whether it is synergistic, redundant, or unique. The BUST map uses (B, U) as a 2D embedding and encodes S (and, where available, T) through bubble size and colour to visualise which pairs are broad, uniform and information-rich.

5

5.5.7. Analysing BUST networks

Composite score construction

To prioritise source pairs that are simultaneously broad, cross-stratum robust, strongly synergistic, and information-rich, we constructed a composite BUST score for network analysis. First, the breadth and uniformity scores (B, U) obtained from the entropy-based metrics were already scaled to $[-1, 1]$ as described above. Second, synergy strength S and total information T were each transformed to $[-1, 1]$ by median-centring and applying a hyperbolic tangent transform to soften the influence of extreme values:

$$S^* = \tanh\left(2\left(\frac{S}{\mathrm{median}(S)} - 1\right)\right), \qquad T^* = \tanh\left(2\left(\frac{T}{\mathrm{median}(T)} - 1\right)\right),$$

with T^* omitted if reliable TMI was unavailable. All four dimensions were then linearly mapped from $[-1, 1]$ to $[0, 1]$ and combined multiplicatively,

$$\mathrm{BUST}_{\mathrm{comp}} = B_{01} \times U_{01} \times S_{01} \times T_{01},$$

where each subscript 01 denotes the corresponding rescaled value in $[0, 1]$. The multiplicative form acts as a soft *AND-gate*: a pair must score reasonably well on *all* dimensions (breadth, uniformity, synergy strength and total information) to attain a high composite value, while a low score in any dimension down-weights the overall importance. This avoids selecting pairs that are extreme on a single axis but weak on the others, and reflects our interest in identifying edges that are jointly broad, robust and information-rich. For network visualisations we defined "strong" BUST edges as those with composite scores above the 75th percentile of $\mathrm{BUST}_{\mathrm{comp}}$. The resulting network structure was robust to alternative percentile cut-offs and to using Gini- or KS-based B and U in place of the entropy-based metrics (see Appendix D.3.1).

Synergistic clique detection and stability analysis
Stable synergistic cliques were identified with a threshold-based criterion applied at four cut-offs. We constructed a weighted graph where nodes represented variables and edges were weighted by the composite BUST scores integrating breadth, uniformity, synergy strength and total information (Section 5.5.7). Cliques (fully connected subgraphs of size ≥ 3) were enumerated independently at four thresholds (67th, 75th, 80th, 85th percentiles of composite scores). Cliques appearing at three or more of the four thresholds were classified as stable. The four thresholded graphs are nested, so a clique that survives a higher cut-off survives every lower one. Requiring three of the four therefore amounts to requiring the clique's weakest edge to clear the 80th percentile. The count records the strength of that weakest edge, not robustness to resampling. Each stable clique was characterised by its mean composite weight, average B and U scores, and total synergy strength, then ranked by a weighted importance score

$$\text{Imp} = 0.4 \times \text{stability} + 0.3 \times \text{size} + 0.3 \times \text{mean weight}.$$

This choice reflects three design considerations. First, we prioritised *stability* across thresholds (weight 0.4), because cliques that recur under different cut-offs are less likely to be artefacts of a particular threshold. Second, we favoured cliques that are *larger* (weight 0.3), as these capture broader constellations of jointly interacting variables. Third, we gave comparable weight to the *mean composite edge weight* (weight 0.3), so that cliques with consistently strong BUST edges are preferred over those formed by weaker links. These weights are heuristic but reflect a conservative preference for robust, sizeable and strongly connected cliques rather than small or unstable motifs driven by a single strong edge.

Rule-based characterisation of joint-state behaviour
To characterise how BUST-identified synergistic constellations behaved with respect to individual outcomes, we examined the distribution of each target variable across all observed joint states of the constellation variables. For each constellation–target combination, we fitted an ordinal regression model with the target as the response and the constellation variables as main effects, and used the model to compute the expected value of the target under an additive assumption. We then quantified, for each joint state, the deviation between its observed mean and this additive prediction. Joint states whose absolute deviation lay in the upper 5% were labelled rule-like configurations. To assess whether these configurations contributed information beyond the main effects, we refitted the ordinal models with binary indicators for the rule-like states and compared model fits using the Akaike Information Criterion. Detailed results for each constellation and target are provided in Section D.6.

5.5.8. Simulation validation of the estimation pipeline

To verify that the QE bias-corrected PID estimation pipeline (Section 5.5.5) and the BUST summary scores (Section 5.5.6) recover known ground truths under realistic conditions,

we conducted three sets of simulation experiments with analytically defined joint distributions whose PID atoms were computed numerically using the `dit` library [166].

Cardinality sweep. We varied the number of levels per variable ($K \in \{2,3,4,5\}$) across four deterministic target functions (modular copy, XOR, minimum, and a redundancy-dominated function). Each function was run at two source-coupling strengths ($c = 0$ and $c = 0.5$) and four sample sizes ($N \in \{500, 1{,}000, 2{,}000, 2{,}636\}$). The largest matches one of the age-group strata in the LASA analyses. For each of the 128 conditions, 50 independent replicates were drawn.

LASA-matched cardinality profiles. Because the empirical cardinalities are asymmetric, we also constructed four profiles matching common LASA variable-cardinality combinations ($3 \times 4 \to 5$, $2 \times 5 \to 5$, $3 \times 7 \to 5$, $5 \times 7 \to 5$). Each profile was run under three information regimes: synergy-rich, function-of-one, and mixed.

BUST score recovery. We then tested whether the breadth (B), synergy strength (S) and total information (T) scores were recovered from estimated rather than ground-truth PID atoms. Three synthetic source pairs with known breadth profiles were used, each evaluated across eight targets.

Bootstrap confidence intervals. To quantify sampling variability, we computed percentile bootstrap 95% confidence intervals for the top-10 BUST-ranked source pairs (1,000 replicates, resampling the full-population dataset with replacement and re-estimating QE-corrected PID atoms per replicate). Full details and results are in Section D.9.

Results of the simulation experiments are presented in Section 5.2.2, with full figures in Section D.7.

5.5.9. Sensitivity to discretisation choices

To assess whether the BUST classification is sensitive to the choice of discretisation strategy, we re-ran the full PID pipeline on the LASA full-population dataset under five alternative binning schemes: (i) the clinical cut-points of the primary analysis, (ii) equal-frequency tertiles, (iii) equal-frequency quartiles, (iv) equal-width tertiles, and (v) a coarsened binary split at each variable's median. For each scheme we recomputed all PID atoms with QE bias correction and rebuilt the BUST scores. Rankings were compared by Spearman correlation of composite BUST scores and Jaccard overlap of the top-10 pairs. Among the four non-binary schemes, correlations ranged from 0.83 to 0.96. Equal-frequency tertiles and quartiles were nearly interchangeable ($\rho = 0.96$, Jaccard top-10 = 0.82). The coarsened binary scheme correlated poorly with all others ($\rho = 0.06$–0.18), as expected from collapsing multi-level variables to two categories. Full results, including BUST map comparisons and rank correlation matrices, are in Section D.8.

5.5.10. Data and preprocessing

All empirical results in this chapter come from one cohort. We analysed data from the Longitudinal Ageing Study Amsterdam (LASA), an ongoing population-based cohort designed to study physical, emotional, cognitive and social functioning in older adults in the Netherlands. The original cohort of 3,107 community-dwelling adults aged 55–85 years was recruited in 1992/93 and has been followed triennially ever since, with refreshment samples drawn in 2002/03 (n = 1,002) and 2012/13 (n = 1,023) to maintain population representativeness [145]. Data are collected through face-to-face structured interviews and self-completed questionnaires.

For the present analyses, we used data from the 1995/96 through 2021/22 waves (nine waves, each approximately three years apart). After excluding interviews with more than 50% missing values and applying random-forest imputation (see below), the analytic sample comprised 14,716 person-wave observations from 4,025 unique participants. Chapter 4 reports 14,717 observations for the same waves and the same exclusion rule, because there the missingness screen covered a partly different variable set. Across all waves these participants were on average 73 (SD 8) years of age, with a range of 57–102 years, 46% were women and 59% had multimorbidity (Chapter 4). We focused on variables that are central to multimorbidity and ageing research and that appear throughout the results. These are symptoms (sleep problems, pain, anxiety, depression), clinical signs (grip strength, body mass index, waist–hip ratio, mean arterial pressure) and behaviours (physical activity, alcohol use). The remaining domains are self-rated health (perceived general health, perceived health relative to peers), cognition (MMSE score, memory complaints, memory problems), sensory function (hearing difficulties) and activities of daily living (functional limitations).

Functional limitations

Functional limitations were assessed with six activities of daily living (ADL). The items were climbing a flight of stairs, dressing/undressing, rising from a chair, cutting toenails, walking outdoors for five minutes, and using transportation. Each item was scored on a 5-point ordinal scale (0 = “yes, without difficulty” to 4 = “no, cannot”). We computed a total score by summing the six items (range 0–24, higher scores indicating worse function). To improve interpretability while maintaining gradation, we defined a five-level variable: “none” (total = 0) and quartiles of the distribution of positive totals among participants with any limitation (total > 0).

Discretisation of symptoms and signs

All variables were analysed as discrete states. Where available, we discretised continuous measures using clinically interpretable cut-points. Otherwise we retained the native ordinal response categories described in LASA documentation. To ensure adequate support for information-theoretic estimation in three-way contingency tables, we applied an additional deterministic post-processing step that merged adjacent categories to eliminate sparse observed triplet states, ensuring no observed triplet state had fewer than

5 observations. Table 5.1 lists the number of levels each variable retained after that step. The rules themselves follow, grouped by domain.

Signs. Body mass index (BMI) was categorised following the World Health Organization classifications [168]: under/normal weight ($< 25.0\,\mathrm{kg/m^2}$), overweight (25.0–29.9 $\mathrm{kg/m^2}$) and obesity ($\geq 30.0\,\mathrm{kg/m^2}$). Underweight was not separated from normal weight, so synergy specific to low BMI cannot be resolved here. For dominant grip strength, we implemented a two-step process based on the EWGSOP2 sarcopenia guidelines [169]. Individuals with grip strength below 27 kg for men and below 16 kg for women were classified as having 'Low' strength. Those at or above these thresholds were divided into quartiles by sex to capture further gradations in strength. The top two quartiles were combined because of data sparsity. Mean arterial pressure (MAP; blood pressure) was dichotomised at 93 mmHg into lower (< 93 mmHg) and higher (≥ 93 mmHg) categories. That cut-point corresponds to a blood pressure of roughly 120/80 mmHg, well below the MAP of approximately 107 mmHg implied by the clinical hypertension threshold of 140/90 mmHg. The upper category therefore indexes comparatively higher blood pressure rather than hypertension.

Symptoms. The sleep quality measure, derived by summing answers to three categorical questions (problems falling asleep, waking through the night and waking up too early), ranged from 3 (no problems) to 12 (many problems). To reduce sparsity, the lowest two scores were combined and the upper tail collapsed, giving states 3–4, 5, 6, 7, 8, 9 and ≥ 10 (i.e., 10–12 combined). Pain was measured with the LASA 6-item Nottingham Health Profile subscale (standing, changing position, sitting, walking, constant pain; the "unbearable pain" item is excluded from scoring). Responses were dichotomous (yes/no) [170]. Following LASA documentation, we recoded no as 1 and yes as 2, summed the five retained items with single-item mean imputation when one response was missing (yielding a 5–10 score), and then linearly rescaled this sum to a 1–6 ordinal score. The two highest levels were combined, yielding five ordered levels (1, 2, 3, 4 and ≥ 5), where higher scores mean more pain symptoms. Depressive symptoms were assessed with the Center for Epidemiologic Studies Depression Scale (CES-D) [171], grouped into two ordered categories: lower (0–19) and higher (≥ 20) symptom burden. Anxiety symptoms were identified using the anxiety subscale of the Hospital Anxiety and Depression Scale (HADS) [172], grouped into lower (0–3) and higher (≥ 4) categories. That split lies below the validated caseness threshold of ≥ 8 for the subscale, so the upper category should be read as comparatively elevated anxiety rather than probable caseness.

Cognition. Cognition was assessed with the Mini-Mental State Examination (MMSE; 0–30). Following LASA documentation, we computed the total by summing 22 items and using the higher of the two attention/calculation alternatives (serial 7s or spelling 'WORLD' backwards). Missing items were assigned 0, and no total was computed if more than 6 items were missing. We then categorised cognition into two strata: impaired

(0–23) versus no impairment (24–30). Memory complaints were determined by a yes/no response to the question: “Do you have complaints about your memory?”. Similarly, memory problems were determined by a yes/no response to the question: “Did you consult a doctor about your memory problems?”.

Behaviours and self-rated health. Physical activity was assessed with a standard LASA activity scale and discretised into tertiles. Alcohol use was categorised from weekly consumption. The three levels were abstinent, light/moderate and heavy. Perceived general health was retained as an ordinal measure, with the two highest response categories combined to give three ordered levels. Perceived health relative to peers was likewise retained as an ordinal measure, with the two lowest response categories combined to give three ordered levels.

Missing data
Missing data were imputed prior to discretisation. Known invalid responses were treated as missing, and interviews with more than 50% missing values across relevant variables were excluded to minimise any imputation bias. Random forest–based imputation was used because it can accommodate complex nonlinear relations. Continuous and categorical variables were imputed using separate models (regressor vs. classifier), each with a maximum tree depth of 5, a minimum of 10 samples per leaf and 100 trees.

Post-hoc merging of sparse categories
After applying the clinically motivated discretisation cut-points described above, we performed an additional merging step to reduce sparsity in the higher-order joint states required for information-theoretic estimation. Using the full-sample dataset, we enumerated all variable triplets and computed the observed 3-way joint frequency table for each triplet (i.e., counts of each observed joint category combination within the triplet). We required every observed 3-way joint state to have at least 5 observations. If any triplet contained joint states with count < 5, we iteratively selected the single *variable–state* pair that appeared in the largest number of sparse triplets and merged that state with an adjacent ordinal neighbour (lower or higher category). When both neighbours were available, the merge direction was chosen to preferentially combine with the neighbour that was itself involved in more sparse triplets. This procedure was repeated until no observed 3-way joint state had fewer than 5 observations. The merged states are shown in Table D.3.

Stratification
All analyses were repeated in strata defined *a priori* by age, sex and multimorbidity to assess whether patterns of synergy were uniform or population-specific. Multimorbidity was encoded as a binary variable, marked present when a person reported two or more chronic diseases. Sex and multimorbidity each defined two strata, and age at interview defined four strata—under 65, 65–74, 75–84, and 85 or older—reported as AgeGroup_0–AgeGroup_3 in Appendix D.

Variables included in the synergy analysis
All variables except age, sex and multimorbidity (used solely for stratification) were included as candidates for both predictors and outcomes in the PID triplets. Treating variables symmetrically allows estimation of how any pair jointly informs each remaining variable, supporting the Breadth, Uniformity and Total components of the BUST representation. Table 5.1 summarises these variables with their domain classification and discretisation basis.

Table 5.1: Clinical variables included in the PID synergy analysis.

Variable	Domain	Levels	Discretisation basis
Sleep quality	Symptom	7	Sum of 3 items; tail-collapsed
Pain	Symptom	5	NHP subscale; rescaled sum
Anxiety	Symptom	2	HADS anxiety subscale
Depression	Symptom	2	CES-D threshold
Hearing difficulties	Sensory	3	Self-report ordinal
Memory complaints	Cognition	2	Self-report (yes/no)
Memory problems	Cognition	2	Self-report (yes/no)
MMSE (cognition)	Cognition	2	MMSE total; clinical cut-point
Grip strength	Sign	4	EWGSOP2 cut-points + sex-specific quartiles
BMI	Sign	3	WHO classification
WHR obesity	Sign	2	Clinical cut-point
Blood pressure (MAP)	Sign	2	Normal vs. elevated/hypertensive
Physical activity	Behaviour	3	LASA activity scale; tertiles
Alcohol use	Behaviour	3	Weekly consumption categories
Perceived general health	Self-rated health	3	Ordinal; top categories combined
Health vs. peers	Self-rated health	3	Ordinal; bottom categories combined
Functional limitations	ADL	3	ADL sum score; none vs. quartiles

Levels is the number of discrete categories. Sparse states were merged post hoc.
Abbreviations. ADL = activities of daily living, BMI = body mass index, MAP = mean arterial pressure, WHR = waist–hip ratio. CES-D = Center for Epidemiologic Studies Depression Scale, HADS = Hospital Anxiety and Depression Scale. MMSE = Mini-Mental State Examination, NHP = Nottingham Health Profile. EWGSOP2 = European Working Group on Sarcopenia in Older People 2, WHO = World Health Organization.
Memory complaints, memory problems and MMSE are grouped under "Cognition". Other classifications are defensible. Memory complaints could count as a symptom and MMSE as a clinical sign. The grouping does not affect the PID or BUST results.

Data and code availability

The LASA data used in this study are not publicly available because they contain sensitive participant information and are distributed under governed access. Researchers may request access to LASA data for specific research questions by submitting an analysis proposal to the LASA Steering Group. If approved, access is provided under a data-sharing agreement. Details and the required proposal materials are available via the LASA website: `https://lasa-vu.nl/en/request-data/`.

The analysis pipeline is publicly available as the open-source Python package `bust-pid` at `https://github.com/cillianhourican/bust-pid`. The package implements permutation screening, QE-corrected PID estimation, BUST scoring, network construction, and visualisation, and includes reproduction scripts for the main paper figures (requiring LASA data access) as well as a synthetic-data demo.



Chapter 6

EMBEDDING-BASED PARTIAL INFORMATION DECOMPOSITION FOR SYMPTOM NETWORKS AT SYSTEM SCALE

Based on:

Hourican, C., Dignum, E., Roy, D., & Quax, R. (2026). Embedding-based partial information decomposition for symptom networks at system scale. Manuscript submitted for publication.

The preceding chapter developed a pair-centric interpretation framework (BUST) that renders higher-order dependence comparable across targets and population strata, but the underlying decompositions remained at the triplet level: two sources and one target, with no explicit representation of the surrounding variable context. In multimorbidity cohort data with a modest number of health-related variables, triplet-level PID can be computed exhaustively; in symptom networks, however, the relevant comparator for a focal symptom is not a single companion but the remainder of the system, and the number of PID atoms grows super-exponentially with the source set. The interpretive question therefore shifts from a triplet-level one—"do these two variables interact for this target?"—to a system-level one: given all remaining symptoms in a network, does a single focal symptom still carry unique or synergistic information about a target, or is its contribution already captured by the rest of the profile? This chapter addresses that question by introducing ePID, an embedding-based partial information decomposition that compresses the multivariate remainder into a supervised discrete representation and computes a tractable two-source PID for each source–target relation. The chapter also marks a domain shift: from the multimorbidity and ageing cohorts of Chapters 4 and 5 to depressive symptom networks in psychopathology. Depressive symptoms are a well-suited test case because the network approach to mental disorders already theorises that symptoms interact as a system rather than serving merely as interchangeable indicators of a latent disease, and the PHQ-9 is a well-validated nine-item instrument whose inter-item structure has been extensively studied. Moreover, two independent population cohorts are available, enabling direct cross-cohort stability testing. The analysis applies ePID to PHQ-9 items in these cohorts, yielding the first network-context redundancy–synergy decomposition at symptom-system scale, and—as a contrasting test case—to the Interpersonal Reactivity Index, an empathy instrument with a very different internal structure. Underpinned by a benchmark spanning synthetic networks and 83 real-world datasets, the results demonstrate that the composition of dependence is reproducible across cohorts and that the same pipeline discriminates redundancy-dominated instruments such as the PHQ-9 from synergy-dominated ones such as the Interpersonal Reactivity Index.

6.1. Introduction

Network models represent multivariate systems as nodes and their statistical associations as edges, offering a framework in which the structure of interest emerges from the observed relationships themselves rather than from an assumed latent cause. This approach has been widely adopted in psychopathology, where symptom networks represent mental-health symptoms as nodes and conditional dependencies as edges [29, 85]. Rather than treating symptoms as passive indicators of a single disease entity, the network perspective treats them as mutually reinforcing elements whose interaction patterns constitute the disorder itself [173, 174]. In the most common formulation, edges quantify pairwise conditional associations between symptoms after adjusting for the remaining symptoms, yielding a sparse network often interpreted as a "direct" dependency structure. The pairwise-insufficiency argument developed for multimorbidity networks in Chapters 1 and 2 applies with equal force here: a scalar edge cannot express how several symptoms jointly inform a third.

For example, after a stressful event, an individual may experience insomnia that induces fatigue, which might affect concentration, in turn contributing to depressed mood and feelings of worthlessness [175]. This perspective has shifted the conceptualisation of mental disorders from a centralised latent-disease model to a decentralised network model. However, commonly employed estimators of symptom networks share a fundamental limitation that no choice of pairwise measure can overcome.

Each edge summarises the association between two symptoms as a single number. Even when that association is estimated without assuming a particular functional form—a useful safeguard—a scalar edge still cannot express how several symptoms act *together* to inform a third. That is the gap we address.

Consider how symptoms behave in context. Two symptoms may act together, so that their combined presence is informative even though neither tells us much alone; or they may act as overlapping sources, each carrying largely the same information about a third symptom. For instance, the sleep and fatigue items of the PHQ-9 remain correlated after conditioning on overall depression severity, so that observing one adds little once the other is known [176, 177]; by contrast, other symptom pairs may carry information about suicidal ideation jointly that neither carries alone, a possibility that pairwise screening has rarely tested [178]. A single number per pair cannot tell these cases apart. They can, however, be operationalised as distinct kinds of information: overlapping information that two symptoms share about a target (*redundancy*) and information that becomes available only when they are considered jointly (*synergy*), alongside the information each contributes uniquely.

This distinction has practical implications under an interventionist reading—one that, we stress, presupposes a causal structure and is vulnerable to hidden confounding, so the readings below are strong causal assumptions rather than established conclusions. If two source symptoms are highly redundant, changing the state of only one would change little for the target, as the other continues to convey the same information. Conversely, if their joint information is primarily synergistic, then whether disrupting either source suffices to break the joint channel depends on the underlying causal system: that the joint configuration carries the information does not by itself license the intervention. These in-

This chapter is based on a manuscript submitted for publication.

terpretations are hypothesis-generating and would require longitudinal or experimental validation; we revisit them in light of the empirical findings in the Discussion.

Information theory supplies the tools to make this operationalisation precise. A source variable provides *information* about a target when observing the source reduces uncertainty about the target's state [38]. Conditional mutual information (CMI) uses this idea to generalise partial correlation (with which it coincides as a conditional-independence criterion under joint Gaussianity) into a model-free (with respect to functional form) measure of conditional dependence. Partial information decomposition (PID), introduced formally in Chapter 1 and applied to multimorbidity cohorts in Chapters 4 and 5, goes further, decomposing the information that two or more *sources* provide about a designated *target* into the non-overlapping components, or *atoms* [22], sketched above—redundant, unique, and synergistic—so that a conditional association can be read as overlapping or interaction-only rather than as a single undifferentiated number.

The obstacle to applying PID is scale—related to but distinct from the combinatorial explosion of candidate variable sets addressed in Chapter 3. The number of atoms grows super-exponentially with the number of sources, from 4 atoms for 2 sources to over 7 million for 6, following the Dedekind numbers $M(N)$ [164]. Estimating so many atoms requires high-dimensional joint distributions that are severely undersampled at typical clinical sample sizes, which in practice restricts most analyses to the triplet level (two sources and one target).

Although triplet-level decompositions are tractable, they miss interactions that emerge only when three or more sources act together. In symptom networks the relevant comparator for a given symptom is its full context—all remaining symptoms—so triplet decompositions cannot answer the system-level attribution question of primary interest: whether a symptom contributes information about a target *beyond what is already present in the rest of the symptom profile*.

To address this, we introduce an embedding-based PID pipeline, which we refer to as *ePID*. For a given pair of symptoms, ePID first compresses the rest of the network into a compact summary variable—a small number of categories learned to preserve as much information as possible about the *target* symptom specifically—and then asks how the focal symptom (i.e., the source in the pair) relates to the target compared with that summary. Because such a decomposition is defined entirely by information about the target, a summary optimised to preserve the remainder's target-relevant information retains what it depends on, while what the summary discards is, by construction, largely irrelevant to the target; we make this approximation, and the information it loses, precise in the Methods. This target-supervised compression is specific to ePID; partial correlation and CMI, by contrast, are not directed at any particular target. The procedure yields four components for each ordered symptom pair: information that only the focal symptom provides (source-unique), information that only the rest of the network provides (remainder-unique), overlapping information (redundancy), and information that emerges only when the focal symptom and the summary are considered together

(synergy). The resulting directed edges reflect the source–target orientation that the decomposition requires.

A complementary strand of psychometric network research tackles higher-order structure with Moderated Network Models (MNM), which add multiplicative interaction terms so that the association between two symptoms can vary with a third [179]. The contrast with ePID is essentially parametric versus non-parametric: MNM provides a parametric test of whether a chosen moderator significantly modifies a given edge, whereas ePID asks, without specifying interactions in advance, how the information underlying each edge is distributed across redundant, unique, and synergistic channels relative to the rest of the symptom profile. We develop the fuller methodological contrast in the Discussion.

This study estimates and compares three approaches to symptom network estimation that progressively relax modelling assumptions and move from pairwise to higher-order characterisations (Figure 6.1). Partial correlations (PC), in their Pearson (PPC) and Spearman (SPC) variants, provide a baseline for conditional linear and monotonic associations; the two variants yield highly similar networks on the ordinal symptom items analysed here (Appendix E.8, Figure E.14), and we therefore present them jointly as PC. CMI generalises to arbitrary conditional dependence while remaining at the pairwise level. To characterise how that dependence is composed, we apply PID via the ePID pipeline, yielding for each source–target pair a decomposition into the four atoms above. The degree to which these methods agree or diverge is itself informative about the statistical structure of the symptom system under study.

We evaluate ePID in two stages. First, we benchmark multiple embedding strategies in simulation settings calibrated to PHQ-9 response distributions, then extend the evaluation across 83 real-world datasets spanning five PID measures to assess generalisability, quantifying how much information the compression loses (the approximation error). Second, we apply the approach to item-level depressive-symptom networks in two datasets (UK Biobank [27, 180] and the Xinxiang student sample [181, 182]), using harmonised PHQ-9 items (excluding suicidal ideation) to assess cross-cohort stability, and to the Interpersonal Reactivity Index [183], a multi-facet empathy instrument that serves as a contrasting psychometric profile. In each application we first treat the agreement between partial correlations and CMI as an assumption check on whether conditional structure is largely monotonic at the item level, and then use ePID to add an interaction-aware interpretation layer, separating redundant from synergistic channels relative to the full symptom context.

By combining model-free measures of dependence with a scalable decomposition method, we characterise what kinds of statistical structure underlie observed symptom networks and assess whether, and to what extent, interaction-aware models are warranted for a given instrument and population. To our knowledge this is the first application and systematic validation of supervised-embedding PID to symptom networks, and although we develop it here for psychometric instruments, the pipeline is generic and should apply to other multivariate systems in which network-context higher-order

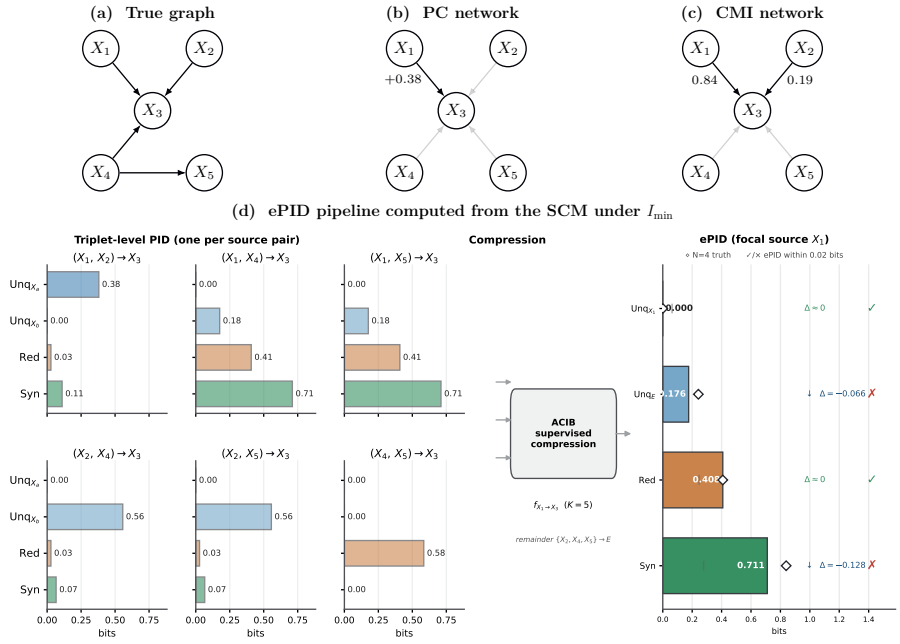


Figure 6.1: **Methods comparison on the toy structural causal model.** The toy system is a five-variable mixture structural causal model ($X_1, X_2, W \sim \text{Uniform}\{0,\ldots,4\}$ independent; $X_4 = X_5 = W$; $X_3 = h(X_1, X_2)$ with probability 0.5 else $X_3 = W$, where h is the non-monotonic stylised mapping defined in Appendix E.10). Top row: **(a)** True graph: $X_1, X_2, X_4 \to X_3$ with $X_4 = X_5$ marked. **(b)** PC network: only the $X_1 \to X_3$ edge survives ($\rho_{ij|\text{rest}} = +0.38$); X_4, X_5 edges are undefined under naive PC because the deterministic equality $X_4 = X_5$ makes the conditioning covariance singular. **(c)** CMI network: $X_1 \to X_3$ and the non-monotonic $X_2 \to X_3$ edge (0.84 and 0.19 bits respectively); X_4, X_5 contributions vanish because conditioning on the remaining system already includes the duplicate $X_5 = X_4$, giving each zero conditional entropy given the rest and hence $I(X_4; X_3 \mid \text{rest}) = 0$; the edges disappear from vanishing conditional mutual information, not from the conditioning rule itself, and the same attenuation arises under near-redundancy when the sources are highly but not perfectly correlated. Bottom row: **(d)** the ePID pipeline. Six triplet-level two-source PIDs (one per source pair targeting X_3, shown at lower opacity to signal pair-dependent attribution) collapse via ACIB-supervised compression of the remainder $\{X_2, X_4, X_5\}$ into a four-atom ePID for focal source X_1. All atom values are computed exactly under $I_{\min}$ (Williams & Beer 2010); the corresponding BROJA decomposition is shown in Appendix Figure E.1. PC = partial correlation; CMI = conditional mutual information; PID = partial information decomposition; ePID = embedding-based PID; ACIB = Agglomerative Conditional Information Bottleneck.

structure is of interest—a generalisation we do not yet validate empirically. We emphasise throughout that these patterns describe statistical associations in cross-sectional data rather than causal mechanisms, and should be read as constraints and hypotheses for future longitudinal and experimental work; their empirical magnitudes, and their behaviour across cohorts and instruments, are reported in the Results.

6.2. Methods

The methodology proceeds as follows. We first introduce notation and then describe two correlation-based methods: Pearson and Spearman partial correlations (PPC and SPC). Next, we explain conditional mutual information (CMI), which relaxes the linearity and monotonicity assumptions of PPC and SPC respectively. These three methods capture only pairwise associations.

Partial information decomposition (PID) goes further by decomposing the information that multiple source variables provide about a target into redundant, synergistic, and unique components, illustrated with a stylised example (Appendix E.10). To make this decomposition tractable beyond a handful of sources, we introduce an embedding-based PID method (ePID) that compresses the remaining variables into a low-dimensional proxy while retaining information about the target (Section 6.2.5). We then describe the simulation and multi-dataset benchmarks used to validate this approximation, define the normalisation and significance-testing framework, and outline the empirical-analysis procedures applied to the two PHQ-9 cohorts and the IRI empathy instrument: permutation testing with false discovery rate control for edge significance, entropy-normalised edge ranking for network visualisation, and cross-cohort stability assessment between the two PHQ-9 cohorts.

6.2.1. Notation

Let $\mathbf{X} = \{X_1, X_2, \ldots, X_p\}$ denote p discrete random variables. Each X_i can represent the presence or severity of symptom i in an individual, for example, or other covariates such as education or neighbourhood characteristics. The cardinality of a discrete variable X_i is denoted $|X_i|$ and equals the number of categories. For any ordered pair (X_i, X_j), we write $Z_{ij} = \mathbf{X} \setminus \{X_i, X_j\}$ for the remaining variables used as a conditioning set. For PID, we consider the triplets $\{X_i, X_j, X_k\}$ to be directed, where we refer to $\{X_i, X_j\}$ (inputs) as sources and $\{X_k\}$ (output) as the target.

For embedding-based decomposition, we consider an ordered source–target pair (X_i, X_k) and denote the remaining variables by

$$Z_{i\to k} := \{X_\ell : \ell \in \{1, \ldots, p\} \setminus \{i, k\}\}.$$

In words, $Z_{i\to k}$ collects every symptom other than the focal source X_i and the target X_k: the rest of the network, whose joint state we will compress into a low-dimensional embedding. We construct a supervised discrete embedding of this remainder,

$$E_{i\to k} := f_{i\to k}(Z_{i\to k}),$$

where $f_{i\to k}$ is fit to predict the target X_k from $Z_{i\to k}$ and $E_{i\to k}$ has finite cardinality $|E_{i\to k}|$. We then compute a two-source PID of $(X_i, E_{i\to k}; X_k)$, interpreting $E_{i\to k}$ as a compressed representation of the network context (the "remainder embedding").

6.2.2. Data

We analyse Patient Health Questionnaire (PHQ-9) item responses from two datasets: the UK Biobank online mental health questionnaire [180] and a large online sample of Chinese university students (Xinxiang Medical University; Su et al. 2024) [181, 182]. The PHQ-9 is a widely used and validated instrument for screening and severity assessment of depressive symptoms in clinical and population settings [184, 185]. Each item asks how often, in the last two weeks, respondents were bothered by a symptom, with ordinal response categories coded as 0 ("Not at all"), 1 ("Several days"), 2 ("More than half the days"), and 3 ("Nearly every day").

For the primary analyses, we focus on eight PHQ-9 items (anhedonia, depressed mood, sleep problems, fatigue/low energy, appetite change, feelings of worthlessness or excessive guilt, concentration problems, and psychomotor changes). We exclude the suicidality item (thoughts of self-harm) because it is extremely sparse in both datasets, leading to unstable estimation in conditional and higher-order analyses. To reduce sparsity in the remaining items and harmonise measurement across datasets, we merge response categories 2 and 3 and analyse a three-level ordinal scale $\{0, 1, 2\}$ for each retained item. Because categories 2 and 3 are the rarest, this adjacent merge preserves ordinality but yields an asymmetric scale on which the lowest-severity level is well sampled and the highest is compressed into a single cell, which reduces the attainable entropy, and hence the absolute mutual information in bits, for high-severity symptom combinations. Because we report atoms as shares of the target entropy (Section 6.2.8), however, this reduction acts largely as a common rescaling that cancels in the shares: recomputing a direct two-source PID at two, three, and four levels yields a near-constant synergy share of 0.193, 0.188, and 0.190 respectively, with the redundancy share stable at 0.534 to 0.542 and hence dominant throughout. We therefore treat the merge as likely conservative for synergy rather than as a fixed downward bias, and return to this point in Section 6.4.4. After preprocessing, the UK Biobank dataset comprised $N = 154{,}291$ participants and the Xinxiang student sample comprised $N = 24{,}292$ respondents.

To assess whether the ePID pipeline generalises beyond depressive-symptom networks to instruments with different higher-order structure, we additionally analyse responses to the Interpersonal Reactivity Index (IRI), a widely used multidimensional measure of dispositional empathy [183]. The IRI comprises 28 items distributed across four 7-item subscales: Perspective-Taking (PT) and Fantasy (FS), indexing cognitive-empathy facets, and Empathic Concern (EC) and Personal Distress (PD), indexing affective-empathy facets [183]. Responses are scored on a five-point Likert scale; for the present analyses we discretise to a three-level ordinal scale to harmonise cardinality with the PHQ-9 protocol. The dataset analysed here comprises $N = 1{,}973$ respondents from the Open-Source Psychometrics Project (`https://openpsychometrics.org/_rawdata/`). We apply ePID using the same protocol as for PHQ-9 ($n_{\text{sources}} = 5$ ACIB-embedded remainder context, I_{mmi} redundancy measure), supplemented by within-subscale and cross-subscale analyses reported in Section 6.3.4.

6.2.3. Correlation Networks

We represent the symptom system as an undirected graphical model in which nodes are symptoms and edges represent conditional dependence after adjusting for the remaining symptoms. Under a Gaussian (or Gaussian-copula) graphical model, the absence of an edge corresponds to conditional independence given all other symptoms, while a nonzero partial correlation indicates a conditional association within that model class.

The (population) partial correlation between X_i and X_j given Z_{ij}, denoted $\rho_{ij|Z_{ij}}$, is defined as the correlation between the residuals obtained by regressing X_i and X_j on Z_{ij}, respectively.

To relax marginal normality and linearity assumptions, we compute rank-based partial correlations (SPC). Specifically, each variable X_k is transformed by its empirical marginal distribution to obtain ranks $R(X_k)$, and partial correlations are computed on the transformed variables. This procedure estimates conditional associations under a Gaussian copula graphical model: dependencies are encoded by a latent multivariate normal structure while marginal distributions are left unrestricted. Rank-based partial correlations therefore capture conditional monotonic dependence, irrespective of marginal scale or skewness.

Partial correlations take values in $[-1,1]$ and define an undirected network in which edges represent conditional associations between symptom pairs.

6.2.4. Information-Theoretic Networks

To quantify conditional dependence without restricting the functional form of associations, we construct information-theoretic networks based on mutual information.

For a discrete random variable X, Shannon entropy $H(X)$ quantifies uncertainty about its outcome on a logarithmic scale and provides a natural uncertainty baseline for categorical data. For a variable with $|X|$ categories, $H(X) \leq \log_2 |X|$. Mutual information between two variables X_i and X_j,

$$I(X_i; X_j) = H(X_i) - H(X_i \mid X_j),$$

measures the reduction in uncertainty about one variable given knowledge of the other. Mutual information is symmetric, non-negative, and makes no assumptions about linearity or monotonicity.

To isolate direct associations in a network setting, we compute conditional mutual information (CMI),

$$I(X_i; X_j \mid Z_{ij}) = H(X_i \mid Z_{ij}) - H(X_i \mid X_j, Z_{ij}),$$

which measures the remaining dependence between X_i and X_j after conditioning on all other variables. Unlike correlation-based measures, CMI detects arbitrary conditional dependence. At the population level, $I(X_i; X_j \mid Z_{ij}) = 0$ if and only if X_i and X_j are conditionally independent given Z_{ij} for discrete variables, so CMI provides a model-free criterion for conditional independence. In finite samples, plug-in estimates of CMI are positively biased; we address this via permutation testing (Section 6.2.9).

Correlation-based measures and CMI estimate the same conditional dependence graph but differ in the class of dependencies they can detect. Partial correlations detect only linear or monotonic conditional associations, whereas CMI provides a necessary and sufficient test of conditional independence for discrete data. Because mutual information is in turn bounded above by the relevant entropies, we report a normalised CMI that can be interpreted as the fraction of remaining uncertainty explained (Section 6.2.8).

6.2.5. Higher-order interactions

Having established the pairwise measures, we now move from pairwise dependence to interactions among more than two variables. We proceed in three steps: (i) define the two-source PID and its redundancy function, (ii) introduce embedding-based PID (ePID), which compresses the multivariate remainder of the symptom system into a supervised discrete embedding $E_{i\to k}$ for each focal source–target pair, and (iii) map the resulting atoms onto directed network visualisations. A stylised three-variable example (Appendix E.10) builds intuition before the formal definitions.

Multivariate extensions of mutual information quantify whether dependence among groups of variables is primarily overlapping or interaction-specific. A core diagnostic is that conditioning can reveal dependencies invisible to marginal analysis: in an exclusive-or (XOR) relationship between two independent, uniformly distributed binary inputs, where the output indicates whether the inputs differ, each input alone carries no information about the output, whereas the joint state is fully informative. We apply partial information decomposition (PID) to obtain non-negative unique, redundant, and synergistic components, using an embedding of the remaining symptoms to make the decomposition tractable at PHQ-9 scale.

Several complementary frameworks provide scalable summaries of higher-order dependence without estimating a full multivariate PID. System-level measures such as the O-information (and related local or dynamical variants) quantify whether multivariate dependence is dominated by redundancy or synergy, but do not attribute information to specific source–target relations [42, 186]. Decision-theoretic and entropy-centric decompositions offer principled alternatives to the standard PID lattice: the redundancy bottleneck recasts redundancy as a constrained optimisation over compressed representations of the sources [187], while partial entropy decomposition operates on joint entropy rather than mutual information [43]. Both can scale via convex optimisation or algebraic structure. Finally, synergy-backbone approaches aim to identify irreducible synergistic subsets in large systems while avoiding the full lattice [188]. These methods highlight a common trade-off between scalability and attributional resolution; here we retain PID's directional, source–target interpretability while scaling to the full symptom context via ePID's remainder embedding (Section 6.2.5).

As a concrete anchor, consider a stylised three-variable system (stress, sleep problems, and concentration) in which concentration tracks sleep under normal conditions but, when sleep is very poor, depends on stress in a non-monotonic way. A partial correlation recovers only the sleep-to-concentration link; conditional mutual information addition-

ally flags the stress-to-concentration dependence; and PID makes the structure explicit, attributing about 71% of the information about concentration uniquely to sleep, 7% to redundancy, and 21% to synergy between stress and sleep (the interaction-only signal that partial correlation misses entirely and conditional mutual information detects only indirectly). The full worked construction is given in Appendix E.10. This pattern is not an artefact of the construction: in the real IRI empathy data, the analogous configuration (sources Fantasy and Personal Distress, target Perspective-Taking) is synergy-dominant in 69.4% of informative cross-subscale triplets, with a mean synergy fraction of about 21%, albeit only about 0.006 bits in total. We now formalise these atoms.

Two-source partial information decomposition

Partial information decomposition (PID) refines mutual information by decomposing the information that two sources (X_i, X_j) provide about a target X_k into four non-negative components (information "atoms"): redundant, unique-to-X_i, unique-to-X_j, and synergistic information. For discrete variables, we write

$$I(X_i, X_j; X_k) = \text{Red} + \text{Unq}_{X_i} + \text{Unq}_{X_j} + \text{Syn}, \tag{6.1}$$

subject to the consistency relations

$$I(X_i; X_k) = \text{Red} + \text{Unq}_{X_i}, \tag{6.2}$$

$$I(X_j; X_k) = \text{Red} + \text{Unq}_{X_j}. \tag{6.3}$$

That is, the four atoms partition the joint mutual information $I(X_i, X_j; X_k)$ into non-overlapping non-negative pieces, and individually they must sum to the marginal informations $I(X_i; X_k)$ and $I(X_j; X_k)$ via the consistency constraints. Here Red is information that both sources provide about the target, Unq_{X_i} and Unq_{X_j} are information that one source provides and the other does not, and Syn is information that only becomes available when the two sources are considered jointly (interaction-only effects).

Different PID measures pin down these four atoms in different ways. Most, including the minimum mutual information (MMI) redundancy measure we use in all embedding-based analyses, define the redundant information first and obtain the others from it; the BROJA measure (below) instead defines the unique information first, as the smallest unique information consistent with the source and target marginals. MMI satisfies self-redundancy, symmetry, and monotonicity and has been widely applied in discrete systems. MMI defines redundancy as the minimum information that any single source provides about the target, $\text{Red}_{\text{MMI}} = \min\{I(X_i; X_k), I(X_j; X_k)\}$, and derives the remaining atoms via the consistency relations above. This definition is intuitive (redundancy cannot exceed what any one source individually reveals) and is computationally efficient for low-cardinality systems. A known limitation of MMI is that it can overestimate redundancy when two sources carry different information about the target of equal magnitude, because the minimum over marginal mutual informations equates equal MI values with shared content regardless of whether the sources encode different aspects of the target [105]. Consequently, MMI may underestimate synergy. Alternative

redundancy definitions address this limitation from different angles. For example, the BROJA measure [105] defines redundancy as the smallest amount of unique information consistent with the observed source–target marginals, expressed as a decision-theoretic optimisation; but each introduces additional computational cost or conceptual trade-offs. To assess sensitivity to the redundancy definition, we additionally compute decompositions using four alternative measures as robustness checks: the $I_{\min}$ (Williams–Beer) measure [22], the original PID redundancy that takes the pointwise minimum specificity across sources; a rescaled redundancy measure I_{rr} [189] that normalises by source entropy; the pointwise partial information measure $I_{\pm}$ [73], which decomposes information at the level of individual joint outcomes rather than averaged distributions; and the Gács–Körner common-information measure $I_{\wedge}$ [190], which retains only the information that two sources jointly encode in a perfectly aligned (deterministic) way. Qualitative patterns in our empirical analyses (in particular the ordering of edges by synergy fraction) were consistent across all five measures.

Computing $I_{\pm}$ and $I_{\wedge}$ for five or more sources via conventional lattice enumeration is computationally prohibitive. The PID redundancy lattice, the hierarchy that organises all non-trivial ways source subsets can overlap in their information about a target, contains 7,579 antichains (the lattice nodes that index distinct PID atoms) at $N = 5$ sources. We bypass this bottleneck using the fast Möbius transform (FMT) [191], which precomputes the lattice structure and obtains each target's PID atoms via a single matrix–vector multiplication, reducing per-target computation from hours to minutes.

Embedding the remaining sources for symptom networks

In principle, PID can be extended to more than two sources, but the number of information atoms grows super-exponentially with the number of sources (from 4 atoms at two sources to over 7,500 at five; Section 6.2.5), quickly becoming computationally and statistically intractable. Two factors compound this. First, the decomposition is not computed once but re-estimated for every directed source–target pair, in each dataset, across the five redundancy measures, and across the calibrated synthetic ensemble used for validation, so the nominal per-target cost is incurred many times over. Second, stable estimation of a high-order joint distribution requires adequate counts in each joint cell: the eight non-target symptoms together with the target (nine three-level variables) span $3^9 \approx 2 \times 10^4$ joint states, leaving on the order of one observation per state in the Xinxiang student sample ($N \approx 24{,}000$) and about eight in UK Biobank ($N \approx 154{,}000$), far too sparse for reliable higher-order estimates. To approximate the high-order structure while retaining interpretability, we adopt an embedding-based strategy. Intuitively, instead of estimating the full joint distribution of all remaining symptoms, we replace them with a single compact summary variable—a learned grouping of the possible remainder configurations that keeps what is informative about the target and discards the rest—so the decomposition need only handle the focal symptom together with this summary. Throughout this section and in all empirical analyses, the two-source decomposition uses the MMI redundancy measure I_{mmi} defined in Section 6.2.5. The four alternative measures introduced there are reported as robustness checks and as a benchmark axis in

the multi-dataset evaluation, whose results are in Appendix E.5.1.

For each ordered symptom pair (X_i, X_k) with $i \neq k$, we treat X_k as the target and X_i as a focal source. The remaining variables $\mathbf{X} \setminus \{X_i, X_k\}$ are compressed into a single discrete embedding

$$E_{i \to k} = f_{i \to k}(\mathbf{X} \setminus \{X_i, X_k\}).$$

This embedding provides a compact proxy for the joint configuration of the remaining symptom network when paired with the focal source in a two-source PID.

We consider multiple constructions of $f_{i \to k}$, including (i) *target-directed* embeddings that learn $f_{i \to k}$ using X_k and (ii) *unsupervised* embeddings that compress $\mathbf{X} \setminus \{X_i, X_k\}$ without access to X_k. All candidate embeddings are evaluated in a simulation benchmark against tractable multivariate ground truth (Section 6.2.5).

Our primary embedding, which we term the Agglomerative Conditional Information Bottleneck (ACIB), groups the joint states of the remainder symptoms into a small number of clusters and then iteratively merges the most similar clusters, choosing each merge so that the conditional information that the remainder carries about the target (given the focal source) is preserved as much as possible. The result is a low-cardinality discrete embedding $E_{i \to k}$ that summarises the rest of the symptom network without discarding the part of its dependence with the target that the focal source does not already explain.

Formally, ACIB adapts conditional information-bottleneck clustering [192] with agglomerative merging [193] to the PID setting. It operates in two phases: first, joint states of the remainder are assigned to clusters by minimising a conditional Kullback–Leibler-divergence distortion; second, the closest cluster pair is greedily merged while constraining the relative loss in conditional mutual information, $1 - I(X_i; X_k \mid E_{i \to k}) / I(X_i; X_k \mid Z_{i \to k})$, to remain below a specified tolerance. Because this constraint operates on conditional rather than unconditional mutual information, the compression preserves the source–target relationship that PID subsequently decomposes.

ACIB is fitted separately for each (X_i, X_k) pair, using $\mathbf{X} \setminus \{X_i, X_k\}$ as inputs. It has three hyperparameters: the conditional-information loss tolerance, the cardinality cap $K_{\max}$, and a Dirichlet smoothing constant, which we set to 5% relative information loss, $K_{\max} = 12$, and 0.5 throughout. A dedicated sensitivity analysis (Appendix E.11) shows that fidelity is essentially insensitive to the loss tolerance and the smoothing constant and that $K_{\max} = 12$ is statistically indistinguishable from neighbouring values; ACIB was selected over alternative embeddings for its superior fidelity at $N \geq 4$ sources (Section 6.2.5).

Given the embedded system $(X_i, E_{i \to k}; X_k)$, we compute a two-source PID using the MMI measure described above. The resulting atoms,

$$\mathrm{Unq}_{i \to k}, \quad \mathrm{Unq}_{E_{i \to k}}, \quad \mathrm{Red}_{i \to k}, \quad \mathrm{Syn}_{i \to k},$$

are interpreted as follows for each target X_k:

- $\mathrm{Unq}_{i \to k}$: information about X_k that is carried uniquely by symptom X_i beyond what is available from the embedded remainder $E_{i \to k}$;

- $\text{Unq}_{E_{i\to k}}$: information about X_k that is carried uniquely by the rest of the network (via $E_{i\to k}$) and not by X_i;
- $\text{Red}_{i\to k}$: information about X_k that is shared between X_i and the embedded remainder;
- $\text{Syn}_{i\to k}$: information about X_k that becomes available only when X_i and the embedded remainder are considered together.

These atoms provide an approximate decomposition of the information that all other symptoms collectively carry about X_k into components attributable uniquely to X_i, uniquely to the remainder, shared, and synergistic channels. Because the embedding is many-to-one, the decomposition is approximate rather than exact for the original multivariate system; its accuracy is characterised in the simulation study below.

In ePID, the remainder-unique atom $\text{Unq}_{E_{i\to k}}$ quantifies predictability of the target from the remainder of the symptom system alone. In contrast, the redundancy and synergy atoms necessarily involve the focal source X_i and therefore quantify *source involvement* in the source→target dependence. That is, $\text{Unq}_{E_{i\to k}}$ quantifies *remainder-only predictability* of X_k, $\text{Red}_{i\to k}$ quantifies *overlap* between the focal symptom and the remainder, and $\text{Syn}_{i\to k}$ quantifies *interaction-only information* that is available only from their joint configuration.

Before computing the two-source PID, each of the three variables (focal source, embedding, and target) is capped to at most five categories: the four most frequent values are retained and all remaining, rarer values are pooled into a single "other" category. Unlike the adjacency-preserving PHQ-9 merge of Section 6.2.2, this is a purely frequency-based cap that does not preserve ordinal adjacency; for the three-level PHQ-9 items it is therefore a no-op on the source and target and binds only on the embedding (a nominal cluster label, for which order is irrelevant), whereas for higher-cardinality instruments such as the IRI it can pool non-adjacent ordinal levels. This cardinality cap ensures that the joint distribution $(X_i, E_{i\to k}, X_k)$ remains well-sampled at typical clinical sample sizes.

Simulation benchmark for embedding validation

Because the embedding-based PID is approximate and the ground-truth multivariate PID cannot be estimated from observational data at scale, we validate against calibrated synthetic systems where exact computation is tractable. This benchmark serves two purposes: (i) quantifying how well ePID recovers network-context unique, redundant, and synergistic structure, and (ii) selecting an embedding method that is accurate under empirically realistic marginal distributions. We generate an ensemble of Bayesian networks calibrated to PHQ-9 data and compare the embedding-derived two-source PIDs to the ground-truth multivariate PID.

Bayesian network ensemble. Using the UK Biobank PHQ-9 item distributions as a reference, we learn a directed acyclic graph over nine discrete variables for benchmarking

purposes. We denote each synthetic Bayesian network by G and write (G, X_t, N) for the configuration in which target X_t and number of sources N are evaluated on network G. The empirical analyses exclude suicidal ideation due to sparsity (only 4.3% and 7.6% of respondents endorsed any thoughts of self-harm at severity ≥ 1 in UK Biobank and the Xinxiang student sample respectively), but including it in the synthetic calibration provides a stress test in the presence of low-prevalence states. Conditional probability tables are estimated from the empirical data and then perturbed with small Dirichlet noise to create 50 distinct Bayesian networks. For each network we draw 600 000 samples by forward sampling. This yields an ensemble of synthetic datasets that preserve the marginal distributions and broad dependency patterns of the PHQ-9 items while allowing exact computation of information-theoretic quantities. Calibration diagnostics confirm that the synthetic ensembles closely match the empirical mutual information structure (mean MI error 6.10%; Kolmogorov–Smirnov statistic 0.166; see Figure E.2).

Ground-truth multivariate PID. For each Bayesian network G, each target X_t (one of the nine variables), and each number of sources $N \in \{3, 4, 5\}$, we select N distinct source variables from the remaining items and compute the full N-source PID of $(X_i, \ldots, X_N; X_t)$ using the MMI PID measure. This yields a redundancy lattice whose atoms specify unique, redundant, and synergistic contributions for all combinations of sources.

Embedding-based PID on the same systems. For the same (G, X_t, N) configurations, we approximate the multivariate PID using two-source decompositions as follows. For each choice of focal source X_i among the N sources, we embed the remaining $N-1$ sources $\{X_j : j \neq i\}$ into a one-dimensional discrete variable E using each of 13 embedding methods spanning three families:

- *Manifold and dimensionality reduction.* ACIB, the agglomerative conditional information bottleneck described above [192, 193]; multiple correspondence analysis followed by k-means clustering of factor scores [194, 195]; non-negative matrix factorisation followed by k-means [196]; sliced inverse regression with quantile binning, which estimates a target-supervised low-dimensional projection [197, 198]; partial least squares with quantile binning, which extracts components maximising covariance with the target [199]; and truncated singular value decomposition followed by k-means [200].

- *Feature selection.* Conditional mutual information maximisation (CMIM), which selects features that maximise CMI with the target conditional on previously chosen features [201]; joint mutual information (JMI), which selects features by their average pairwise joint information with the target [202]; a greedy CMI-based selector that adds variables one at a time to maximise the running conditional mutual information [203]; and ReliefF, which scores features by their ability to distinguish nearest hits from nearest misses across classes [204, 205].

- *Clustering.* Agglomerative clustering with Hamming distance over the joint state vector [206]; k-modes, the categorical analogue of k-means using mode-based centroid updates [207]; and spectral clustering with a Hamming-distance-based affinity matrix [208].

We then compute a two-source PID for $(X_i, E; X_t)$ using the MMI measure.

Amalgamation of multivariate PID. To compare the two-source PID on $(X_i, E; X_t)$ with the N-source PID on $(X_1, \ldots, X_N; X_t)$, we collapse the many atoms of the N-source decomposition into four two-source atoms via a structural amalgamation that groups the $N-1$ non-focal sources into a single composite source, so that the projected atoms answer the same question as ePID: how much information about X_t is attributable to the focal source X_i beyond what the remainder provides. The mapping is conservative with respect to synergy, classifying an atom as synergistic only when it necessarily requires contributions spanning both the focal source and the remainder group, so the amalgamated synergy is a lower bound on the true interaction-only information between X_i and the remainder. The grouping preserves the two-source consistency relations. The full antichain classification rule and its derivation are given in Appendix E.2.6. Figure 6.3 visualises this amalgamation step on the toy structural causal model used in the methods-comparison figure in this chapter.

Error metrics. For each PID atom $c \in \{\mathrm{Unq}_{X_i}, \mathrm{Unq}_E, \mathrm{Red}, \mathrm{Syn}\}$ and each embedding method, we compute the absolute error

$$\mathrm{AE}_c = \left| \mathrm{PID}_c^{(N\text{-source})} - \mathrm{PID}_c^{(\mathrm{embed})} \right|, \tag{6.4}$$

where the first term is the amalgamated ground-truth value and the second term is the corresponding atom from the two-source PID on $(X_i, E; X_t)$. In words, the absolute error captures the per-atom gap between embedding-based and ground-truth values; smaller values indicate that the embedding more faithfully reproduces the multivariate decomposition for that atom. We summarise errors across all (G, X_t, N) configurations and focal sources. Figure 6.4 displays the mean absolute error by embedding method and PID atom; Appendix Figure E.8 compares approximated versus ground-truth synergy for one representative method from each family tier (ACIB, JMI, k-modes); and Appendix Figure E.9 shows how synergy error scales with the number of sources.

For the multi-dataset benchmark (Appendix E.5.1), we additionally report the total absolute error $\mathrm{TAE} = \sum_c \mathrm{AE}_c$ summed across all four PID atoms, and the relative error $\mathrm{RE} = \mathrm{TAE} / I(X_i, E; X_t)$, where the denominator is the ground-truth total mutual information that the focal source and its embedded remainder jointly carry about the target (equivalently, the sum of the four ground-truth atoms). RE expresses the approximation error as a fraction of the total information being decomposed, enabling comparison across datasets with different baseline information levels.

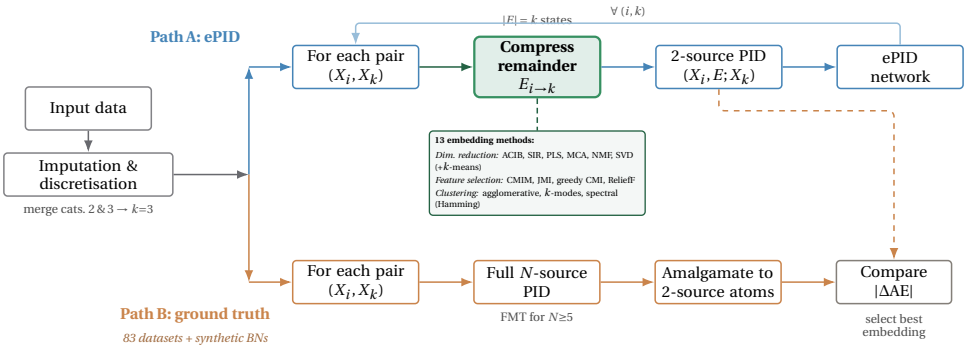


Figure 6.2: **Overview of the ePID evaluation pipeline.** An end-to-end view of how raw symptom data become a PID-decomposed network: a scalable embedding-based path used for the main analyses runs in parallel with a small-scale ground-truth path used to validate it. Raw data are preprocessed (imputation, discretisation) and then processed along these two paths. *Path A (top, blue):* for each source–target pair, the remainder symptoms are compressed into a discrete embedding using one of 13 methods, and a two-source PID is computed on (source, embedding, target). *Path B (bottom, orange):* for benchmark validation, the full N-source PID is computed via the fast Möbius transform (FMT) and amalgamated to two-source atoms for comparison with Path A.

The simulation study serves two purposes. First, it establishes that embedding-based PID can approximate the multivariate decomposition with low error when suitable embeddings are used. Second, it provides empirical guidance for choosing embeddings in the empirical PHQ-9 analyses, where the ground-truth PID is unknown.

Applicability beyond PHQ-9. The ePID pipeline is designed to accommodate datasets with varying numbers of variables, response cardinalities, and sample sizes. For variables with many categories, the general pipeline supports cardinality capping that merges infrequent categories to reduce the dimensionality of joint distributions before embedding and PID computation. This step is exercised in the multi-dataset benchmark (Section 6.2.6), where datasets vary in cardinality. For the PHQ-9 items analysed here, the three-level ordinal scale after category merging is already low-cardinality, making this additional step unnecessary. Computational details for the embedding benchmark, including the subsampling strategy for distance-based methods, are provided in Appendix E.2.6. Figure 6.2 provides an overview of the full ePID evaluation pipeline, including both the embedding-based path and the ground-truth benchmark path.

6.2.6. Multi-dataset embedding benchmark

The synthetic Bayesian network benchmark (Section 6.2.5) provides controlled validation with known ground truth but is limited to a single data-generating mechanism calibrated to PHQ-9. To assess generalisability, we additionally evaluated embedding fidelity across 83 real-world datasets spanning clinical, epidemiological, and psychometric domains.

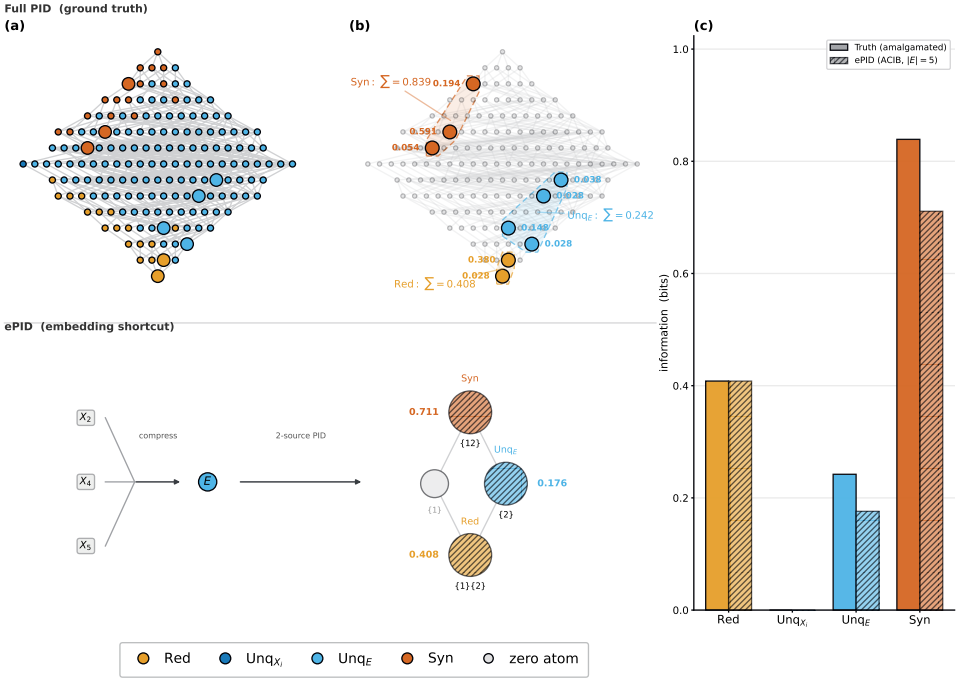


Figure 6.3: **ePID on the toy SCM: redundancy lattice, amalgamated atoms, truth-versus-ePID recovery, and the embedding shortcut.** Same SCM as Figure 6.1; visualises the four-source PID lattice that the main pipeline amalgamates over. **(a)** Full $N = 4$ antichain lattice for $\{X_1, X_2, X_4, X_5\} \to X_3$ under $I_{\min}$: 166 atoms in total, of which 9 carry non-zero information (sum = 1.489 bits). Node area scales with atom value; colour encodes the type-1 amalgamation class. **(b)** The nine non-zero atoms grouped by their type-1 amalgamation class (synergy $\Sigma = 0.839$, remainder-unique $\Sigma = 0.242$, redundancy $\Sigma = 0.408$ bits), each labelled with its individual value. **(c)** Amalgamated four-atom view for focal source X_1: ground truth versus ePID (ACIB embedding, $|E| = 5$); the arrow marks the amalgamation step that collapses the lattice into four atoms. ACIB recovers Red exactly, Syn at 85%, and Unq$_E$ at 73%, for 87% of total joint information. The lower row makes the ePID shortcut explicit: rather than enumerating the full lattice, the remaining sources $\{X_2, X_4, X_5\}$ are compressed into a single embedding E and a two-source PID is computed on $(X_1, E; X_3)$, recovering the amalgamated atoms directly (synergy 0.711, remainder-unique 0.176, redundancy 0.408 bits). Grey nodes carry zero information. ACIB = Agglomerative Conditional Information Bottleneck.

Dataset assembly and preprocessing. We assembled 83 datasets from publicly available repositories spanning clinical, epidemiological, and psychometric domains, with 5–56 variables per dataset and sample sizes ranging from 32 to 445,000 observations. Full details of dataset assembly, preprocessing, imputation, and discretisation are provided in Appendix E.1.1. In total, the benchmark comprises 2,477,784 individual comparisons across 5 PID measures, 83 datasets, 13 embedding methods, and $N \in \{3, 4, 5\}$ sources.

Embedding hyperparameters. The 13 embedding methods were configured as follows: ACIB used a loss tolerance of 5% relative information loss with maximum embedding cardinality $K_{\max} = 12$; clustering methods (k-modes, spectral Hamming, agglomerative Hamming) used $K = 6$ clusters; SIR and PLS used 2 latent dimensions with $B = 4$ bins; MCA, SVD, and NMF used 2 components with $K = 6$ k-means clusters; ReliefF used 100 neighbours with $k = 3$ selected features; JMI and CMIM selected $k = 3$ features. For distance-based methods (spectral and agglomerative Hamming), datasets with more than 5,000 observations were subsampled to 5,000 rows for distance matrix computation. All methods used a fixed random seed of 42 for reproducibility.

Error metrics. For each embedding method and dataset, we computed the ground-truth N-source PID via the fast Möbius transform [191] and amalgamated it to a bivariate decomposition using the conservative amalgamation scheme described in Section 6.2.5. Embedding error was quantified using the total absolute error (TAE) and relative error (RE) defined in Section 6.2.5.

6.2.7. Multi-measure robustness framework

PID decompositions are measure-dependent: the choice of redundancy function determines how atoms are partitioned. To assess the robustness of embedding fidelity across different PID axiomatisations, we computed ground-truth N-source PIDs using the five redundancy measures introduced in Section 6.2.5 (I_{mmi}, I_{min}, I_{rr}, $I_\wedge$, $I_\pm$); we adopt I_{mmi} as the primary measure throughout and treat the remaining four as benchmark axes here.

For $N \geq 5$ sources, $I_\pm$ and $I_\wedge$ were computed via the fast Möbius transform [191] using precomputed antichains and Möbius matrices for the redundancy lattice, reducing per-target computation from hours to minutes.

For each dataset, target, number of sources $N \in \{3,4,5\}$, and PID measure, we computed the full N-source ground-truth PID, amalgamated the resulting atoms onto two-source equivalents using the conservative scheme described in Section 6.2.5, and compared these against the corresponding embedding-based two-source PID. This yields error profiles stratified by embedding method, PID measure, and number of sources, enabling a systematic assessment of which embedding methods are robust across different axiomatisations of redundancy and synergy.

Dataset coverage. Coverage was identical across all five measures: complete ground-truth PIDs were available for 83 datasets at $N = 3$, 79 at $N = 4$, and 77 at $N = 5$, the shortfall at $N = 4$ arising from four NHANES datasets that lack sufficient variables.

6.2.8. Normalisation and effect-size scaling

Because information-theoretic measures are expressed in bits and are bounded above by the uncertainty of the target, we report normalised effect sizes—a target-normalised CMI (nCMI) and entropy-share PID atoms—to facilitate comparisons across targets and

cohorts; the definitions, equations, and the network-visualisation pruning rule are given in Appendix E.2.7.

6.2.9. Significance testing for pairwise measures

Statistical significance of pairwise conditional-dependence edges (PC and CMI) was assessed using permutation tests. For each source–target pair (X_i, X_j), the target column X_j was permuted across participants while all other columns (including X_i and the conditioning set Z_{ij}) were held fixed. This preserves the joint distribution of the non-target variables while destroying the focal source–target association. The pairwise statistic was recomputed on each of $B = 1{,}000$ permuted datasets to form an empirical null distribution. The one-sided p-value was computed as $p = (m+1)/(B+1)$, where m is the number of replicates with a test statistic at least as large as the observed value. Multiple testing across all edges was controlled using the Benjamini–Hochberg false discovery rate (FDR) procedure at $q = 0.05$ [209].

Cross-cohort stability. To assess replicability of the PID decomposition, we match directed source–target edges that are significant in both UK Biobank and the Xinxiang student sample and compute Spearman rank correlations of PID atom fractions (redundancy, synergy, remainder-unique) across matching edges. Mean absolute differences between dataset-specific fractions are also reported.

6.3. Results

We present results in four stages. First, we benchmark candidate embeddings against ground-truth PID on synthetic Bayesian networks calibrated to PHQ-9. Second, we extend that benchmark across 83 real-world datasets and five PID measures to assess generalisability. Third, we apply the validated pipeline to the empirical PHQ-9 networks in UK Biobank and the Xinxiang student sample, comparing pairwise conditional-dependence estimates from partial correlations and CMI before decomposing each source–target dependence via ePID into redundancy, synergy, and remainder-unique channels. Fourth, we apply the same pipeline to the Interpersonal Reactivity Index as a contrasting psychometric instrument.

6.3.1. Embedding benchmark for ePID

The central question for the benchmark is whether this compression preserves the resulting PID atoms. This subsection serves to evaluate the quality of preserving PID atoms for different embedding algorithms and determine the optimal embedding algorithm for all subsequent analyses.

To determine which embedding methods provide reliable approximations to high-order PID, we evaluated 13 candidates on an ensemble of 50 Bayesian networks $\{G_1, \ldots, G_{50}\}$ calibrated to the joint distribution of the PHQ-9 items. BN calibration statistics are shown in Figures E.2, E.3. For each network G, each target X_t, and each num-

ber of sources $N \in \{3,4,5\}$, we computed (i) the exact N-source PID using the MMI-based redundancy measure in `dit`, and (ii) a family of two-source PIDs where one source was a focal variable and the other was a discrete embedding of the remaining $N-1$ sources. We then amalgamated the ground-truth N-source atoms onto the four components of a two-source PID and quantified the absolute error for each atom (source-unique, remainder-unique, redundancy, synergy) and each embedding method (Methods 6.2.5). This synthetic benchmark uses MMI (minimum mutual information) throughout, both for the ground-truth N-source decomposition and for the embedding-based approximation, so comparisons are internally consistent. Because these results characterise approximation fidelity conditional on a given PID measure, they do not speak to the separate question of which measure best captures the information structure of interest; we return to measure dependence in the Discussion (Section 6.4.4). Cross-measure generalisation, comparing each of five PID measures against its own ground truth, is evaluated separately in Appendix E.5.1.

Method-family stratification Errors stratified strongly by embedding family (Figure 6.4). Manifold and dimensionality-reduction methods (ACIB, SIR, PLS, MCA+k-means, SVD+k-means, NMF+k-means) achieved the lowest overall mean absolute error across components (≈ 0.014 bits), ahead of information-theoretic feature selection approaches (CMIM, JMI, greedy CMI, ReliefF; mean error ≈ 0.018 bits) and clustering methods (agglomerative Hamming, spectral Hamming, k-modes; mean error ≈ 0.022 bits). Among all methods, the Agglomerative Conditional Information Bottleneck (ACIB) embedding showed the highest overall fidelity. At $N = 5$, ACIB achieved a mean total absolute error (summed across all four PID atoms) of 0.048 bits, or about 19% of the total mutual information (TMI); synergy-specific errors remained small in both mean (0.0095 bits) and median (0.008 bits).

Component-resolved performance. Most methods reproduced the source-unique and redundancy atoms with near-numerical precision (median errors $< 10^{-10}$ bits for 11 of 13 methods), whereas discrepancies concentrated in the remainder-unique and synergy atoms (bottom panels of Figure 6.4). A relative-error view of the same data, in which per-atom errors are normalised by the magnitude of the corresponding ground-truth atom, is reported in Appendix Figure E.6. For these challenging atoms, ACIB again performed best: the median error for the remainder-unique atom was 0.0163 bits and the median synergy error was 0.0058 bits, whereas typical feature-selection methods exhibited larger median synergy errors. We report these errors as fractions of TMI, the exact ground-truth total mutual information for each edge (the sum of its four ground-truth atoms), rather than as fractions of each atom's own true magnitude, because the synergy and remainder-unique atoms are themselves near-zero in many of the calibrated networks: a per-atom denominator is then numerically unstable and can inflate relative errors without bound, whereas TMI provides a stable common scale across atoms. The complementary per-atom relative-error view is the convention used in Appendix Figure E.6, where exactly this

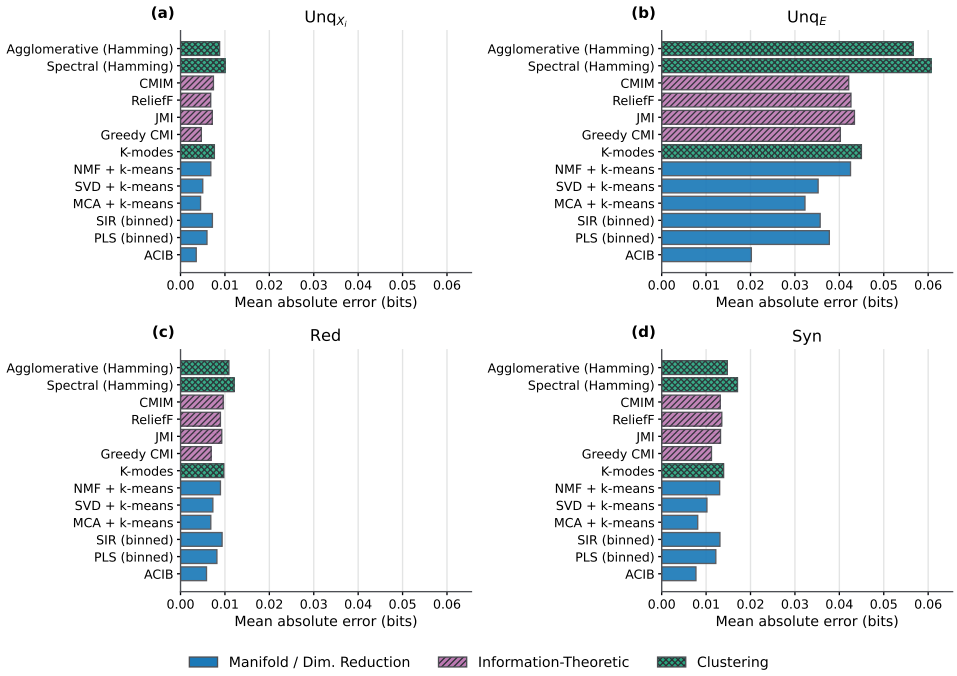


Figure 6.4: **Mean absolute error by embedding method and PID atom.** Each panel shows mean absolute error (in bits) between the amalgamated ground-truth N-source PID and the embedding-based two-source PID approximation, for one of the four PID atoms: **(a)** source-unique Unq$_{X_i}$, **(b)** remainder-unique Unq$_E$, **(c)** redundancy Red, **(d)** synergy Syn. Methods are coloured by family: *manifold / dim-reduction* (blue: ACIB, SIR, PLS, MCA+k-means, SVD+k-means, NMF+k-means), *information-theoretic feature selection* (pink: CMIM, JMI, ReliefF, Greedy CMI), and *clustering* (green: Agglomerative Hamming, Spectral Hamming, k-modes). Errors are pooled across 50 synthetic Bayesian networks calibrated to PHQ-9, nine targets, and three source counts ($N \in \{3,4,5\}$); within each panel, methods are sorted by mean error. Within the manifold family, ACIB (the recommended default) achieves the lowest error across all four atoms. The clustering family is heterogeneous: k-modes performs comparably to manifold methods on Red and Unq$_{X_i}$, while Hamming-based clustering collapses Syn toward zero (see also Figure E.4). ACIB = Agglomerative Conditional Information Bottleneck; JMI = Joint Mutual Information; CMIM = Conditional Mutual Information Maximisation.

instability is visible for the near-zero atoms. This pattern indicates that the main difficulty in approximating PID lies in preserving higher-order and source-specific information; embeddings that explicitly optimise information preservation fare better than those that optimise only geometric cohesion or pairwise criteria.

Synergy fidelity across method families. Per-pair comparison of approximated versus ground-truth synergy illustrates the family-level spread (Appendix Figure E.8). For ACIB (manifold/dim-reduction; Pearson $r = 0.92$), synergy estimates aligned closely with the identity line, indicating that the one-dimensional embedding captured the range of synergistic information present in the original multi-source systems. For JMI (feature-selection; $r = 0.76$), estimates were more scattered around the identity line but no longer strongly attenuated. For k-modes (clustering; $r = 0.75$), estimates aligned reasonably well, illustrating the upper end of clustering performance. Under this calibration the Hamming-based agglomerative and spectral clustering methods no longer collapse synergy toward zero (agglomerative $r = 0.68$, spectral $r = 0.62$; see Figure E.4); they recover a moderate share of the synergistic structure, though still less than the manifold methods.

Feature selection methods (CMIM, JMI, greedy CMI, ReliefF) showed a distinct failure mode. For systems with three sources, these methods approximated synergy almost perfectly (errors near machine precision). This is expected: with $N = 3$ sources the remainder consists of $N - 1 = 2$ variables, and feature-selection embeddings that retain up to two features can store both losslessly, so the joint state of the remainder is preserved exactly and the resulting two-source PID coincides with the three-source ground truth. The embedding step is therefore active (two sources are still compressed into a single discrete embedding for the PID), but no information is lost in that compression. When the number of sources increased to five, synergy errors rose to ≈ 0.017 bits, a more modest increase, but still indicating that methods which select or weight features via greedy, pairwise, or univariate criteria capture the emergent joint information less completely as more sources interact simultaneously.

Scaling with the number of sources. We next examined how approximation error changed with the number of sources N. For ACIB, mean synergy error increased smoothly from 0.0046 bits at $N = 3$ to 0.0095 bits at $N = 5$, remaining below 0.01 bits throughout. Because the UK Biobank-calibrated networks carry little synergy per edge, these small absolute errors correspond to a larger fraction of the (small) per-edge synergy: the synergy atom's median per-edge error, normalised by each edge's synergy, rose from about 21% at $N = 3$ to 76% at $N = 5$. We therefore emphasise the absolute error, which stays below 0.01 bits; the relative figure is inflated by the near-zero synergy denominator in these low-synergy networks. A sliced inverse regression (SIR) embedding with binning showed synergy error that increased with N (from ≈ 0.008 bits at $N = 3$ to ≈ 0.016 bits at $N = 5$), comparable to the other manifold methods, whereas clustering methods had both larger baseline errors and steeper increases with N. Thus, ACIB provides the best absolute fidelity, while SIR offers the most stable performance as systems become more complex.

The extended multi-dataset benchmark (Appendix E.5.1) confirms that ACIB's advantage persists at higher N, with nominally the shallowest error growth slope among all methods.

Embedding choice for symptom networks. Based on these benchmarks, we use ACIB as the primary embedding in all empirical symptom-network analyses. The manifold/dimensionality family consistently preserved the structure of unique, redundant, and synergistic information with low error, whereas clustering and feature-selection methods either underestimated synergy or collapsed it toward zero. This chapter therefore focuses on ACIB-based two-source PIDs.

To confirm that embedding-based ePID generalises beyond the PHQ-9-calibrated synthetic networks, we ran an extended benchmark of all 13 embeddings across 83 real-world datasets (clinical, epidemiological, and psychometric; 5–56 variables) and all five PID measures at $N \in \{3,4,5\}$ sources—over 2.4 million atom-level comparisons. The best-performing embedding varied from dataset to dataset, but a single manifold-family method, ACIB, was the most consistent choice, ranking first on 75–84% of datasets at $N = 5$ for $I_{\min}$, I_{mmi}, and $I_{\wedge}$; we adopt it as the default and report the full method-by-dataset comparison in Appendix E.5.1. The one measure no embedding approximates well is the signed $I_{\pm}$, whose pointwise structure does not survive compression.

6.3.2. PHQ-9 symptom networks

Having validated the embedding pipeline on synthetic and real-world data, we apply it to the empirical PHQ-9 networks in UK Biobank and the Xinxiang student sample. We report two complementary views of the same data. Section 6.3.2 characterises pairwise conditional dependence using partial correlations and CMI; Section 6.3.2 then decomposes those dependencies via ePID. Reporting both side by side allows the ePID atoms to be read against the conventional partial-correlation network used elsewhere in the symptom-network literature.

Pairwise conditional dependence: partial correlations vs conditional mutual information We found that PHQ-9 items in both cohorts have largely monotonic conditional dependence at the level detectable by Spearman partial correlation (SPC), while CMI provides a model-free scale in bits for the same conditional-dependence structure. Figure E.12 (Appendix E.8) compares pairwise conditional-dependence estimates derived from partial correlations and conditional mutual information (CMI) in the Xinxiang student sample. Pearson (PPC) and Spearman (SPC) partial correlations were highly similar at the edge level (Fig. E.14), and both aligned closely with CMI. Permutation tests with false discovery rate control ($q = .05$) yielded statistically significant CMI for 56/56 directed source–target pairs in UK Biobank and 54/56 in the Xinxiang student sample (though many of the UK Biobank edges have nCMI well below 0.005, reflecting the high power afforded by $N \approx 154{,}000$; see Appendix Table E.7 for effect-size–pruned counts).

In the Xinxiang student sample the CMI and Spearman partial-correlation networks were comparably dense (54 significant directed CMI edges versus 54 for SPC at $q = .05$),

indicating that monotonic partial correlations and CMI flag a similar set of conditional dependencies at the item level. Normalised CMI values (nCMI = $I(X;Y \mid \mathbf{Z})/H(Y)$) ranged from 0.02 to 0.07 in the Xinxiang student sample and from 0.0003 to 0.06 in UK Biobank. For context, the largest nCMI in the Xinxiang student sample indicates that the source symptom carries conditional information about the target amounting to about 7% of the target's marginal entropy, after accounting for the rest of the network; most values are far smaller. Conditional dependencies at the item level are therefore modest in absolute magnitude.

Network-context decomposition via ePID
We now decompose each conditional dependence into redundancy, synergy, and remainder-unique channels, asking whether this finer-grained view reveals structure that pairwise methods cannot express.

Aggregate atom distributions and cross-cohort stability. The composition of source–target information was similar in UK Biobank and the Xinxiang student sample: redundancy accounted for 51% of total mutual information (TMI) in UK Biobank and 45% in the Xinxiang student sample, synergy for 6% and 8%, and remainder-unique information for 42% and 47%, respectively. Under MMI with a target-optimised embedding the source-unique atom was zero or near-zero for all edges, so the decomposition effectively partitions dependence into three non-trivial channels: redundancy, synergy, and remainder-unique. We report atom fractions as shares of TMI, $I(X_i, E_{i\to k}; X_k)$, so that shares sum to one for each edge; target-entropy-normalised shares (Eq. E.2) are reported alongside in bits when needed. Edge-level ranges were also comparable across datasets: redundancy 11–79% (UK Biobank) and 19–68% (Xinxiang student sample), synergy 0.4–28% and 1–20%, and remainder-unique 20–79% and 13–80%.

The cross-dataset similarity of atom shares was matched by edge-level concordance. Across the 54 directed edges that were significant in both datasets (FDR-corrected $q < .05$), PID atom fractions were positively correlated (Spearman ρ: redundancy = 0.59; synergy = 0.78; remainder-unique = 0.68), with mean absolute differences in atom share of 0.05, 0.02, and 0.06 (i.e. 5, 2, and 6 percentage points of TMI) respectively. The redundancy concordance ($\rho = 0.59$) is more modest than the synergy concordance. Together, these aggregate and edge-level convergences indicate that the relative importance of redundant, synergistic, and remainder-unique channels is a replicable feature of PHQ-9 symptom networks rather than a property of either dataset.

Network visualisation. To make the PID networks interpretable at a glance, we report two complementary visualisations. In the *full* view (Fig. E.13a, Appendix E.8), each directed edge $X \to Y$ is segmented into remainder-unique, redundancy, and synergy; edge thickness encodes the target-normalised total information $I(X, E_{X\to Y}; Y)/H(Y)$. The *no-remainder-unique* view (Fig. E.13b) removes the remainder-unique atom and rescales the remaining segments, so that colour differences directly compare redundant

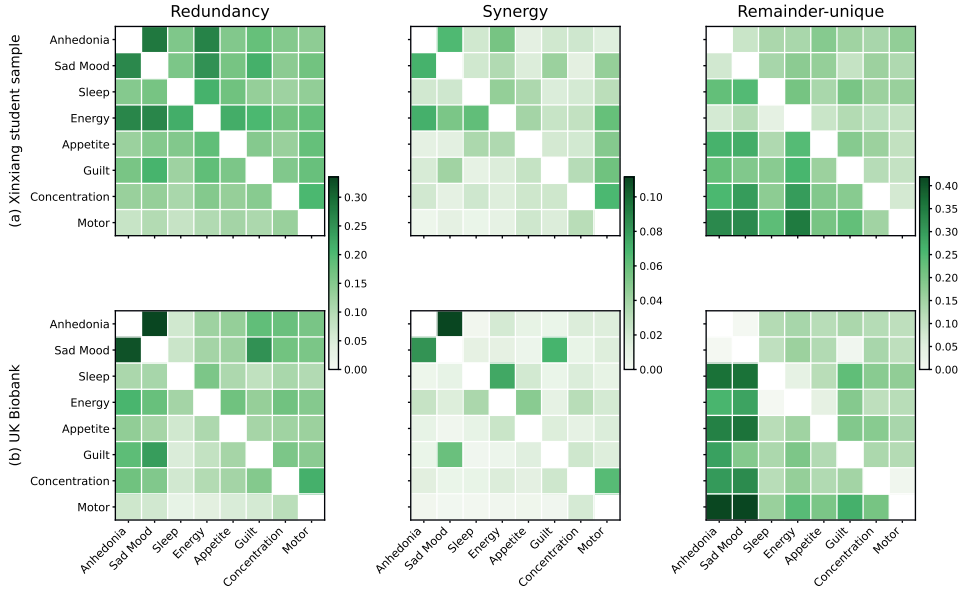
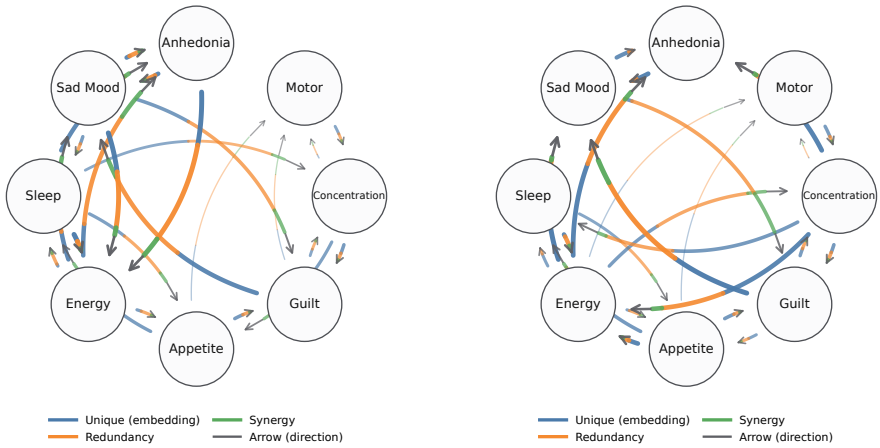


Figure 6.5: **Entropy-normalised PID atom shares across source–target pairs.** Each cell is the atom fraction $I_{\text{atom}}(X \to Y)/H(Y)$ for a directed pair among the eight PHQ-9 symptoms (suicidality excluded); rows are sources, columns are targets, and diagonals are blanked. Top row (a): Xinxiang student sample; bottom row (b): UK Biobank. Columns show the three non-trivial PID channels under ACIB (redundancy, synergy, and remainder-unique) with shared colour scales within each atom to make cross-cohort comparisons visible. The source-unique atom is zero or near-zero across all edges under MMI with a target-optimised embedding (see text above) and is omitted; the conditional-mutual-information share at the edge level is visualised in Fig. 6.6.

versus synergistic source involvement. Read together, these views distinguish targets that are predictable primarily from the broader symptom context from targets where the focal symptom participates through overlap (redundancy) or interaction-only structure (synergy).

Figure 6.6 shows the resulting networks for UK Biobank and the Xinxiang student sample. PID decomposes each retained source–target relation into overlapping versus interaction-only channels. Across retained edges, most dependence is carried by redundant and remainder-unique components, indicating that the apparent influence of a focal symptom on a target typically reflects information already present in the surrounding symptom context. Nonetheless, several edges display substantial synergistic components, consistent with the presence of interaction-dependent configurations in which a symptom becomes informative about a target primarily in conjunction with the remainder of the network.



(a) Xinxiang student sample

(b) UK Biobank

Figure 6.6: **Replication of ePID networks (top-3 incoming edges per target) in UK Biobank and the Xinxiang student sample.** Nodes are eight PHQ-9 symptoms (suicidality excluded). Edges were first screened using permutation-tested conditional mutual information (CMI) with Benjamini–Hochberg FDR control ($q = .05$) and then pruned for visualisation by retaining, for each target symptom Y, the three incoming edges $X \to Y$ with the largest entropy-normalised conditional mutual information $\mathrm{nCMI}_{X\to Y} = I(X;Y \mid \mathbf{Z})/H(Y)$. For each retained directed edge, we compute an embedding-based two-source partial information decomposition (PID) on $(X, E_{X\to Y}; Y)$ using an ACIB embedding of the remaining symptoms. Edge thickness is proportional to total mutual information $I(X, E_{X\to Y}; Y)$. Coloured segments indicate the fraction of this total carried uniquely by the embedded remainder (blue), redundantly by X and the remainder (orange), or synergistically (green); a terminal black segment indicates direction. Edges were selected for display using CMI; decompositions for all 56 directed pairs are reported in Appendix E.8. Because nCMI conditions on the full remaining symptom set, it captures the source's unique-plus-synergistic information and excludes redundancy; since the source-unique atom is near zero throughout these networks, the top-three-per-target display, if anything, emphasises synergistic edges rather than hiding them. The reported atom fractions (Fig. 6.5) are in any case computed from all 56 directed decompositions and all informative edges, not only the displayed subset.

6.3.3. Depression-status contrast in PHQ-9 network composition

To test whether the compositional balance of the PHQ-9 network depends on symptom severity, we split the UK Biobank sample by the standard PHQ-9 cutoff (total score ≥ 10 versus < 10) and estimated the ePID network separately in each group. To remove sample size as a confound, the two arms were equal-N matched at $n = 8{,}879$ per group and each decomposition was averaged over five random subsamples of the larger (non-depressed) group. Relative to non-depressed respondents, the depressed network shifts away from redundancy toward unique and synergistic information (mean shares across the displayed top-three-per-target edges: redundancy 0.34 → 0.15, unique 0.47 → 0.59, synergy 0.19 → 0.26; Fig. 6.7), the clearest signal being the contraction of redundancy. Because the absolute synergy share is sample-size dependent (the full-cohort PHQ-9

synergy share is ≈ 0.08), these values should be read as a matched-N *contrast* between depression groups rather than as absolute synergy levels; the same qualitative shift holds over all 56 significant edges (Fig. E.16).

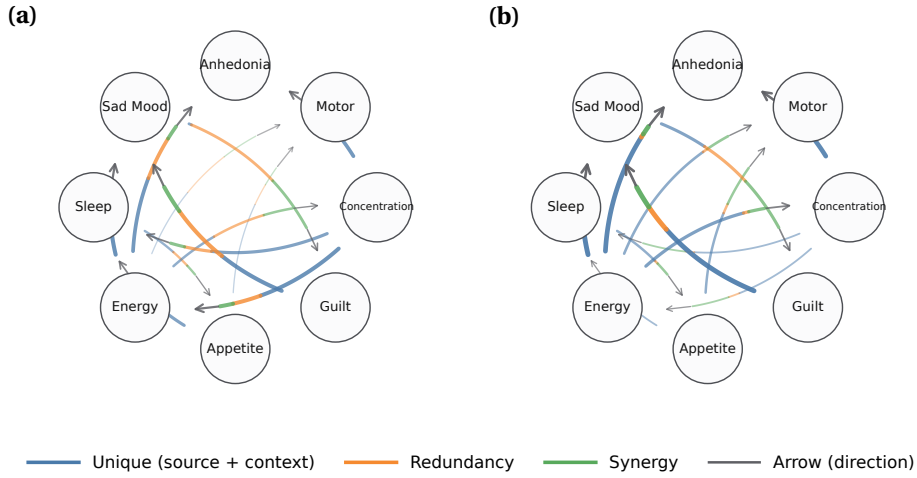


Figure 6.7: **ePID symptom networks in non-depressed versus depressed UK Biobank respondents.** Directed embedding-PID networks for the eight retained PHQ-9 symptoms (suicidality excluded), estimated separately by depression status. **(a)** non-depressed (PHQ-9 total < 10); **(b)** depressed (PHQ-9 total ≥ 10). The groups were equal-N matched at n = 8,879 per arm and each decomposition was averaged over five random subsamples of the larger group. Both panels display the same 24-edge skeleton (the top three incoming edges per target by entropy-normalised CMI, pooled across groups). For each directed edge, coloured segments partition the pair–target total mutual information into unique information carried by the source plus embedded remainder context (blue), redundancy (orange), and synergy (green); a terminal black segment indicates direction, and edge thickness is proportional to total mutual information on a scale shared across both panels. Relative to non-depressed respondents, the depressed network shifts from redundancy toward unique and synergistic information (mean shares: redundancy 0.34 → 0.15, unique 0.47 → 0.59, synergy 0.19 → 0.26); the clearest visual signal is the contraction of the orange (redundancy) segments. Because the absolute synergy share is sample-size dependent (full-cohort PHQ-9 synergy ≈ 0.08), the figure communicates the matched-N *contrast* between groups rather than an absolute synergy level. All 56 significant directed edges are shown in Fig. E.16.

6.3.4. PID and ePID applied to the Interpersonal Reactivity Index

The PHQ-9 atom decomposition (Fig. 6.5) returned a modest synergy share: 6–9% of pair–target information across the cohorts analysed, with no individual edge synergy-dominant. This raises an interpretive question. Does ePID systematically return low synergy, or do PHQ-9 items share substantial overlapping information about a common depressive-symptom construct by design? To distinguish these possibilities we applied the identical pipeline to the Interpersonal Reactivity Index (IRI; Section 6.2.2) [183]. Using protocol parameters matched to the multi-dataset benchmark ($n_{\text{sources}} = 5$ ACIB embedding, I_{mmi} measure), the global 28-item IRI yielded a mean synergy fraction of

23.1%, with 71% (95/133) of informative directed pairs synergy-dominant.

The IRI analyses combine three protocols matched to system size. For within-subscale analyses (7 items per subscale, applied to each of the four subscales separately), we apply ACIB-based ePID at $n_{\text{sources}} = 5$ and pool atom fractions across the four subscales for reporting. For cross-subscale triplet analyses (one item from each of three different subscales as X_1, X_2, target), we report exact two-source PID without embedding, since the triplet system has only two sources and the embedding step is unnecessary. For the cross-subscale triplet analysis, synergy-dominance percentages are computed on the subset of triplets with total mutual information $I(X_i, X_j; X_k) \geq 0.01$ bits, excluding those whose four atoms are individually too small to distinguish synergy and redundancy reliably. For the full 28-item analysis, we report two complementary runs: a primary protocol matched to PHQ-9 ($n_{\text{sources}} = 5$ ACIB-embedded remainder, I_{mmi} measure) supplies the headline atom fractions reported below, and a full-context run (conditioning on all 26 remaining items) supplies the per-pair atom heatmaps and the source–target retention diagnostic in this section. The two full-instrument runs differ in conditioning-set saturation: at $|Z| = 26$ many directed pairs have $I(X_i; X_k \mid \mathbf{Z}) \approx 0$ before any embedding, so the full-context run is reported chiefly as a structural diagnostic rather than as the source of the synergy-dominance claim. Atom fractions across these protocols are comparable in interpretation only when read together as a gradient (Table 6.1).

The atoms reported here use the same two-stage pipeline as the PHQ-9 application: ACIB compresses each remainder context into an embedding with $n_{\text{sources}} = 5$, after which a two-source PID is computed using the I_{mmi} redundancy measure. For the full-context run (conditioning on all 26 remaining items), we restrict the atom-fraction analysis to pairs whose ACIB embedding respected its 5% conditional-mutual-information loss tolerance, retaining 133 of 756 directed source–target pairs.

Configuration	n	Synergy (mean)	Redundancy (mean)	Synergy-dominant pairs
IRI within-subscale ($n_{\text{sources}} = 5$, pooled)	140	15.5%	29.2%	4%
IRI global 28-item	133	23.1%	18.0%	**71%**
IRI cross-subscale triplet (raw 2-source PID)	1,903	21.8%	22.6%	45%

Table 6.1: **Scale-dependent synergy-dominance gradient across IRI configurations.** Synergy fraction increases from within-subscale IRI through cross-subscale triplet IRI to the global 28-item IRI; redundancy decreases in mirror image. Atom fractions are means over informative directed pairs whose ACIB embedding respected its 5% conditional-mutual-information loss tolerance (where applicable); synergy-dominant pairs are those for which the synergy share exceeds the redundancy share.

Atom fractions in Table 6.1 change systematically with analysis scale. Within a single 7-item subscale (e.g. Perspective-Taking analysed in isolation), items are designed to converge on a single empathy facet: redundancy is high (29%) and synergy is modest (16%). At the full 28-item scale, synergy reaches 23% and dominates in 71% of pairs, while redundancy contracts to 18%. When the analysis widens to triplets that cross subscales, redundancy drops to 23% and synergy rises to 22%. This pattern is consistent with the multi-facet construct theory of empathy [183]: within-facet convergent validity manifests

as information-theoretic redundancy, and cross-facet integration manifests as synergy.

Three patterns within the IRI atoms warrant a closer reading. First, among cross-subscale source pairs, the combination of Fantasy and Personal Distress targeting Perspective-Taking items is the most strongly synergy-dominant configuration: 69% of such triplets cross the synergy–redundancy threshold. Mature cognitive empathy is informed synergistically by the affective extremes: neither Fantasy nor Personal Distress alone is strongly informative, but their joint state is. Second, the Fantasy subscale, rather than Davis's cognitive and affective division, organises redundancy. Every source pair containing Fantasy ranked above every pair without it (24.9% versus 20.3% of TMI; difference 4.6 percentage points, 95% CI [2.1, 6.9], cluster bootstrap over target items), although individual pairs within each group were not separable. Perspective-Taking paired with Empathic Concern was the least redundant combination (17.2%; lowest in 94% of replicates). Fantasy items therefore duplicate information the rest of the scale already carries, consistent with long-standing doubts about the subscale's validity [210]; the cognitive and affective core of the instrument contributes the most distinct information. Third, when pooled across all source pairs targeting any Perspective-Taking item, 49% of pairs are synergy-dominant, the highest subscale-level synergy-dominance rate. Mature cognitive empathy emerges as the IRI facet most synergistically informed by the others, consistent with developmental accounts in which cognitive empathy emerges later than its affective counterpart and comes to integrate it [211].

The dominance regime is not an artefact of pipeline choices. The global 28-item mean synergy is stable across $n_{\text{sources}} \in \{3,4,5\}$ (range 22.2–24.2%) and across the twelve embedding methods compared in the multi-dataset benchmark (range 22–28%), indicating that neither the ACIB cardinality nor the supervised-embedding family drives the result.

6.4. Discussion

Symptom networks have become a widely used framework for representing psychopathology, with edges encoding the (conditional) dependence between symptom pairs. Yet a scalar edge cannot express how several symptoms act together to inform a third, and partial information decomposition (PID), the natural tool for recovering that higher-order structure, becomes intractable and hard to interpret beyond a handful of sources. Embedding-based partial information decomposition (ePID) makes network-context partial information decomposition tractable for psychometric instruments: by compressing the multivariate remainder of a symptom network into a target-directed discrete proxy, it decomposes each directed pair's conditional dependence into redundant, synergistic, and unique channels at realistic item counts. In both PHQ-9 datasets, redundancy and remainder-unique channels jointly dominated pair–target information, with a smaller but consistently non-zero synergistic component, and this compositional profile replicated across two large, demographically distinct datasets (synergy fraction cross-dataset $\rho = 0.78$; redundancy $\rho = 0.59$), one a population cohort (UK Biobank) and the other an open online sample (the Xinxiang student sample); the redundancy concordance is

more modest than the synergy concordance. Applied without parameter changes to the Interpersonal Reactivity Index, the same pipeline instead returned a synergy-dominated decomposition, indicating that ePID resolves structural differences between instruments rather than imposing them. That this added structure sits atop a monotonic pairwise backbone is an assumption check rather than a result in its own right: the close agreement of the Pearson and Spearman partial-correlation networks (Fig. E.14), together with the rarity of correlation-invisible non-linear dependencies across datasets (Appendix E.7), shows that ePID's contribution is the compositional layer, not a relaxation of pairwise linearity. After situating ePID against existing approaches, we organise the discussion around three questions: what PHQ-9 redundancy dominance implies for depression-instrument design, what the IRI contrast implies about construct heterogeneity, and what the principal methodological limitations and next steps are.

6.4.1. What ePID adds, and how it relates to existing approaches

PPC, SPC, and CMI quantify whether a symptom pair is conditionally associated, but not whether that association reflects information shared with the broader symptom profile or information expressed only when symptoms are considered jointly. PID converts the scalar edge weight into a compositional profile and so represents this distinction directly: two pairs with identical CMI may differ in whether their information is largely redundant with the surrounding context (suggesting substitutability) or partly synergistic (suggesting interaction-dependent structure). The cross-cohort stability of these fractions supports that the decomposition captures reproducible structure rather than cohort-specific noise.

Triplet-level PIDs provide a useful local lens but cannot attribute information to the remainder of the symptom system (all symptoms except the focal source and target), so information redundant with unmodelled symptoms can be misclassified as unique or synergistic within a restricted triplet. The embedding-based approach fills this gap by including a proxy for the remainder as a source, yielding decompositions aligned with the network question: how a focal symptom behaves *in context* of the full symptom set.

Embedding-based PID provides an information-theoretic analogue of "incremental contribution" that is not restricted to linear models and does not require enumerating all higher-order interactions. The simulation results indicate that manifold and dimensionality-reduction methods, in particular target-directed ones such as ACIB, preserve synergy structure, whereas Hamming-based clustering substantially distorts it (Fig. 6.4).

The stylised symptom triplet (Appendix E.10) illustrates this principle in miniature: when synergy is high, nonlinear models dramatically outperform linear models, and partial correlations fail to detect the underlying interaction. Although the deterministic structure of the example exaggerates the effect relative to empirical data, the implication carries over: edges with high synergy fractions in the PHQ-9 networks may represent dependence patterns that additive clinical models cannot capture, motivating the predictive validation proposed in Section 6.4.4.

A central methodological contribution is the validation of ACIB as a practical em-

bedding for network-context PID. The simulation benchmark demonstrates that this supervised design choice is consequential: target-directed embeddings preserve synergy with high fidelity, whereas similarity-based clustering collapses synergy toward zero. This supervised-embedding step therefore provides a general recipe for extending interaction-aware information decompositions to symptom systems of realistic size.

Recent work has argued that common dimensionality-reduction techniques, particularly variance-preserving methods such as PCA, can preferentially retain redundant structure while failing to preserve synergistic higher-order dependencies [212]. Our embedding-based PID does not treat the symptom system as lying on a low-dimensional manifold and does not optimise a geometric or variance criterion. Instead, we use a target-directed discrete coarse-graining of the remaining symptoms to form a proxy conditioning variable and then quantify unique, redundant, and synergistic contributions in the embedded two-source system. This distinction matters for interpretation: our goal is not to recover a continuous latent geometry of symptoms, but to obtain an interpretable approximation to the distribution of information about a target across a focal symptom and the remainder of the network.

Several complementary frameworks quantify higher-order statistical structure in multivariate systems, and situating ePID relative to these clarifies its specific contribution. Unlike system-level measures such as the O-information (Section 6.2.5), ePID provides per-edge resolution, decomposing each directed pair's dependence into redundant, synergistic, and unique channels. In neuroscience, PID has been applied to characterise information processing in neural circuits, revealing synergistic coding and redundant representations across neural populations [213]. Symptom networks differ from neural data in important respects: fewer variables, ordinal rather than continuous measurement, and clinical constraints on data collection. ePID addresses the scalability challenge differently from neural applications, by embedding the multivariate remainder rather than restricting analysis to small subsets of variables, making it applicable to the moderate-dimensional, discrete systems typical of psychopathological research.

Among psychometric methods specifically, the closest existing approach to higher-order structure is the Moderated Network Model (MNM) framework [179], which warrants direct contrast on three axes. First, MNM is parametric: it extends a mixed graphical model with multiplicative interaction terms and reads edge moderation off estimated coefficients, whereas ePID is model-free and information-theoretic, with synergy and redundancy obtained from joint distributions via PID atoms and no functional-form assumption on the interaction. The trade-off is between interpretable coefficients with standard parametric inference on the one hand, and freedom from any assumed functional form for the interaction on the other. Second, MNM requires the analyst to nominate one or a few candidate moderators and enter them explicitly, which suits confirmatory questions about specific moderators; ePID instead treats the entire remainder of the symptom profile simultaneously as conditioning context via the target-directed embedding, which suits exploratory characterisation of system-wide higher-order structure. Third, the two methods identify different objects: MNM tests *which* edges are moderated by a chosen

variable and by how much, yielding a moderated-edge map, whereas ePID quantifies *how much* of each edge's information is interaction-dependent (synergy), overlapping with the remainder (redundancy), or carried uniquely: a per-edge compositional profile. The two outputs are not interchangeable: a moderated edge under MNM and a synergistic edge under ePID need not coincide. MNM and ePID therefore occupy complementary positions in the modelling-flexibility and pre-specification trade-off, and the choice between them depends on whether the analyst seeks confirmatory tests of specific moderators or exploratory decomposition of system-level higher-order information.

6.4.2. What does PHQ-9 redundancy dominance imply for instrument design?

In both cohorts the decomposition was dominated by redundancy and remainder-unique channels, and this composition is itself informative about how the instrument is built. It is the information-theoretic signature of a well-targeted, largely unidimensional severity scale whose items function as partial substitutes for one another, which aligns with longstanding critiques of depression sum scores assembled from overlapping, content-similar items [177, 214]. What that redundancy looks like in detail, and where synergy nonetheless appears, is the question we now address, beginning with the dominant channels and working towards the synergistic edges that pairwise measures cannot represent.

Redundancy and remainder-unique channels dominate. In both cohorts, redundancy and remainder-unique channels jointly accounted for the vast majority of pair–target information (the Xinxiang student sample: 45% redundancy, 47% remainder-unique; UK Biobank: 51% redundancy, 42% remainder-unique), with synergy contributing 6–9%. This finding should be interpreted as a structural characterisation of the symptom network rather than a limitation of the method. High redundancy between a focal symptom and the remainder means that the information the focal symptom carries about the target is also available through other routes in the symptom network, regardless of the underlying generative model. This is a statement about the information topology of the system: many symptoms share overlapping information channels about a given target. The practical value of redundancy maps is that they identify which edges are "substitutable" (where the focal symptom's contribution to predicting the target is largely duplicated by other symptoms) and which carry unique or interaction-dependent information that would be lost if the focal symptom were removed from the model. This distinction is not available from PC or CMI alone.

Specific synergistic symptom pairs. Synergistic contributions were not uniformly distributed across targets, but the distribution proved reproducible without being attributable to particular symptoms. The edge-level ranking of synergy agreed closely between the two cohorts (Spearman $\rho = 0.78$), and the agreement was insensitive to normalisation ($\rho = 0.76$ with synergy expressed in bits), so the relative distribution of

synergy is a reproducible property of the PHQ-9 network rather than of the chosen scale. Per-symptom rankings, by contrast, were not stable. The target ranked highest on synergy fraction differed between cohorts (psychomotor disturbance in the Xinxiang student sample, energy in UK Biobank), and both rankings reordered substantially when synergy was expressed in bits rather than as a fraction of pair–target information (Fig. 6.5). We therefore report the reproducibility of the synergy profile and refrain from a symptom-specific interpretation, which the present sample sizes and the sensitivity to normalisation do not support. We record this as a negative result, since naming individual synergistic symptom pairs is the use to which a decomposition of this kind would most naturally be put: ePID recovers the composition of dependence reliably at the level of the whole network, but these data do not identify which PHQ-9 pairs carry the synergy. Localisation of that kind will require either uncertainty intervals on individual atoms, which demand participant-level resampling, or instruments whose synergy is concentrated rather than diffuse, as the Interpersonal Reactivity Index proved to be. In contrast, for many pairs the embedded remainder alone captured most of the pairwise information, indicating that these targets are largely determined by the broader symptom context rather than by any single additional source.

Clinical reading. The introduction states that "if two source symptoms are highly redundant, changing the state of only one source would change little to the target's state." The empirical results allow us to revisit this observation concretely. Redundancy-dominated edges suggest that the focal symptom's predictive contribution to the target is largely duplicated by other symptoms; removing it from a predictive model would change little, because the same information is available through other routes. Synergistic edges suggest a different pattern: the information about the target would be lost if either the focal symptom or the remainder configuration were unobserved, because it arises only from their joint state. These are hypothesis-generating insights rather than treatment recommendations, given the cross-sectional design and the observational nature of the data. Nonetheless, they illustrate how the redundancy–synergy distinction can generate qualitatively different clinical hypotheses from the same conditional-dependence backbone.

Whether such hypotheses generalise depends on whether the compositional profile itself is reproducible across samples, an empirical question that connects the within-cohort findings above to the broader symptom-network replicability literature.

The cross-cohort stability of PID atom fractions speaks to the broader replicability debate in symptom network research. Previous work has raised concerns about the replicability of network edge weights and centrality indices across samples [177, 215]. The *composition* of dependence, and not only its magnitude, replicates across two large and demographically distinct datasets. It does not, however, replicate better than the pairwise quantities it decomposes. Over the same 28 symptom pairs, cross-cohort rank agreement was $\rho = 0.82$ for CMI and $\rho = 0.74$ for both partial-correlation variants, against $\rho = 0.76$, 0.59 and 0.59 for the synergy, redundancy and remainder-unique fractions

respectively. The compositional layer should therefore be read as reproducible to a degree comparable with standard edge weights, rather than as a more stable representation of network structure.

6.4.3. What does the IRI contrast imply about construct heterogeneity?

Synergy dominance and the construct-structure hypothesis. The cross-instrument application to the Interpersonal Reactivity Index (Section 6.3.4) demonstrates that the redundancy dominance observed in PHQ-9 is not an artefact of the embedding pipeline. The same ACIB-based ePID, applied without parameter changes, returned a synergy-dominated decomposition on the IRI: 71% of informative directed pairs had higher synergy than redundancy, against 0% for PHQ-9. The monotonic gradient across analysis scales (within-subscale 16% synergy; cross-subscale triplets 22%; full instrument 23%) indicates that the dominance regime is shaped by the construct structure of the instrument rather than by the embedding choice. The same ePID pipeline that returned a redundancy-dominated PHQ-9 recovered a synergy-dominated IRI without parameter changes, indicating that the dominance regime is a property of the instrument rather than of the method. Together with the PHQ-9 anchor, these findings indicate that ePID discriminates between redundancy-dominated and synergy-dominated psychometric instruments using identical machinery: the decomposition captures structural differences in how information is distributed across items, rather than artefacts of the instrument or the embedding choice.

Specific synergy patterns. Three patterns within the IRI atoms warrant a closer reading. First, in the cross-subscale raw 2-source PID (no remainder embedding), the combination of Fantasy and Personal Distress targeting Perspective-Taking items is the most strongly synergy-dominant configuration: 69% of such triplets cross the synergy–redundancy threshold. This suggests that mature cognitive empathy is informed synergistically by affective extremes: neither Fantasy nor Personal Distress alone is strongly informative, but their joint state is. Second, also within the cross-subscale 2-source PID, redundancy is organised by the Fantasy subscale rather than by Davis's cognitive and affective division: pairs containing Fantasy carried reliably more redundant information than pairs without it, while pairs within each group were not separable, and Perspective-Taking with Empathic Concern was the least redundant combination. Fantasy items therefore duplicate information the rest of the scale already carries, consistent with long-standing doubts about the subscale's validity [210]. Third, in the full 28-item ePID ($n_{\text{sources}} = 5$ ACIB embedding, I_{mmi}), 49% of source pairs targeting any Perspective-Taking item are synergy-dominant, the highest subscale-level synergy-dominance rate, consistent with developmental accounts in which cognitive empathy emerges later than its affective counterpart and comes to integrate it [211].

Theoretical implications and limitations. The redundancy–synergy distinction therefore captures structural differences between psychometric instruments, not just between

symptoms within an instrument. Instruments designed around a single underlying construct (PHQ-9 depressive symptoms) yield high redundancy because items are partial substitutes for one another, whereas multi-facet instruments (IRI's four empathy facets) yield high synergy because diagnostic information about a target item often resides in the joint configuration of items from different facets. This suggests an information-theoretic axis for psychometric construct validation that complements conventional reliability and factor-analytic criteria. Because self-report items are imperfectly reliable, and measurement noise attenuates higher-order structure more than linear structure, by analogy with the reliability attenuation of degree-k interactions [216] the reported synergy fractions are best read as a plausible lower bound. Two caveats qualify the IRI-specific findings. The IRI sample is drawn from the Open-Source Psychometrics Project (a self-selected online sample whose composition differs from clinical or community samples), and, in the full-context analysis (conditioning on all 26 remaining items), the atom-fraction analysis is restricted, by design, to the 133 of 756 directed pairs (17.6%) for which the ACIB embedding preserved at least 95% of the source–target conditional mutual information. The 5% loss tolerance is applied because the atoms of a PID are unstable summaries when the embedding has thrown away an appreciable share of the quantity being decomposed; the remaining pairs therefore represent the IRI source–target combinations on which the decomposition can be read with confidence. The IRI patterns should therefore be read as a within-instrument illustration of discriminative validity rather than a population-level inference about empathy structure.

6.4.4. Methodological limitations and next steps

The choice of PID measure is itself a non-trivial modelling choice, since PID has no single canonical measure; we therefore report I_{mmi} as primary with I_{min} as a robustness check, and the qualitative patterns (edge rankings by synergy fraction, cross-cohort stability) held across both. The multi-dataset benchmark across 83 real-world datasets and five PID measures provides a comprehensive assessment of embedding fidelity for this class of methods. Three findings deserve emphasis.

First, ACIB's dominance is robust across four of five measures (I_{mmi}, I_{min}, I_{rr}, $I_{\wedge}$) and across diverse data structures, from small clinical samples to large population surveys. At $N = 5$, ACIB won 84.4% of per-dataset contests for $I_{\wedge}$ [95% CI: 78.4, 91.1], 75–78% for $I_{\text{min}}/I_{\text{mmi}}$, and 64.9% for I_{rr} [56.2, 73.3] (its most modest margin), with the lowest error growth slope among all methods (0.041 [0.035, 0.046] bits per additional source). This consistency suggests that target-directed coarse-graining is a generally effective strategy for preserving the information structure that PID quantifies, not an artefact of the synthetic benchmark's calibration to PHQ-9.

Second, the striking failure of ACIB on the $I_{\pm}$ (pointwise partial information) measure, where ACIB won only 2.6% of contests, has a specific mechanistic origin. Unlike non-negative measures, $I_{\pm}$ produces signed atoms: 77.7% of ground-truth unique-source values were negative at $N = 5$ (compared to 0.7% for I_{min}). The embedding compression disrupts the structured cancellation between positive and negative contributions,

inflating absolute error beyond the total signal (RE > 100%). This finding has practical implications: researchers who wish to use pointwise PID measures should consider alternative embeddings such as ReliefF or SVD+k-means, though errors remain substantial.

Third, at $N = 3$ all methods were statistically indistinguishable (Appendix E.5.1), consistent with the minimal compression required when only two sources are embedded into one variable. Method choice becomes consequential only at $N \geq 4$, where ACIB's information-bottleneck objective provides a clear advantage.

More broadly, the finding that no single embedding method dominates all PID measures cautions against treating any embedding as a universal approximation. The choice of embedding should be aligned with the choice of PID measure, and sensitivity analyses across both dimensions are advisable.

Limitations. Three limitations qualify the interpretation of these findings. First, embedding fidelity has been validated for $N = 3$ to $N = 5$ sources, whereas the empirical setting embeds $N = 6$ remainder symptoms. Computing a ground-truth PID at $N = 6$ is infeasible (the redundancy lattice contains approximately 7.8×10^6 antichains). The sub-linear error growth from $N = 3$ to $N = 5$ and the bounded atom magnitudes ($\leq H(X_k)$) support but do not guarantee extrapolation to $N = 6$; accordingly, the empirical atom fractions should be interpreted as approximate compositional profiles rather than exact quantities.

Second, sampling uncertainty in the atom estimates themselves is not formally quantified. We report point estimates from plug-in (frequency-based) estimators, for which the entropy term is negatively biased in finite samples while the mutual-information functionals built from it are typically biased upward [217]; participant-level bootstrap confidence intervals for PID atoms would be informative but are computationally demanding because each resample requires re-fitting the ACIB embedding for all source–target pairs. The confidence intervals we do report, for contrasts between groups of source–target pairs, are obtained by resampling analysis units (target items or directed edges) rather than participants, and therefore describe variability across pairs within a cohort, not estimation error in any individual atom. They should not be read as sampling intervals on the atom values. Bias is partially mitigated by the large sample sizes ($N > 24{,}000$ for the Xinxiang student sample; $N > 154{,}000$ for UK Biobank) and the low cardinality (three levels per item), but no formal bias correction was applied. Small differences in atom fractions across edges or cohorts should therefore be interpreted cautiously.

Third, three design choices in the empirical pipeline systematically affect atom estimates. We merged the two highest PHQ-9 response categories to reduce sparsity, which may underestimate synergy by collapsing fine-grained joint configurations that distinguish interaction-dependent patterns from additive ones. Separately, the ACIB embedding is optimised to preserve information about the target, which by design ensures the remainder embedding is maximally informative about X_k and could inflate the remainder-unique component relative to a target-agnostic embedding. The decomposition should therefore be interpreted as conditional on these design choices: estimated synergy fractions are likely conservative, and the atoms describe how information about

X_k is distributed between the focal source and a *target-optimised* summary of the remainder. A third design choice concerns the 5% relative-CMI tolerance used by ACIB, which serves both as an internal stopping rule and as a downstream retention filter in the full-context IRI analysis. The 17.6% retention rate (133 of 756 pairs) reflects principally that conditioning on 26 of the 28 IRI items saturates the conditioning set: most directed pairs have $I(X_i; X_k \mid \mathbf{Z})$ at or below numerical zero before any embedding step, and the filter therefore drops mathematically vacuous pairs rather than badly compressed ones. The PHQ-9 application restricts the conditioning set to the five highest-mutual-information sources and so does not encounter this saturation regime (45/56 directed pairs retained); the divergent retention rates between PHQ-9 and IRI reflect the protocol's interaction with instrument size, not a property of the embedding family. The 5% threshold is supported by the sensitivity analysis in Appendix E.11, where the atom estimates are insensitive to the tolerance and to the smoothing prior across the tested grid.

A further interpretive limitation concerns sign. Because the measures used in the main analyses (I_{mmi}, I_{min}, I_{rr}, $I_{\wedge}$) are non-negative, they carry no sign: unlike partial correlation, ePID has no analogue of a negative or inhibitory association and cannot, on its own, distinguish facilitative from suppressive relations. The signed $I_{\pm}$ measure is the only exception, and for the reasons above it is not used in the main analyses.

Directedness does not imply causality. Directed edges in the ePID framework are a computational convention required by PID's source–target formulation, not causal claims. The decomposition is not symmetric: PID($X_i \to X_k$) and PID($X_k \to X_i$) generally yield different atom profiles because the remainder embedding is constructed separately for each target. This asymmetry reflects the information structure of the system (how predictive each symptom is of a given target in the context of the remaining symptoms) rather than causal direction. Cross-sectional observational data cannot distinguish causal influence from confounding, reverse causation, or shared latent causes. The clinical hypotheses described in the introduction (e.g., that insomnia can induce fatigue) motivate the choice of symptom instrument and provide domain context, but they are not tested or validated by the present analyses.

Next steps. Two extensions would most directly strengthen the ePID framework. First, bootstrap or permutation-based confidence intervals for PID atoms would enable formal statistical inference on atom differences across edges, targets, and cohorts, addressing the estimation uncertainty noted in Section 6.4.4. Second, a systematic sensitivity analysis of category merging (varying the number of ordinal levels from 2 to 4) would quantify how discretisation affects atom estimates, particularly synergy. The integration of these symptom-network findings with the multimorbidity results of Chapters 2–5 is taken up in the General Discussion (Chapter 7).

6.5. Conclusion

ePID makes network-context partial information decomposition tractable for psychometric instruments by compressing the multivariate remainder of a symptom network into a target-directed discrete proxy, validated up to $N = 5$ remainder sources. A comprehensive benchmark, spanning calibrated synthetic data and 83 real-world datasets and five PID measures, establishes that ACIB preserves the information structure PID quantifies while identifying measure-specific failure modes, notably for the signed $I_{\pm}$. Applied to two PHQ-9 datasets and to the Interpersonal Reactivity Index, the method shows that decomposing conditional dependence into redundant and synergistic channels adds interpretive value beyond scalar association measures: the composition of dependence replicates across datasets within an instrument, the same pipeline discriminates redundancy-dominated from synergy-dominated instruments, and the redundancy/synergy distinction generates qualitatively different clinical hypotheses about symptom substitutability and interaction dependence. Together, these results support embedding-based PID as a practical complement to standard symptom-network methodology.



Chapter 7

DISCUSSION

Clinical data rarely reduce cleanly to pairs: symptoms, diseases, and functional outcomes are entangled in combinations that pairwise analysis alone cannot represent. Mathematical tools now exist that can encode these higher-order combinations directly, such as hypergraphs, simplicial objects, and multivariate extensions of mutual information [32, 42]. However, the ability to represent a pattern is not the same as knowing what it means. A recurring theme across this thesis is that the practical obstacle lies elsewhere: in the weighting and interpretation logic that turns higher-order structure into something actionable. What does a hyperedge represent, how can it be estimated robustly under cohort-scale constraints, and how can the resulting structure be compressed into something clinically interpretable rather than a catalogue of motifs?

The contribution of this thesis is therefore not that we should "use hypergraphs", but that we should "make higher-order structure measurable and readable". The thesis provided information-theoretic weights of synergy and redundancy relative to explicit targets, together with pipeline procedures (screening, bias control, overlap handling, and structured summarisation), so that higher-order relations could become data-driven and interpretable. At a broader level, the thesis represents a return to symptom-centred reasoning (from symptoms to diagnoses, and now back to symptoms), this time equipped with quantitative tools that distinguish substitutable measurements from context-dependent synergy gates.

The five empirical chapters addressed a linked sequence of questions—*why* higher-order analysis is necessary, *how* to find higher-order structure efficiently, *whether* it matters for clinical endpoints, *how* to summarise it interpretably, and *how* to scale it to network contexts—delivering on the *weighting and interpretation programme* introduced in Chapter 1. The central research question was: *How can information-theoretic notions of higher-order dependence, especially synergy, be operationalised in health data to produce scalable, interpretable, and clinically relevant insights for multimorbidity and symptom networks?* Five sub-questions structure the answer.

Answer to SQ1 Pairwise networks are structurally insufficient. By construction they cannot represent synergistic links, and they can mislead intervention reasoning. In the synthetic model of Chapter 2, pairwise analysis alone produced three disconnected components, whereas including synergistic associations recovered the complete network structure. The same chapter identified 184 statistically significant synergistic triplets in the EASYcare data, demonstrating that the structure is not confined to the synthetic setting. Chapter 2 further argued that symptoms, signs, and disease states belong in one and the same network, because the combinations carrying this information can cross the boundary a disease-only model draws.

Answer to SQ2 Stochastic search made synergy discovery feasible under combinatorial growth (50–80× speed-up over brute-force enumeration). The *echo problem*—identified in Chapter 3 and there mitigated heuristically—established that apparent higher-order structure must be distinguished from redundancy-driven inflation before interpretation begins.

Answer to SQ3 Higher-order analysis demonstrated clinical relevance: symptoms provided unique and synergistic information about daily functioning that disease labels

alone did not capture, suggesting that diagnostic labels may plateau in informativeness under multimorbidity (Chapter 4; 34.1% unique symptom information; κ improvement from 0.208 to 0.459).

Answer to SQ4 Interpretation at scale requires structured summarisation. The BUST (Breadth, Uniformity, Synergy, Total) workflow (Chapter 5) showed that high-synergy and high-total-information features were empirically distinct and surfaced stable cross-domain cliques that mapped onto interpretable informational roles.

Answer to SQ5 PID-style reasoning extended to the symptom-network scale when the multivariate context was compressed into a tractable representation. The ePID framework (Chapter 6) preserved synergy structure under compression (Pearson $r = 0.92$ between approximated and ground-truth synergy). The PHQ-9 decomposition replicated across independent cohorts (synergy $\rho = 0.78$ between UK Biobank and the Xinxiang student sample) and confirmed that redundancy dominated the within-instrument information landscape of depression (45–51%). Applied unchanged to the Interpersonal Reactivity Index, the same pipeline returned a synergy-dominated decomposition, showing that ePID discriminates between redundancy- and synergy-dominated psychometric instruments.

7.1. What synergy and redundancy can (and cannot) tell us

Taken together, the five answers yield three synthesis points. First, the synergy–redundancy lens enables a role-based taxonomy that treats symptoms as informational actors rather than additive indicators. Second, credible higher-order claims require a systematic workflow—screening, bias control, and context-aware estimation—rather than a single metric. Third, synergy analysis is not always worth the additional complexity: its yield depends on the data regime and the research question. Section 7.2 then traces how these findings translate into implications for PID theory, symptom-network methodology, multimorbidity classification, and clinical practice.

The following analysis rests on three clarifications. First, the term “synergy gate,” used throughout this discussion, denotes a statistical property of a pair of variables in a population, not a fixed feature of any patient. It flags a source pair (for example, fatigue and reduced physical activity) that carries joint information about a specified target such as functional decline. Neither variable carries that information alone. Formally, the pair’s PID synergy atom exceeds a threshold, evaluated against a stated target and variable set. Second, observed synergy does not establish a mechanism or causal direction; the appropriate reading is *that this pair of variables jointly carries information about this target in this setting*. Third, the system-level statistic used for initial screening in Chapters 2–3 (O-information) is conservative for synergy detection. Because it summarises the whole system with a single sign, it can report net redundancy even when synergistic triplets are present. Screening on it can therefore miss synergy masked at the system level. These distinctions constrain what the empirical findings below can and cannot imply.

7.1.1. Informational roles for symptoms and clinical features

A first synthesis point is conceptual: this thesis treated symptoms not merely as additive indicators but as informational actors within a coupled system. The synergy–redundancy lens enabled a role-based taxonomy. It assigns each feature a role according to how its target-relevant information is shared with, or unlocked by, the features around it.

In this perspective, some features act as *general indicators*. These are high-redundancy variables whose target-relevant content is widely shared with other measures, often because they encode broad system-level states such as general frailty, systemic inflammation, or diffuse functional limitation. Other features look weak in isolation and become informative only in combination. They enter through the *synergy gates* defined above: pairs or small sets whose target-relevant information is accessible only when the components are considered jointly. Gates differ in how many targets they serve and in how stable they are across subpopulations. A gate that is narrow but stable acts as a *disambiguator*, resolving ambiguity about particular targets for the population as a whole. Gates formalise the clinical intuition that the meaning of a symptom, sign, or measurement can shift sharply depending on what else is present.

The BUST archetypes introduced in Chapter 5 provide an operational bridge between synergy profiles and informational roles. The mapping is not one-to-one. The BUST map positions only pairs that already carry synergy. General indicators are picked out instead by the total-information (T) dimension. A pair high in total information but low in synergy shares much information about the target and carries little of it jointly. Among the gates, broad-uniform (BU) pairs—those whose synergy spans many targets and remains stable across strata—behave as integrative indicators. Narrow-uniform (NU) pairs, with focused but stable synergy, act as disambiguators for specific targets. Broad-specific (BS) and Narrow-specific (NS) pairs, whose synergy varies across demographic or multimorbidity strata, are gates whose relevance is population-dependent.

A concrete example from Chapter 5 illustrates how these categories translate into practice. The pair 'perceived health relative to peers' and 'physical activity' achieved the highest composite BUST score in the cohort. It combined 0.20 bits of synergy strength with positive uniformity ($U = 0.27$) and slightly below-median breadth ($B = -0.17$), both scores median-centred across the 135 source pairs. Neither variable was a strong marginal predictor of most health targets on its own—'perceived health' is a subjective comparator, 'physical activity' a behavioural summary—yet their joint state resolved ambiguity about functional and pain-related targets that neither resolved alone. This profile classified the pair as a *disambiguator*: narrow in scope but stable across age, sex, and multimorbidity strata. By contrast, 'memory complaints' paired with 'memory problems'—a pair that ranked among the highest in total mutual information—exhibited modest synergy, consistent with the high redundancy expected from overlapping cognitive self-reports. The informational-roles lens thus separated pairs that appeared important by different mechanisms: one contributed context-dependent, jointly emergent information, the other contributed broadly shared, substitutable information.

7.1.2. Empirical application requires systematic workflows

A second synthesis point is epistemic: higher-order analysis in health data must be treated as a workflow rather than a single metric [113, 218]. Discovery without screening invites overinterpretation because combinatorial search inevitably produces apparent structure. Interpretation without disciplined discovery becomes either selective (limited by compute) or unmanageable (a long list of candidates).

Five obstacles make this workflow non-optional. Finite-sample bias inflates synergy in sparse samples [113], and the configuration space grows super-exponentially with the number of variables. PID constructions are not unique [40, 44], support is sparse in high-dimensional joint distributions, and health research has no standardised protocol for reporting PID results. We handled each in turn. Bias was corrected by quadratic extrapolation in $1/N$ against a target-shuffle null (Chapter 5), and combinatorial growth by stochastic search with echo filtering (Chapter 3). Sparse support was met by domain-level aggregation and clinically motivated cutpoints (Chapters 4–5), and reporting by the fixed set of quantities the BUST summary records for every pair (Chapter 5). Non-uniqueness needs spelling out.

"PID synergy" is not a single fixed number. Its value depends on the redundancy definition and on the estimator. Chapter 4 used MMI (I_{mmi}), and Chapter 5 used I_{min} [22] as its primary measure, with I_{mmi} and BROJA [106] as sensitivity checks. Chapter 6 returned to I_{mmi}, with I_{min} among its robustness checks. Each choice was made for tractability and interpretability in the setting at hand, and qualitative conclusions were preserved across all three constructions. The O-information [42] used for screening in Chapters 2–3 is a system-level index rather than a decomposition, and its values were not compared with PID atoms. Section 7.3 discusses this robustness in more detail.

The choice of comparator matters as much as the choice of construction. A triplet-level decomposition scores a source pair against one companion variable, not against the rest of the profile. Information that is in fact redundant with unmodelled symptoms can therefore be misclassified as unique or synergistic. Chapter 6 computed its decomposition in network context for that reason: triplet-level PID overestimated synergy relative to network-context PID because it could not attribute information to the remainder of the network. Together with the echo problem (Chapter 3, Section 7.2), this indicated that context-aware workflows are essential to avoid systematic overestimation of synergy.

Applying these safeguards across a combinatorial candidate space begins to resemble an engineering problem. The stochastic search of Chapter 3 and the embedding step of Chapter 6 were both introduced to keep the computation affordable. "Best" is nevertheless not defined by speed or predictive accuracy. Methods should be judged by calibration and stability across cohorts and strata, by reproduction of known structures in simulation benchmarks, and by whether outputs remain interpretable under realistic uncertainty.

The yield of this workflow depends on the type of dataset. In the depression symptom network of Chapter 6, most of the shared information in the PHQ-9 was duplicated across items. That is the pattern expected when a single underlying construct (depression sever-

ity) drives most of the covariance—the reading the network approach to psychopathology was formulated to contest [85]. In the multimorbidity data of Chapters 4–5 the balance tipped, because information about the target became accessible only through specific joint configurations of symptoms and measures across body systems. At the domain level, pure synergy accounted for 34.7% of the total information about functioning in the whole sample, for symptoms and for diseases alike, and 9.3% for signs. PID therefore seems more informative when a system spans several interacting domains than when a single latent construct dominates. Whether that gain justifies the additional complexity is the next question.

7.1.3. Is synergy worth it?

The third synthesis point is pragmatic: is the additional complexity of a PID-based analysis justified relative to simpler alternatives? Pairwise associations, product-term interactions, and marginal-attribution methods such as SHAP [82] are computationally cheaper and statistically better understood. When the goal is purely predictive, or the data are too sparse for higher-order estimation, they are preferable.

What PID uniquely provides is a decomposition of shared information into qualitatively distinct *modes*. These modes are substitutable (redundant), exclusive (unique), and jointly emergent (synergistic). In Chapter 5, synergy strength and total mutual information were only weakly correlated across the 135 source pairs (Pearson $r = 0.31$), and the two composite rankings they produced shared no pairs in the top 10. Total information sharing and PID synergy therefore flag different source pairs as important. No method that evaluates features individually can register information that exists only in the joint state. The informational-roles taxonomy cannot be derived from pairwise statistics or Shapley-style credit allocation, because it requires knowing whether two sources carry the *same* or *different* context-dependent information about a target.

Furthermore, the echo problem (Chapter 3) showed that apparent higher-order structure can be generated by pairwise redundancy alone. Product-term interactions cannot make that distinction. In a linear regression, a product term tests whether the conditional mean of Y departs from additivity in X_1 and X_2. That is, it tests whether the effect of one predictor varies across levels of the other, with an analogous interpretation on the link-function scale for generalised linear models. PID synergy, by contrast, quantifies information about Y that is accessible only when both sources are observed jointly, drawing on the full joint distribution rather than conditional means alone.

The two measures can dissociate in both directions. A product term can be significant when PID synergy is zero, because misspecified nonlinear main effects leave unexplained curvature for the product term to absorb. Conversely, synergy can be nonzero when no product term is significant, because joint-only information can reside in how outcomes are distributed across joint states rather than in shifts of conditional means.

This distinction is consequential in practice. If two predictors share substantial mutual information and one participates in a genuine interaction with a third variable, the other will also yield a significant product term. Regression provides no principled

decomposition that identifies the second interaction as a statistical echo propagated through source–source redundancy. Collinearity diagnostics may flag instability, but they do not attribute the source of the apparent interaction. PID's explicit decomposition of shared versus complementary information is what makes this distinction possible [41].

In the multimorbidity domain, these distinctions yielded the role-based taxonomy, the BUST structured discovery workflow, and the ePID network-level decomposition—none available from pairwise methods. Tree-based and neural models can implicitly *exploit* higher-order interactions for prediction, but they do not *decompose* why a combination matters. That distinction is essential for measurement design rather than prediction alone. The answer is therefore conditional: synergy analysis is worth the investment when the question extends beyond prediction to understanding how information is distributed across variables and contexts, and when the system contains context-dependent structure. For multimorbidity, these conditions are met. For single-construct scales dominated by a common factor, they may not be.

7.2. Implications and clinical translation

The synergy–redundancy vocabulary and the workflow developed in this thesis have implications that extend beyond the specific cohort analyses reported. This section traces those implications across four domains: PID theory, symptom-network methodology, multimorbidity classification, and clinical translation.

7.2.1. What this thesis contributed back to PID theory

The thesis was designed to apply PID rather than to advance PID axiomatics, but the empirical programme produced observations relevant to the information-theory community. First, sensitivity analyses across $I_{\min}$ [22], I_{mmi}, and BROJA [106] showed that qualitative conclusions were preserved across constructions—an empirical robustness finding whose regime-dependence is discussed in Section 7.3. This observation is relevant to the ongoing theoretical debate about which redundancy measure to adopt [40, 44, 96], as it suggests that in some data regimes the choice of construction matters less for rank-ordering than theoretical divergences would predict.

Second, the echo problem identified in Chapter 3 is the thesis's most practically important contribution to PID methodology. Stated precisely: given sources (X_1, X_2, X_3) and target Y, if $I(X_2; X_3)$ is large and the triplet $(X_1, X_2; Y)$ is synergistic, then replacing X_2 with X_3 typically yields a second synergistic triplet $(X_1, X_3; Y)$, because X_3 acts as a noisy proxy for X_2. The synergy of that second triplet is not mis-estimated. It is a genuine property of those three variables, but it is not independent evidence of a second higher-order structure. What echoes inflate is the number of synergistic sets returned rather than the synergy of any one of them. Screening for echoes therefore reduces the number of correlated candidates carried into significance testing, which is what makes multiplicity correction feasible over a combinatorial candidate set. The screening logic developed for this, redundancy subtraction and overlap filtering, is a heuristic with no guarantee

that it removes only echoes. Its threshold has no principled selection criterion, and it was evaluated only against brute-force rankings on the same data. It remains an open problem to separate synergy that arises from dependence among the sources alone from synergy that requires a genuinely non-additive mechanism. Settling it would need benchmarking: synthetic data with planted synergistic triplets and known redundant proxies, on which true- and false-positive rates could be measured as a function of the overlap threshold. Despite these open questions, the identification and formalisation of this failure mode have implications beyond health data. The same failure arises wherever strong pairwise associations coexist with higher-order structure, including neuroscience [118, 219] and ecology. That generality has not been tested outside the health-data context of this thesis.

Third, the scalable approximation in Chapter 6 raised the question of how context compression interacts with decomposition quality [187]. Compressing multivariate remainder variables into a low-dimensional context representation is pragmatic, but it highlighted a gap in PID theory: there is currently no principled criterion for when such an approximation preserves the qualitative structure of a decomposition. The empirical benchmarking in Chapter 6 provided a first set of data points, but a theoretical treatment of approximation error in PID would be valuable. The information bottleneck guarantees that a compressed representation that preserves $I(E;Y) \geq I(Z;Y) - \epsilon$ will also preserve downstream quantities that depend on $I(Z;Y)$, but it is not known whether this preservation extends to the individual PID atoms (redundancy, uniqueness, synergy) or only to their sum. Establishing such bounds—or demonstrating that they cannot hold without additional structural assumptions—is an open mathematical problem. Moreover, because the ACIB embedding is target-directed, it may systematically preserve redundant information (shared with the remainder's prediction of the target) at the expense of synergy; the high Pearson correlation ($r = 0.92$) between approximated and ground-truth synergy is reassuring but describes agreement across source–target pairs, not the accuracy of any individual atom magnitude.

In practical terms, the Agglomerative Conditional Information Bottleneck (ACIB) embedding validated in Chapter 6 offers a reusable template. Because it learns a target-directed discrete coarse-graining rather than a system-specific compression, the same supervised-embedding step can in principle extend interaction-aware information decompositions to symptom systems beyond the PHQ-9, including the cross-domain multimorbidity variable sets analysed at the domain level in Chapters 4–5. The choice of target-directed rather than variance-preserving compression is not incidental. Variance-preserving methods such as PCA can preferentially retain redundant structure while failing to preserve synergistic higher-order dependencies [212]. The distinction between generic dimensionality reduction and target-directed embedding therefore bears directly on whether a compressed representation preserves the decomposition the analysis is designed to recover.

Finally, the BUST summarisation (Chapter 5) relates to broader efforts in information theory to develop interpretable summaries of multivariate dependence [47, 188]. BUST can be seen as a domain-specific instance of a more general problem: how to com-

press a high-dimensional decomposition landscape into a small number of interpretable dimensions that preserve the distinctions that matter for downstream decisions.

7.2.2. Implications for symptom-network methodology

The informational-roles taxonomy and the BUST workflow transfer directly to psychiatric symptom networks. The pairwise conditional-dependence models that dominate the network approach to psychopathology [36, 85] face the structural limitation addressed here: they cannot represent information that is unlocked only in combination. Chapter 5 showed that pairwise importance and configurational importance were empirically dissociable. Centrality measures in pairwise symptom networks, widely used as proxies for clinical importance [220], may therefore miss features whose relevance is context-dependent. A synergy-augmented symptom network [103]—in which edges carry conditional-dependence weights alongside a redundancy–synergy decomposition relative to a clinical target—would be a richer and more clinically aligned representation. Chapter 6 provided a first implementation. The ePID network decomposed each conditional-dependence edge in the PHQ-9 network into redundancy, synergy, and remainder-unique components relative to the full symptom context. The proposal is therefore technically feasible, and the resulting decomposition replicated across independent cohorts.

Chapter 6 analysed two contrasting psychometric instruments rather than a cross-domain multimorbidity symptom set. The PHQ-9 was chosen because it offered large samples and cross-cohort replication opportunities. The Interpersonal Reactivity Index was chosen because its multi-facet structure provided a synergy-dominated contrast to the redundancy-dominated PHQ-9, showing that the same ePID pipeline discriminates between instrument types. Applying ePID to cross-domain sets—combining pain, fatigue, mood, and cognitive items—would close the loop between Chapter 4's domain-level synergy findings and Chapter 6's variable-level method, directly testing whether the informational-roles taxonomy is recoverable at symptom-network resolution.

The network-versus-latent-variable debate in psychopathology [85] maps onto the redundancy–synergy distinction. If symptom covariance is driven primarily by a common latent factor, the decomposition should be redundancy-dominated. If direct causal interactions among symptoms play a substantial role, synergy should be more prominent. The PHQ-9 analysis (Chapter 6) offered a first empirical window: redundancy accounted for 45–51% of total mutual information and synergy for 6–9%, a balance more suggestive of a common factor than of pervasive causal coupling. That this reflects the instrument rather than the method is underscored by the contrasting Interpersonal Reactivity Index, where the identical pipeline returned a synergy-dominated decomposition. ePID registers synergy dominance when it is present rather than defaulting to redundancy. This finding must be read cautiously: the PHQ-9 is a short screening instrument whose items were selected for diagnostic coverage rather than causal specificity [221]. Second, the PHQ-9 is a short screening instrument whose items were selected for diagnostic coverage rather than causal specificity [221]. A more fine-grained symptom space could shift the balance.

What ePID contributes is not a verdict on the debate but a tool that makes the question answerable with data.

The critical debate about network approaches in psychopathology bears on these contributions. Three criticisms apply to varying degrees to any network-oriented analysis, including PID: the limited replicability of network parameters across samples [215], the argument that strength centrality reduces to factor loadings rather than providing novel structural information [222], and the over-interpretation of cross-sectional network models [223, 224]. PID, by contrast, does not rely on graph-topological centrality measures as its primary interpretive output: importance is always target-conditioned and decomposed into qualitatively distinct modes rather than derived from network topology. While centrality metrics could, in principle, be computed on networks derived from PID decompositions, the core framework itself sidesteps the critique that centrality reduces to factor loadings [220, 222], because its explicit redundancy atom makes the latent-variable concern transparent rather than concealing it behind edge weights. PID does, however, introduce its own replicability burden. The combinatorial parameter space is larger than in pairwise networks, making single-cohort synergy findings more fragile than single-cohort edge estimates. Replication across instruments and populations, beyond the initial cross-cohort results reported in Chapter 6, remains essential.

7.2.3. Implications for multimorbidity classification

Current multimorbidity indices are predominantly additive [16]: they count diagnoses, weight them by severity or resource use, and sum them. In Chapter 4's total sample, the information that symptoms, signs, and diseases carried about functional limitation was almost entirely interaction-driven. A count cannot represent that. Additive indices are therefore not wrong but structurally incomplete: they under-describe the dependence that drives patient-centred outcomes such as daily functioning [225]. The remedy is to complement counts with configuration-aware summaries that distinguish patients whose conditions interact from those whose conditions merely coexist.

Pattern-based approaches to multimorbidity classification already move past simple counts. Latent class analysis, cluster analysis, and network-based methods identify which conditions co-occur in clinically meaningful groupings [226]. What PID adds is a distinction these approaches cannot make: whether co-occurrence is informationally redundant or synergistic for a specific outcome. Two conditions in the same cluster might be largely redundant (reflecting the same underlying pathology, such as shared cardiovascular-metabolic risk) or synergistic (creating emergent risk that neither condition produces alone). Clustering identifies the grouping, not the informational role.

The same argument applies to frailty. The Rockwood Frailty Index (FI) [136] captures the accumulation of deficits across multiple domains but does not characterise the informational structure of that accumulation. Two patients with identical FI scores can differ substantially in their synergy profiles: in one the deficits are largely redundant, many expressions of a single underlying process, and in the other their combination creates a vulnerability that no deficit carries alone. A synergy-augmented index would

separate those two patients. The proposition is testable: comparing the predictive validity of a standard frailty index against a synergy-augmented version for outcomes such as hospitalisation, falls, or mortality would give direct evidence on whether informational structure adds prognostic value beyond the count itself.

The most stable synergistic constellations identified in this thesis spanned measurement domains. Chapter 5's recurrent multidomain clique linked general health, pain, grip strength, and cognition. In Chapter 3, metabolite triplets drawn from different functional groups tended to be more strongly synergistic than triplets confined to a single group. Both patterns are consistent with geroscience frameworks: shared pathological processes—inflammaging, metabolic dysregulation, and cellular senescence—may give rise to multi-system dependencies that simultaneously affect mood, cognition, physical function, and disease progression [227, 228]. The thesis cannot confirm these mechanistic pathways from observational data alone, but the alignment between informational patterns and known biology provides face validity that is essential for the geriatric community to engage with PID-based analysis. This multi-system perspective also reframes what personalised measurement requires in practice, a theme developed in the next section.

7.2.4. Towards context-sensitive assessment in multimorbidity

Personalised medicine is frequently invoked as an aspiration, yet what it demands of measurement design is rarely made explicit. Before treatment can be tailored, assessment itself must be tailored: deciding which questions to ask, which tests to order, and which follow-up to pursue given what is already known about a patient. Consider fatigue in a 78-year-old with diabetes, COPD, and mild depression. Fatigue could reflect poor sleep, medication side-effects, deconditioning, or low mood. Its informational value for predicting functional decline depends on which other symptoms have already been observed. The synergy–redundancy lens makes this intuition operational. *Synergy* flags combinations of symptoms, signs, or measures whose joint presence carries information about an outcome that neither provides alone, motivating targeted follow-up. *Redundancy* identifies items that largely duplicate each other's contribution, justifying parsimony where burden must be minimised.

Operationally, a redundancy-driven perspective supports deliberate pruning of assessment batteries. When two symptoms, signs, or instruments share most of their target-relevant information within a subgroup, collecting both increases burden and cost without a commensurate gain in inferential power. In a setting where both are routinely assessed, one could plausibly be dropped with minimal information loss for the targets examined, provided that the retained measure adequately covers the clinical decisions for which the dropped measure was originally collected. This is a claim about measurement design under constraints rather than a claim that any symptom is clinically unimportant: pruning decisions should be validated against clinical safety, downstream performance, and the possibility that a measure is required for decisions beyond the focal target. For instance, the 'memory complaints' and 'memory problems' items in Chapter 5

clustered together and ranked high on total mutual information while contributing only modest synergy, suggesting that both need not be retained when the assessment target is functional limitation.

These observations suggest an adaptive assessment pathway. A small core set of broad, low-burden indicators can be collected routinely to establish a baseline context. Additional questions or tests are then triggered only when they are expected to add incremental information for the target in that context. This logic extends the principles of computerised adaptive testing (CAT) [229]—where item selection is driven by expected information gain for a single latent trait—to multivariate, multi-domain health states. Synergy can be read as analogous to a value-of-information quantity: how much an additional measurement is worth given what has already been observed. A sequential rule built on it would collect the next item only when the expected synergistic gain exceeds the burden of asking. The value-of-information reading is a structural analogy rather than a demonstrated property of these estimates. Whether the cohort-level informational roles identified here operate within an individual through time is an open empirical question requiring intensive longitudinal designs [230].

7.2.5. Toward prospective validation

Translating cohort-level findings into personalised assessment requires a staged pathway, beginning with the large-scale cohort analyses reported in Chapters 4–5: mapping redundancy structure and identifying a limited set of stable synergy gates for specific targets such as daily functioning. Those cohort-level signatures then motivate prospective monitoring studies—ideally within-person—where symptoms and relevant signs are captured at higher cadence around clinically meaningful transitions such as exacerbations, recovery episodes, and treatment changes. The aim is to test whether the same gates recur over time and whether they anticipate outcomes. Only where stability and utility have been demonstrated can the gates be embedded into decision support as context-sensitive prompts: recommending the next measurement or follow-up question conditional on the current evidence state, and communicating uncertainty when the expected information gain is low. How such a system should be evaluated for decision impact is taken up in Section 7.4, Programme 3.

The reasoning supported by synergy-informed assessment extends beyond measurement to intervention planning. Chapter 2 demonstrated this directly: in a synthetic Bayesian network, a pairwise representation prescribed four separate treatments for a patient's active signs because it could not detect the connections among them, whereas the hypernetwork revealed that a single intervention on S_2 deactivated the synergistically linked signs S_4 and S_6 [48]. At the population level, the same chapter showed that nudge interventions propagate unexpectedly through synergistic pathways: nudging S_2 shifted the marginal distributions of nodes invisible to the pairwise graph. Public-health strategies built on pairwise evidence may therefore underestimate or misdirect their downstream effects. Both demonstrations rest on a single nine-node topology. They establish that such failures can arise by construction, not how often they arise in real

data, and they do not transfer directly to clinical decision-making. The structural claim nonetheless holds: when dependence is irreducibly multivariate, measurement and intervention reasoning must account for synergistic pathways to avoid redundant effort or missed consequences.

The simulations also suggest a design principle for strategic multi-interventions in multimorbidity. Rather than treating each active condition independently, synergy-informed intervention planning would identify nodes whose activation propagates through higher-order pathways, targeting a smaller set of leverage points that jointly cover more of the dependence structure. This remains speculative, as the simulations used a synthetic network, and translating the principle into real clinical decisions requires the causal validation outlined in Programme 2. Synergy-informed multi-intervention design is therefore a long-term goal for multimorbidity care: a move from treating each active condition in isolation toward coordinated interventions at a small set of leverage points identified by their higher-order reach. It is not a near-term proposal.

A clinical vignette illustrates how synergy-informed prioritisation might work in practice. A primary-care clinician reviews the same 78-year-old patient at a routine consultation. In current practice, comprehensive geriatric assessment (CGA) already embodies a multi-domain, interaction-aware approach, assessing functional, cognitive, nutritional, social, and psychological domains together. What the synergy framework adds is not a replacement for that clinical judgement but a quantitative prioritisation logic. Drawing on population-level patterns such as the perceived-health/physical-activity synergy gate identified in Chapter 5, such a system might flag that the combination of fatigue and reduced physical activity carries information about functional-decline risk beyond what either indicator signals alone—recommending a brief follow-up question about activity levels and triggering a medication review to identify iatrogenic contributors to fatigue. This is most valuable in primary care or resource-constrained settings where full CGA is not feasible and clinicians must decide which assessment domains to prioritise under time pressure. This vignette illustrates reasoning, not a proposed system: the synergy gates described are population-level distributional signatures whose within-person predictive validity remains untested, and any deployed system would need to communicate its confidence and be evaluated for error rates before clinical adoption.

Synergy-informed network visualisations could also serve a pedagogical purpose. They present students and trainees with an empirically grounded map of which symptom combinations carry emergent information for specific outcomes, making higher-order interaction concrete and domain-specific.

Only the first stage of this pathway has been completed here. Chapters 4–6 delivered the cohort-level analyses. No prospective monitoring study, decision-impact evaluation, or power calculation for higher-order effects in a longitudinal design was undertaken. Realising the later stages also demands engineering, ethical, and regulatory work that lies outside the analytic scope of this thesis.

7.3. Limitations and boundary conditions

The implications outlined above assume that the higher-order signatures identified in this thesis are trustworthy. Those signatures are bounded by the regimes in which multimorbidity data are currently collected, and several constraints are especially consequential.

7.3.1. Domain-level aggregation and system size

Chapter 4's analysis was constrained to a system of four aggregate variables: symptoms, signs, diseases, and functioning. The restriction was a deliberate trade-off. The small system made three-source PID tractable and interpretable, but it could not reveal which individual variables within each domain drove the synergy or how those contributions were distributed across specific symptoms, signs, and diseases. That limitation motivated the progression from domain-level to variable-level resolution in Chapters 5 and 6, which disaggregated the domains and examined informational roles at the level of individual clinical features. Chapter 4 established that domain-level synergy exists, and the chapters that followed asked where, specifically, it resides.

A second feature of that design warrants comment. The Chapter 4 estimates were computed on 14,717 observations contributed by 4,025 participants, merged across waves and treated as a single cross-sectional dataset. As that chapter states, the models do not account for the dependence in observations within persons over time. Chapter 4 probed this by repeating the random-forest models within the first and last waves separately. Participant-clustered resampling—grouped cross-validation for the classification results, a cluster bootstrap for the PID atoms—was not performed. Until it is, the domain-level estimates are best read as describing person-waves rather than independent individuals.

Domain-level aggregation may also mask within-domain heterogeneity. If synergy is concentrated among a small subset of symptoms or signs, averaging across an entire domain can dilute the signal or, conversely, amplify it if aggregation conflates informative and uninformative variables. The variable-level analyses in Chapters 5 and 6 partially addressed this concern by revealing which individual features carried synergistic information, but the domain-level findings of Chapter 4 should be interpreted with this caveat in mind.

A related constraint concerns the breadth of the measurement instruments themselves. LASA and EASYcare capture a bounded set of common symptoms, signs, and chronic conditions, whereas real clinical encounters feature a much longer tail of rare or idiosyncratic presentations. The informational roles identified here should therefore be read as properties of common presentations, and their transfer to rarer configurations remains an open question.

7.3.2. Temporal mismatch and aliasing

Multimorbidity processes operate on timescales ranging from hours to years, whereas the cohorts analysed here sample far more slowly. LASA waves are roughly three years apart. Dependence that unfolds between measurement occasions can therefore be aliased into apparently contemporaneous dependence. The PID decompositions reported in

this thesis describe what is identifiable at the observed cadence, not the process that generated it.

The dependence signatures identified in this thesis are cohort-level statements about how information is organised across individuals at a measurement occasion. They do not establish that the same informational roles and gates operate within an individual through time. Within-person designs are essential to distinguish trait-like co-encoding from state-dependent interaction structure. Without them, projecting between-person patterns onto individual trajectories remains an assumption rather than a finding. The mismatch is sharpest for symptoms. They change rapidly and are among the most patient-centred variables in multimorbidity, yet they remain comparatively under-measured. The transition-focused, symptom-forward designs this requires are set out in Section 7.4, Programme 1.

7.3.3. Treatment processes as unmeasured context

The variables entering the PID analyses—symptoms, signs, cognitive scores, functional limitations, and chronic-condition indicators—described the patient's observed health state, not the clinical management surrounding it. Medication use, prescribing changes, deprescribing episodes, and care utilisation were not included as sources or conditioning variables. Some LASA waves collected them, at least partially. In older populations with multimorbidity, polypharmacy is the norm and treatment regimens shift frequently. These processes shape the symptom landscape continuously. A synergy gate linking sleep quality and physical activity, for example, may partly reflect a recent medication adjustment that affects both sleep and mobility rather than a biological interaction between the two indicators. Because treatment processes were absent from the variable set, the PID decomposition could not distinguish information arising from disease-state interactions from information arising from care-delivery patterns.

Polypharmacy and prescribing cascades supply a concrete mechanism. A beta-blocker that causes fatigue and a diuretic that causes dizziness can jointly raise fall risk in a way that neither produces alone. That is a synergistic pattern driven by prescribing rather than by disease interaction. Chapter 1 opened with a case of the same shape: co-trimoxazole and an ACE inhibitor turn hazardous in a patient with reduced renal function. Medication data never entered the empirical chapters. Applying PID to symptom-medication-diagnosis triplets is therefore a concrete next step, and the prevalence of polypharmacy in the study populations makes it a high-priority one.

This limitation matters most for the deployment scenario outlined in the clinical translation section above. If synergy-informed assessment guides clinical decisions, clinicians need to know that the gates they interpret are properties of the observed system, *including* its clinical management context, not of the underlying biology alone. This does not invalidate the identified gates. A gate shaped by treatment logic still signals that clinical decisions in that population are context-dependent, which is itself informative for care-pathway design. It does, however, constrain their causal interpretation. Treatment-mediated synergy remains clinically actionable: it marks where deprescribing

or medication review should target the *combination* rather than individual medications in isolation. Disentangling biological interaction from treatment-mediated covariance requires the intervention-aware designs and explicit treatment-process capture called for in Section 7.4, Programme 2. That is a necessary step before synergy gates can be attributed to specific pathological mechanisms.

7.3.4. PID non-uniqueness and estimator sensitivity

PID constructions differ in their axioms and can produce different numerical decompositions [40, 44, 96]. The divergence is not merely numerical but semantic. $I_{\min}$ defines redundancy via the pointwise minimum of specific information, I_{mmi} via the minimum mutual information between any source and the target, and BROJA via a constrained optimisation that preserves the marginal source–target distributions [105, 106]. These are different operational definitions of what it means for two sources to "say the same thing." On pathological distributions, including those with asymmetric marginals or exclusive-OR-like joint configurations, the constructions can disagree qualitatively rather than only quantitatively. This thesis addressed the resulting non-uniqueness by replicating across constructions rather than by adopting a single "correct" measure. Qualitative conclusions were preserved across each set of constructions that was actually compared: $I_{\min}$, I_{mmi}, and BROJA in Chapter 5, and I_{mmi} against $I_{\min}$, I_{rr}, $I_{\pm}$, and $I_{\wedge}$ in Chapter 6. Chapter 4 used I_{mmi} alone and was never checked against an alternative redundancy definition. Its Random Forest replication provides an independent, non-PID check on the same conclusion. This agreement is an empirical observation about the data regime of health cohorts—discrete variables with low-to-moderate cardinality and large samples—and not a theoretical guarantee. No result establishes that the constructions must converge in this regime. The robustness observed here is therefore regime-dependent and should not be assumed to generalise.

PID decompositions are also conditional on, and therefore sensitive to, the included variable set. Adding a variable to the source set can redistribute information across atoms: a pair that appeared synergistic may become redundant once a previously omitted variable captures the same joint information. This is not only a practical inconvenience. A bivariate decomposition computed on a pair drawn from a larger system is not guaranteed to agree with a decomposition of the system that contains it. The pairwise atoms reported in Chapter 5 are therefore statements about the selected pair and target, not fragments of one decomposition of the full 17-variable set. Variable selection in this thesis was guided by domain knowledge and instrument availability, and the analyses were not exhaustive. The multi-chapter strategy addressed the concern only partially, in that different variable sets in different chapters converged on similar qualitative conclusions. A leave-one-variable-out and add-one-variable perturbation, reporting rank stability and top-10 overlap as the binning-sensitivity analysis of Chapter 5 already does for discretisation, would turn this open challenge into a measured quantity. That experiment was not run.

Discretisation of continuous variables, required by the plug-in entropy estimators used in Chapters 4–6, introduces researcher degrees of freedom that interact with synergy

estimation. The expected direction of bias is informative. Coarse discretisation tends to reduce synergy by collapsing jointly informative states and is therefore conservative, while fine discretisation tends to inflate synergy through estimation noise and is, therefore, liberal. Most LASA and EASYcare variables were natively binary or ordinal, which limits discretisation sensitivity for the majority of analyses. The metabolomic analyses in Chapter 3, which relied on quantile binning of continuous nuclear magnetic resonance (NMR) data, are more exposed to this concern. The thesis addressed discretisation partly through bias correction (quadratic extrapolation in Chapter 5) and partly through clinically motivated cutpoints where available. Chapter 5 is the only chapter to re-run its full pipeline under alternative binning schemes. Resolution was not varied systematically elsewhere.

7.3.5. Self-report and measurement validity

A further source of correlation is the measurement process itself. The majority of variables in the LASA and EASYcare analyses were self-reported: symptoms, functional limitations, and chronic conditions all relied on patient recall and were subject to response biases. Correlated response biases such as acquiescence and social desirability inflate redundancy. Spurious synergy, by contrast, requires no special error structure. Independent measurement noise on either source is by itself sufficient for tripartite measures—partial correlation, variance partitioning, and PID alike—to return false-positive synergy and redundancy [231]. The large samples analysed in Chapters 5 and 6 are therefore not protective. Neither are the target-shuffle nulls of Chapter 5, which preserve the joint distribution of the sources by design and so leave intact the source–source noise structure that generates the artefact. The PID decomposition cannot separate information arising from the underlying health state from information arising from the measurement process. Biomarkers such as CRP, IL-6, HbA1c, and metabolic panels are available in some LASA waves and extensively in UK Biobank, and were not used in this thesis. Adding objectively measured variables of this kind alongside self-report would show how far the identified informational roles reflect clinical reality rather than measurement artefact.

7.3.6. Observational dependence is not causality

The limitations above delimit the causal reach of PID-based synergy: domain-level aggregation, low measurement cadence, unmeasured treatment context, estimator and variable-set sensitivity, and reliance on self-report. None of the cohorts analysed here was interventional. Yet these limitations also clarify synergy's proper role in causal work. Although synergy is not itself a causal interaction term, stable synergy gates are informative about where purely dyadic causal representations are likely to be misspecified. A gate indicates that target-relevant information emerges only jointly, but conjunctive causation, effect modification, and latent confounding all produce that signature, and cross-sectional PID cannot tell them apart. The same configuration troubles causal discovery itself: XOR-like dependence, which is pure synergy in PID terms, violates the standard faithfulness assumption, and recovering such edges requires explicitly weak-

ened variants of it [232]. Read this way, synergy is a disciplined precursor to causal discovery [233]. It narrows an otherwise unbounded interaction search to a small set of candidates with demonstrated predictive relevance and stability in the observed distribution. In practical terms, a stable gate is a candidate for interaction terms in structural equation models, for tests of heterogeneous treatment effects in trial data, or for non-additive pathways in directed acyclic graphs. The step from synergy to causality requires temporal ordering, intervention or quasi-experimental variation, and explicit confounding control—requirements that define a concrete next-step programme rather than an in-principle limitation.

7.3.7. Alternative explanations

The thesis's interpretive framework also makes claims that the statistical decompositions alone cannot establish. Three alternative readings warrant scrutiny: one conceptual, one statistical, one empirical. First, the informational-roles taxonomy imposes categorical labels—general indicators, disambiguators, synergy gates—on continuous BUST scores. Whether these categories mark genuinely distinct modes of information organisation or are merely convenient labels along a continuum is an open question, and resolving it would require distribution analysis and differential external validation. Second, the cross-domain synergy advantage reported in Chapters 4–5 is consistent with genuine multi-system interaction. However, less-redundant source pairs mechanically leave a larger share of mutual information available for synergy in any PID decomposition. Part of the advantage may therefore reflect variable heterogeneity rather than biological interaction, a confound the present analyses did not control for. Third, the divergence between synergistic and pairwise importance rankings noted above ($r = 0.31$) shows that PID identifies different variables as important, but divergence alone does not demonstrate clinical value. The decomposition's meaningfulness currently rests on face validity and theoretical plausibility, not on prospective evidence that synergy-ranked features predict patient outcomes better than simpler alternatives. Each alternative is testable. The evidence reported here is consistent with the thesis's interpretation but does not exclude the alternatives, and the programme outlined in Section 7.4 is designed in part to resolve them.

7.4. Future directions

Each of the limitations identified above point toward a tractable next step. One of those steps requires no new data and is a prerequisite for the rest. The remaining questions fall into three research programmes.

A prerequisite: inferential calibration. Finite-sample bias, null-model choice, and multiplicity can all be addressed with the cohorts already in hand [113]. Three steps would do most of the work. First, the target-shuffle null used in Chapter 5 removes estimator-specific bias but leaves the sources untouched, so it cannot rule out apparent

synergy generated by correlated error on the source side. Generative nulls that reproduce that error structure would provide a stronger test. Second, multiplicity is controlled at the screening stage, where Benjamini–Hochberg is applied to a joint mutual information test, rather than at the level of the synergy atom finally reported. Echo filtering (Chapter 3) reduces the candidate count but supplies no error rate of its own. Third, resampling should be carried out at the participant level wherever a cohort contributes repeated measures, as it does in the pooled datasets of Chapters 4 and 5. Those datasets draw about 14,700 observations from 4,025 participants, so observation-level bootstrapping treats correlated records as independent and understates uncertainty. None of this requires new data collection. Each step would tighten the inferential claims the three programmes below are designed to extend.

Programme 1: Within-person, dynamic higher-order inference. The most immediate extension is from cross-sectional, between-person analysis to within-person longitudinal designs. This requires dense symptom time series—ideally event-triggered around clinical transitions—and dynamic models capable of representing time-varying higher-order dependence. Methodologically, the appropriate models are closer to switching hypergraphs or regime-aware state-space formulations than to smoothly evolving pairwise graphs, because the dependence structure itself may change across clinical states. Time-windowed and lagged PID estimators [110, 234] offer a pragmatic entry point, allowing synergy and redundancy to be tracked as functions of the temporal lens. An important connection exists to the multimorbidity trajectory literature, which has identified distinct longitudinal patterns of disease accumulation and transition [235]. Synergy-informed trajectory analysis could reveal whether patients following the same trajectory class differ in their higher-order dependence structure, and whether shifts in synergy precede clinically recognised trajectory transitions.

Programme 2: Causal and intervention-aware extensions. Stable synergy gates are candidate inputs for causal modelling [233]: they identify where interaction structure is worth testing in causal outcome models, or where to look for heterogeneous treatment effects. At a minimum, causal interpretation requires repeated measurement with clear temporal ordering, detailed capture of treatment processes, and adequate measurement of plausible confounders. Designs that introduce exogenous variation—randomised or quasi-experimental interventions embedded in routine care—are particularly valuable because they allow the causal status of a candidate gate to be tested rather than inferred.

Programme 3: Clinical evaluation and deployment. The staged evidence pathway proposed in this thesis runs from stable cohort-level signatures, through within-person validation, to decision-impact evaluation in pragmatic trials. It provides a tangible programme for clinical translation. The highest-value outputs for clinical deployment are sequential and conditional. Given what is already known today, what is the next best question, test, or follow-up that is worth the burden for this person *in this context*? A

synergy-aware system can support that framing, but only if implementation respects burden constraints, logs decisions and outcomes for continual auditing, and updates cautiously when data drift or changes in practice shift the joint distribution. In routine care, deployment must also address the communication contract: the system should state what it is confident about, what it is uncertain about, and what additional information would meaningfully reduce that uncertainty.

7.5. Toward adaptive, data-rich multimorbidity research

All three programmes presuppose richer data than the cross-sectional cohort designs that supplied the analyses in this thesis. Meeting that requirement is not a matter of collecting *more* data but of collecting *smarter* data, with deliberate attention to the informational structure that higher-order analysis demands. What follows is not a fourth programme. The first three requirements below are the measurement infrastructure the programmes presuppose. The fourth is the clinical endpoint toward which Programme 3 is directed.

Richer, replicated cohort data. Higher-order analysis needs richer variable sets *and* replication across populations. Synergy estimation is combinatorial, which makes single-cohort findings more fragile than pairwise results. The cross-cohort replication reported in Chapter 6 was possible only because both cohorts happened to include the PHQ-9 [184]. For the cross-domain multimorbidity analyses in Chapters 4–5, no comparable replication dataset currently exists. The implication for research infrastructure is deliberate overlap: harmonised variable coding and cross-domain symptom batteries in new cohort waves. A single validated ten-item battery, shared by two cohorts, would create the common variable set that cross-domain replication currently lacks.

Context-triggered measurement. Passive sensing technologies—accelerometers, GPS traces, sleep-wake detection—capture physical activity and mobility at a granularity that cohort questionnaires cannot match [236, 237]. However, they cannot capture the subjective symptoms that carried the strongest synergistic signals in this thesis: perceived health, pain interference, mood states, and cognitive complaints. Those require active self-report, which introduces a respondent burden. The opportunity lies in combining passive sensing with *context-triggered* active prompting. A passively detected state change, such as a sustained drop in activity alongside deteriorating sleep, becomes the trigger for a brief symptom query. The synergy–redundancy framework provides a principled basis for deciding *which* questions to ask and *when*: ask when the expected synergistic information gain exceeds the respondent burden. This connects to just-in-time adaptive intervention frameworks [238], which require precisely this kind of context-sensitive triggering logic.

Synergy-aware instrument design. Standard psychometric practice optimises instruments for internal consistency, retaining items that correlate highly with one another

and producing scales dominated by redundant information [239]. The scientific question in multimorbidity research concerns context-dependent, cross-domain interactions. Optimising for internal consistency may therefore exclude precisely the items that carry synergistic information. The PHQ-9 illustrates the tension from the other direction. Its nine items map one-to-one onto the diagnostic criteria for major depression [184] rather than being selected to maximise internal consistency, yet Chapter 6 found redundancy dominating its information landscape at 45–51% of total mutual information. Even criterion-driven item selection can produce a redundancy-dominated instrument. That strengthens the case for a complementary synergy-aware criterion, one that evaluates whether a candidate item contributes non-redundant, context-dependent information about a clinically relevant target. Existing instruments need not be discarded. They could instead be augmented with items that appear weakly predictive alone but unlock joint information in the presence of specific co-occurring symptoms. Seven common depression scales collectively encompass 52 largely non-overlapping symptoms [240], so the raw material for such augmentation already exists.

From characterisation to prevention. These infrastructure elements converge on a clinical ambition that extends beyond characterisation. The prevention pathway is sequential: identify high-synergy configurations at the cohort level, validate them as prospective predictors of adverse transitions in longitudinal data, embed the validated configurations as triggers in adaptive monitoring systems, and deliver targeted preventive action when a synergistic configuration is detected in an individual's data stream. The intervention logic is qualitatively different from conventional screening: it is triggered not by crossing a single threshold but by detecting a risk-relevant *joint configuration*. The thesis provided the first steps in this pathway: characterisation of informational structure, and the computational methods that make that characterisation tractable and reproducible. Prospective validation, intervention design, and pragmatic evaluation are the bridge from descriptive information theory to actionable clinical prevention.

7.6. Conclusion

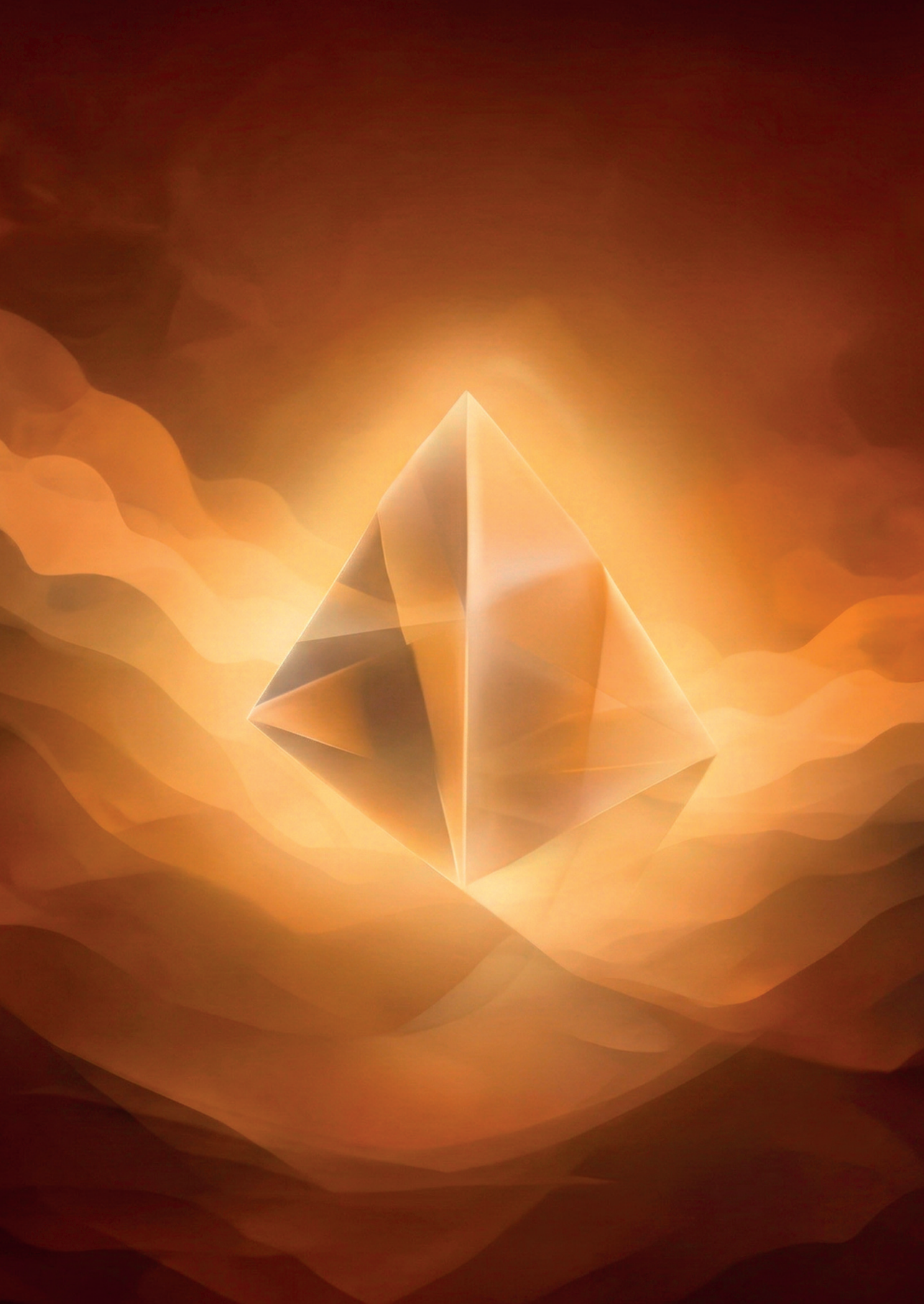
This thesis showed that pairwise associations do not exhaust the information symptoms carry about patient-centred outcomes. Some of that information is synergistic. It emerges only from specific joint configurations and is invisible to any analysis that examines variables one at a time. That finding has consequences for how multimorbidity is measured, modelled, and eventually managed.

The contribution was threefold. The thesis established that clinically relevant higher-order dependence exists and is not an estimation artefact (Chapter 4); it provided the computational infrastructure—scalable search (Chapter 3), structured summarisation via the BUST workflow (Chapter 5), and network-context approximation via ePID (Chapter 6)—to make that dependence tractable; and it introduced a conceptual vocabulary of informational roles that reframes multimorbidity measurement as an information-design problem. Cross-cohort replication and cross-measure sensitivity analyses provided con-

vergent evidence that these findings reflect stable features of the joint distribution, not artefacts of a particular dataset or estimator.

What the thesis did *not* provide is a clinical decision tool, a causal model, or proof that synergy-informed care improves patient outcomes. Those claims require within-person longitudinal validation, causal designs, and pragmatic trials. The clearest next step is to apply ePID to repeated cross-domain measurements in an ageing cohort such as LASA. The question is whether synergy gates identified at baseline predict incident functional decline over subsequent waves. If configuration-dependent information anticipates transitions that pairwise summaries miss, the case for synergy-informed assessment moves from plausible to evidentially grounded.

The return to symptom-centred reasoning formalised here serves a patient-centred goal. That goal is to understand the person's health as an integrated whole rather than a list of separable conditions. What the patient experiences—and what the clinician encounters first—is not a set of independent diagnoses but a configuration of interacting symptoms and signs whose meaning depends on context. The tools developed here provide a quantitative way to take that clinical reality seriously.



SUPPLEMENTARY MATERIAL FOR CHAPTER 2

Three passages in Chapter 2 defer to this appendix. The chapter's primer on synergy points here for the mutual information variants, the O-information, partial information decomposition, and the synergy constructs used in other domains. Its finding that pairwise associations alone split the synthetic network into three disconnected components is revisited below using thresholded partial correlations. Its analysis of the EASYcare data points here for the synergistic triplets themselves, tabulated below alongside those from the synthetic model. The remaining sections give an example of synergy that the O-information misses and the reasons binary variables suffice. We begin with two worked examples of synergistic triplets.

A.1. Intuitive examples of synergistic triplets and XOR gates

Chapter 2 derives the mutual information terms of an XOR gate. The table below restates that example with two switches and a lamp. An XOR gate returns one when its two binary inputs differ, and zero when they are the same. Let the third column (Lamp) be the output we want to predict, and columns one and two be independent inputs, each with probability 0.5.

Table A.1: An example of an XOR gate.

Switch 1	Switch 2	Lamp
Off	Off	Off
Off	On	On
On	Off	On
On	On	Off

We know the probability the lamp is on is 0.5. Learning only one switch status gives us no information about the lamp being on/off. We gain information about the lamp only if the state of both switches is known. Here, information about the lamp is not contained in either switch (no pairwise/dyadic mutual information), but is contained synergistically in their combination.

Table A.2: An example of a relationship with both pairwise and synergistic associations.

Sex	Intercourse	Pregnant
Male	No	No
Male	Yes	No
Female	No	No
Female	Yes	Yes

In the second table the output we want to predict is whether a person is pregnant. The baseline probability is 0.25. Once we know the sex, the pregnancy probability changes to 0 for male and 0.5 for female, and the same holds if we know intercourse but not sex. Thus, each input alone provides information about the output, but there is also synergistic information in their combination. Using synthetic data, we find an O-information score of approximately -0.19 for this table. The negative sign indicates that synergy outweighs redundancy in this triplet, not that pairwise information is absent.

A.2. Mutual information variants and O-information

Mutual information between two random variables X and Y can be written as:

$$MI(X;Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X) = H(X) + H(Y) - H(X,Y).$$

A

Conditional mutual information is:

$$MI(X;Y \mid Z) = H(X \mid Z) - H(X \mid Y,Z).$$

Two multivariate extensions of mutual information are the Total Correlation (TC) and the Dual Total Correlation (DTC), which can be written as

$$TC(X_1,\ldots,X_n) = MI(X_1;X_2) + MI(\mathbf{X}_1^2;X_3) + \cdots + MI(\mathbf{X}_1^{n-1};X_n),$$

$$DTC(X_1,\ldots,X_n) = MI(X_1;X_2 \mid \mathbf{X}_3^n) + MI(\mathbf{X}_1^2;X_3 \mid \mathbf{X}_4^n) + \cdots + MI(\mathbf{X}_1^{n-1};X_n),$$

where $\mathbf{X}_i^j = (X_i,\ldots,X_j)$ is a shorthand notation. Both quantities equal the mutual information for two variables. TC and DTC are zero if, and only if, the random variables are independent. TC accounts for collective constraints, while DTC measures the shared randomness [42]. The O-information is defined by their difference:

$$\Omega_n(X_1,\ldots,X_n) = \mathrm{TC}(X_1,\ldots,X_n) - \mathrm{DTC}(X_1,\ldots,X_n),$$

where $\Omega_n < 0$ implies dominance of higher-order constraints (synergy) and $\Omega_n > 0$ implies dominance of low-order constraints (redundancy). A more complete introduction can be found in [42, 112].

A.3. An example of synergy missed by O-information

The O-information identifies the primary trend of information in data and classifies it as either synergistic or redundant. In cases where both trends coexist, they may negate each other, causing some synergy to go undetected. For instance, let X_1 and X_2 be independent binary variables, each with probability 0.5, and let $Y = \mathrm{XOR}(X_1,X_2) + X_1$. The triplet $(X_1,X_2;Y)$ has an O-information of zero. This can be checked directly: the four states (X_1,X_2,Y) are $(0,0,0)$, $(0,1,1)$, $(1,0,2)$ and $(1,1,1)$, each with probability $1/4$, which gives a total correlation of 1.5 bits and a dual total correlation of 1.5 bits. The O-information is their difference, and so is exactly zero. The triplet is dominated by neither synergistic nor redundant information. However, Y is composed of synergistic information from the XOR gate and redundant information from the copy of X_1.

A.4. Partial Information Decomposition

Partial information decomposition (PID) aims to decompose mutual information into unique, redundant, and synergistic components. For a triplet $(X_1,X_2;Y)$ the decomposition is

$$MI(X_1,X_2;Y) = U_1 + U_2 + R + S,$$

where U_1 and U_2 are the unique contributions of each source, R the redundant, or shared, information both carry, and S the synergy carried only by the pair. Three axioms constrain any candidate redundancy measure $I_\cap$, written here for the target Y and source sets $A_1, \ldots, A_k$ [73]. Symmetry requires $I_\cap(A_1, \ldots, A_k; Y)$ to be invariant under any permutation of the sources. Monotonicity requires $I_\cap(A_1, \ldots, A_{k-1}; Y) \geq I_\cap(A_1, \ldots, A_k; Y)$, with equality when A_k contains one of the earlier sets, so redundancy cannot increase as sources are added. Self-redundancy requires $I_\cap(A_i; Y) = MI(A_i; Y)$ for a single source set. No measure is known to satisfy all three axioms for an arbitrary number of sources and a single target [72].

A.5. Synergy constructs in other domains

Capturing synergistic interactions is relevant in many domains such as drug-drug combinations, disease co-occurrences, and functional activity in the brain [41, 76–78]. In AI and machine learning, Shapley values quantify feature importance and can be used to reveal interactions and synergistic associations [241].

A.6. The sufficiency of binary variables

The metrics used in Chapter 2 rely on computing mutual information. Binary data is sufficient because mutual information quantifies statistical dependence between discrete variables, and binary variables make the required joint probabilities easy to estimate.

A.7. Reconstructing the synthetic network structure using pairwise associations

We reconstructed the association network of the synthetic model from pairwise associations, computed as partial correlations on simulated data. The network of all non-zero partial correlations contains many spurious and non-significant edges, which can be removed by thresholding correlation strength or by regularisation. Retaining only correlations with a magnitude of at least 0.2 leaves three disconnected pairwise networks rather than the single hypernetwork of the ground truth. Synergistic associations are needed to fully characterise all correlations in the data.

A.8. O-information triplets from the synthetic model

The top-ranking synergistic triplets (O-information below approximately −0.07) are:

A.9. Possible synergistic triplets in the EASYcare data

A permutation test of O-information values computed on triplets with low dyadic mutual information found 184 triplets to be significantly synergistic (p-values < 0.05), with the Benjamini-Hochberg procedure controlling the false discovery rate at 0.15. Examples are shown in Table A.4.

Table A.3: Synergistic triplets from the synthetic toy model along with a description of their functional form.

Triplet	O-information	Functional description
$\{S_1, S_4, D_1\}$	-0.2302	OR
$\{S_1, S_2, S_4\}$	-0.1417	XOR
$\{S_1, S_3, S_4\}$	-0.1114	Noisy XOR
$\{S_2, S_6, D_3\}$	-0.0992	Noisy XOR
$\{S_1, S_4, D_3\}$	-0.0912	Noisy XOR
$\{S_2, S_3, S_6\}$	-0.0815	XOR (broken)
$\{S_5, S_6, D_2\}$	-0.0787	OR

Table A.4: Examples of possible synergistic triplets from the EASYcare data.

Synergistic triplet (A, B, C)	Sum dyadic MI	$MI(A, B; C)$	$MI(A, C; B)$	$MI(B, C; A)$
Stiff restricted, UTI, CIRS-G Lower GI	0.0082	0.0263	0.0283	0.0244
Movement limitation, Stomachache, Mood complaints	0.0344	0.0493	0.0290	0.0388
Weight loss, Fear of fall, CIRS-G Lower GI	0.0169	0.0275	0.0276	0.0239
Fear of fall, Vertigo, Other skin condition	0.0160	0.0255	0.0290	0.0206
Miction problems, UTI, Coughing	0.0058	0.0181	0.0199	0.0160
Movement limitation, Hearing problem, CIRS-G Heart	0.0158	0.0288	0.0236	0.0210
Urinary incontinence, other skin condition, CIRS-G ENT	0.0039	0.0163	0.0148	0.0175
Pain locomotor, Miction problems, CIRS-G Psychiatric illness	0.0169	0.0198	0.0241	0.0289
Vertigo, presyncope complaints, CIRS-G Neurologic	0.0071	0.0152	0.0200	0.0177
Weight loss, sputum, CIRS-G Lower GI	0.0050	0.0177	0.0134	0.0172
Presyncope complaints, UTI, Chest pain	0.0171	0.0295	0.0169	0.0261
Strength loss, Heartburn, Anxiety complaints	0.0119	0.0243	0.0182	0.0191
Fatigue, Presyncope complaints, CIRS-G Neurologic	0.0156	0.0201	0.0230	0.0260
Walk problem, Restricted action radius, Heartburn	0.1555	0.0134	0.1679	0.1672

Synergistic triplet (A,B,C)	Sum dyadic MI	$MI(A,B;C)$	$MI(A,C;B)$	$MI(B,C;A)$
Fall, Urinary incontinence, CIRS-G Upper GI	0.0219	0.0248	0.0222	0.0330
Weight loss, Presyncope complaints, Heartburn	0.0028	0.0139	0.0146	0.0128
UTI, Heartburn, CIRS-G Endocrine and breast	0.0012	0.0128	0.0125	0.0128
Stiff restricted, other skin condition, CIRS-G Neurologic	0.0276	0.0359	0.0161	0.0383
Syncope, UTI, CIRS-G Vascular	0.0068	0.0175	0.0165	0.0144
Fear of fall, Visual problem, CIRS-G Psychiatric illness	0.0123	0.0144	0.0218	0.0232
Pain locomotor, Visual problem, CIRS-G Psychiatric illness	0.0158	0.0181	0.0215	0.0266
Strength loss, Change defecation, Hearing problem	0.0136	0.0132	0.0240	0.0247
Fatigue, Weight loss, Heartburn	0.0074	0.0114	0.0187	0.0186
Weight loss, Fall, Heartburn	0.0034	0.0116	0.0146	0.0145
Urinary incontinence, Miction problems, CIRS-G Genitourinary	0.2168	0.2265	0.0748	0.1658



SUPPLEMENTARY MATERIAL FOR CHAPTER 3

B.1. Biological Clustering and Synergistic Interactions in Lipid Metabolism

The following tables provide additional details on the biological clustering of lipid metabolism variables and the synergistic interactions among selected triplets with significant improvement factors.

Tables B.5 and B.6 complement the main analysis by detailing how specific combinations of variables contribute to a more comprehensive understanding of lipid metabolism. The improvement factor, which quantifies the ratio between the joint mutual information of two variables about the third and the sum of their pairwise mutual information, is high for the triplets listed here, indicating that these combinations carry more information about lipid homeostasis and transport mechanisms than their pairwise associations alone.

Table B.5: Biological clustering of lipid metabolism variables in the selected triplets with significant synergy and large improvement factors. Each variable is categorized based on its biological function into one of three clusters: cholesterol transport and esterification, lipoprotein particle concentration and size, and phospholipids and lipid composition. These triplets were selected based on their high synergistic interactions and large improvement factors, which highlight how certain combinations of variables provide more information about lipid metabolism than individual pairwise associations. The grouping provides insight into the roles of the variables and helps interpret their complex interactions within lipid transport.

Variable Name	**Description**	**Cluster/Grouping**
Total Concentration of Lipoprotein Particles	The total number of lipoprotein particles across all lipoprotein classes.	Lipoprotein Particle Concentration and Size
Concentration of Medium VLDL Particles	The concentration of medium-sized VLDL particles, responsible for triglyceride transport.	Lipoprotein Particle Concentration and Size
Cholesteryl Esters in Medium HDL	Esterified cholesterol stored in medium-sized HDL particles.	Cholesterol Transport and Esterification
Phospholipids in Very Large HDL	Phospholipids found in very large HDL particles, important for structural integrity.	Phospholipids and Lipid Composition
Concentration of Small HDL Particles	The concentration of small HDL particles, important in cholesterol efflux.	Lipoprotein Particle Concentration and Size
Total Lipids in Very Large HDL	Total lipid content in very large HDL particles, contributing to lipid transport.	Phospholipids and Lipid Composition
Total Cholesterol	The total amount of cholesterol in all lipoprotein particles.	Cholesterol Transport and Esterification
Total Lipids in Medium LDL	Total lipid content in medium-sized LDL particles, involved in cholesterol transport.	Phospholipids and Lipid Composition

Variable Name	Description	Cluster/Grouping
Phospholipids in Large HDL	Phospholipids found in large HDL particles, involved in cholesterol transport.	Phospholipids and Lipid Composition
Total Free Cholesterol	Unesterified cholesterol circulating in lipoproteins, available for uptake by tissues.	Cholesterol Transport and Esterification
Cholesteryl Esters in Small LDL	Esterified cholesterol stored in small LDL particles.	Cholesterol Transport and Esterification
Concentration of VLDL Particles	The concentration of VLDL particles, responsible for triglyceride transport.	Lipoprotein Particle Concentration and Size
Phospholipids in Medium VLDL	Phospholipids found in medium-sized VLDL particles, involved in lipid transport.	Phospholipids and Lipid Composition
Phospholipids in Small VLDL	Phospholipids found in small VLDL particles, important for triglyceride transport.	Phospholipids and Lipid Composition

Figure B.1 gives the full distribution of improvement factors summarised in Chapter 3, and the ten triplets with the largest improvement factors are shown in Figure B.2. This pattern implies that in lipid metabolism, higher-order interactions frequently contribute more information than the sum of individual pairwise interactions. Additionally, the extended tail of the distribution signifies that a subset of triplets possesses notably high improvement factors, emphasizing that certain interactions exhibit exceptionally high synergy. This distribution reinforces the value of exploring beyond pairwise associations to capture complex interdependencies. The top ten triplets, in terms of improvement factors, are given in Figure B.2.

B.2. Clique Algorithm Results

B.2.1. Clique Algorithm Complexity

Without a limit on the number of cliques returned, the time complexity of the clique-based algorithm depends on the size of the graph and the clique size k. After each edge addition, the algorithm examines the local neighbourhood of the two connected nodes to check whether a new k-clique has been formed. For triangle detection ($k = 3$), this requires checking $O(n)$ common neighbours per edge, and a fully connected graph has $O(n^2)$ edges, so the worst case for the whole graph is $O(n^3)$. For 4-cliques ($k = 4$), the algorithm must examine pairs of neighbours instead, giving $O(n^4)$. Thus, in the absence of any clique limit, the algorithm has a complexity of $O(n^k)$ for k-cliques, scaling with the graph's density.

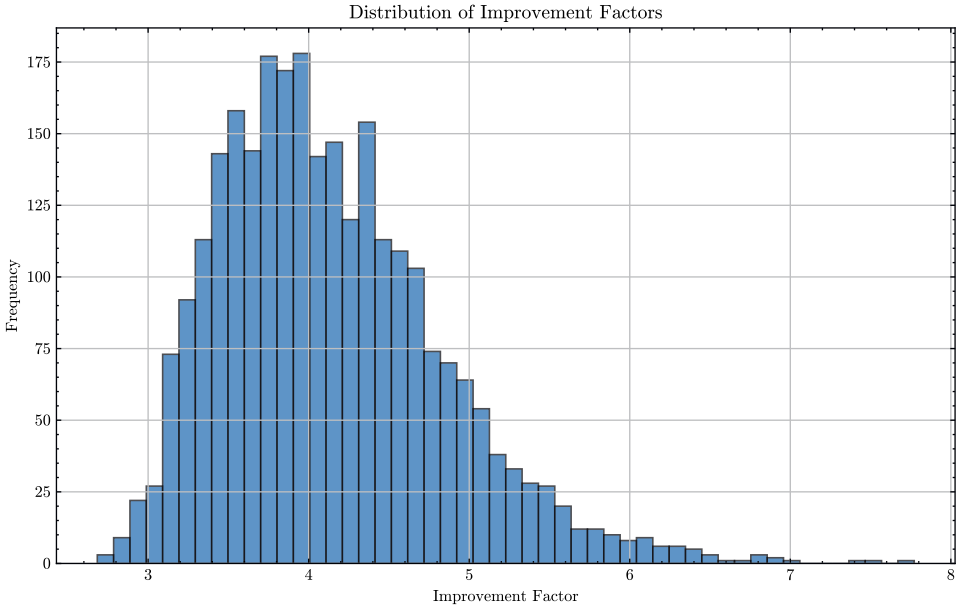


Figure B.1: Distribution of improvement factors across strongly synergistic triplets, taken to be those with bias-corrected O-information below −0.4. The improvement factor is the ratio of the joint mutual information of two variables about the third to the sum of their pairwise mutual information. Most of these triplets have improvement factors between 3 and 5, with a peak around 4 (mean 4.13, standard deviation 0.69). The tail extends towards 8, where the joint mutual information is eight times the sum of the pairwise values. The histogram therefore describes these triplets rather than the full distribution.

However, the aim of the clique algorithm is to quickly find a sample of strongly synergistic sets and so a constraint is imposed to only find the top p sets. This significantly reduces the computational complexity. For example, if $p = 1000$, the algorithm halts after identifying 1000 cliques, regardless of the total number of possible edges in the graph. For triangle detection ($k = 3$), each edge addition involves checking for new triangles in $O(n)$, so that inserting m edges before the limit of 1000 cliques is reached costs $O(m \cdot n)$. For 4-cliques ($k = 4$), each edge addition requires $O(n^2)$ operations, leading to a total complexity of $O(1000 \cdot n^2) = O(n^2)$. Therefore, with a limit of $p = 1000$, the complexity is reduced to $O(p \cdot n^{k-2})$, making the algorithm more scalable and efficient for large graphs.

B.3. PSO Training

Figure B.3 compares the initial and final O-information scores for each particle in the PSO algorithm under a single parameter configuration. Each particle begins with an initial score (shown in blue) and progresses towards a lower final score (shown in red) as the algorithm iterates. The overall trend of decreasing scores suggests that PSO effectively

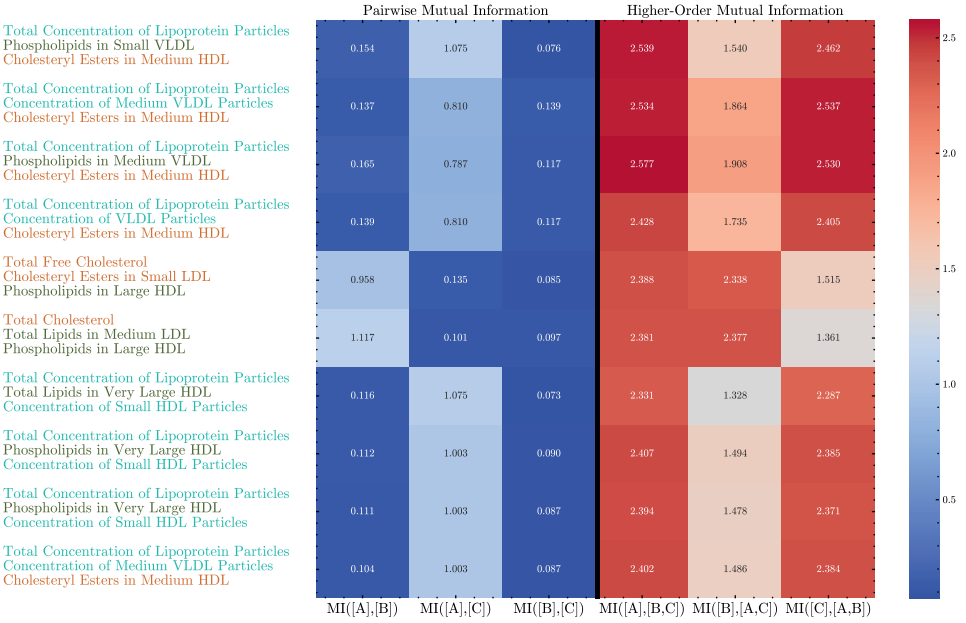


Figure B.2: Heatmap of pairwise and higher-order mutual information (MI) for lipid metabolism triplets with significant synergy and large improvement factors, categorized by biological function. The heatmap presents both the pairwise MI (columns 1–3) and the higher-order MI (columns 4–6) for each triplet. Variables are colour-coded according to their biological cluster, showing that these triplets exhibit strong synergistic interactions and large improvement factors. The analysis reveals that interactions between variables from different clusters, such as lipoprotein particle concentration and cholesterol transport, exhibit stronger synergies, suggesting coordinated regulation of lipid transport and metabolism.

guides particles to explore lower O-information regions, thereby enhancing optimisation performance. Notably, the distribution of final scores below the initial scores shows that most particles improved on their starting positions under this configuration; this is consistent with an ability to escape local minima, a common limitation observed in SA, although that comparison rests on the distributions of best scores obtained by the two algorithms rather than on this figure. This capacity for avoiding local optima is critical for accurately identifying synergistic interactions, further validating PSO as a suitable method for exploring complex interaction landscapes.

B.4. SA Training over Epochs

Figure B.4 presents the training trajectories of three distinct runs of the SA algorithm, highlighting its performance and sensitivity within the optimisation landscape for identifying synergistic triplets. Panel (a) shows a run in which no improvement is achieved. The lack of progress indicates the algorithm's difficulty in escaping local minima in this

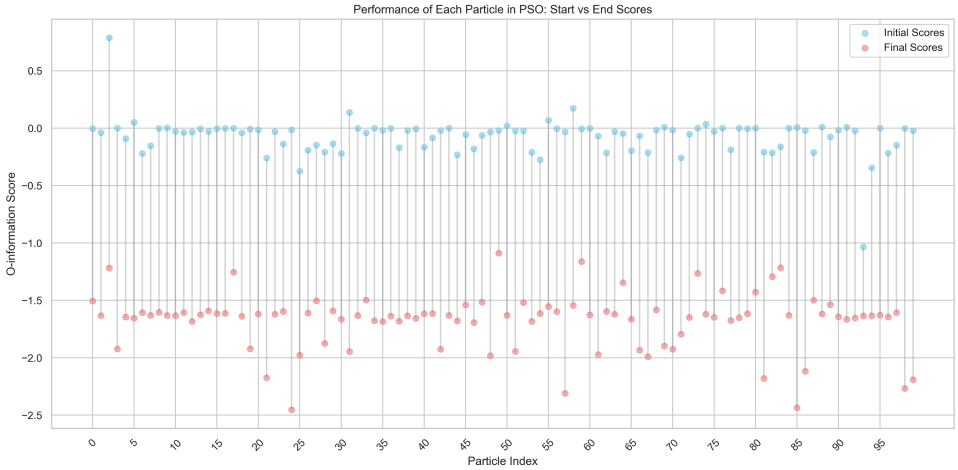
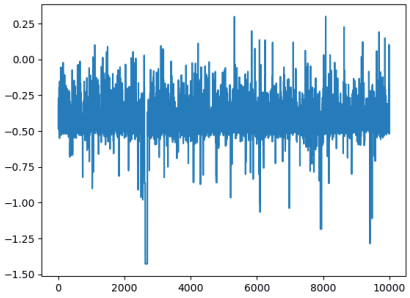


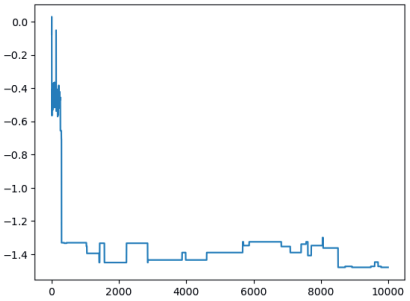
Figure B.3: Initial and final scores for each particle in the PSO algorithm for one parameter setting. The plot shows that the particles generally improve their positions, moving towards lower O-information scores, indicating that the PSO algorithm effectively prevents particles from getting stuck in local minima compared to the SA approach.

particular run. Panel (b) shows the SA initially struggles to find an optimal path, leading to high score variability. However, after approximately 4000 steps, a gradual improvement emerges, enabling the algorithm to converge on a highly synergistic triplet. In panel (c), the SA algorithm rapidly identifies a highly synergistic triplet within the first few thousand steps and maintains this low score with minimal fluctuations, demonstrating a successful escape from local minima.

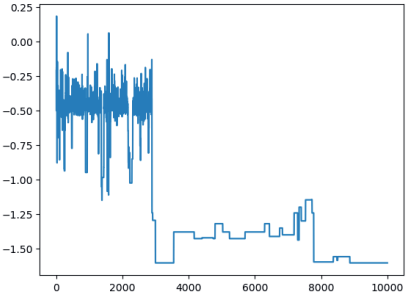
The three runs use the same configuration with different random initialisations, and their outcomes range from no improvement at all to convergence within the first few thousand steps. A single run therefore says little about SA's expected performance. Chapter 3 accordingly reports SA results over 100 independent runs per configuration rather than for one run.



(a) No significant improvement: In this plot, the SA algorithm remains trapped in a region of suboptimal solutions throughout the run, failing to find any highly synergistic triplets. The scores fluctuate around the same range, indicating that the algorithm is unable to escape local minima in this particular run.

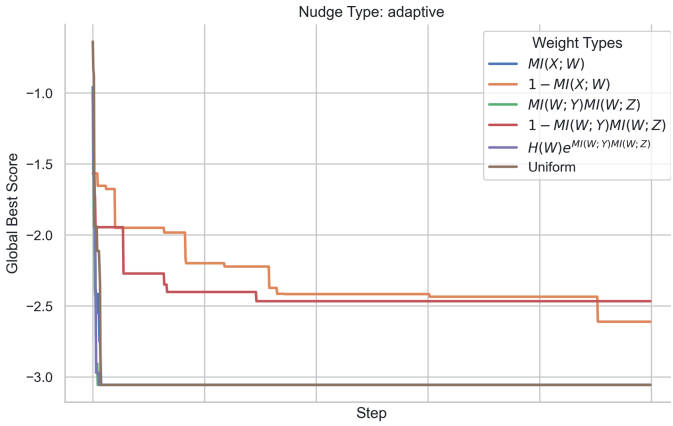


(b) Slow initial improvement: This plot demonstrates a scenario where the SA algorithm initially struggles to find a good solution, resulting in high variability in scores. However, after approximately 4000 steps, the algorithm starts to improve steadily, eventually finding a highly synergistic triplet.

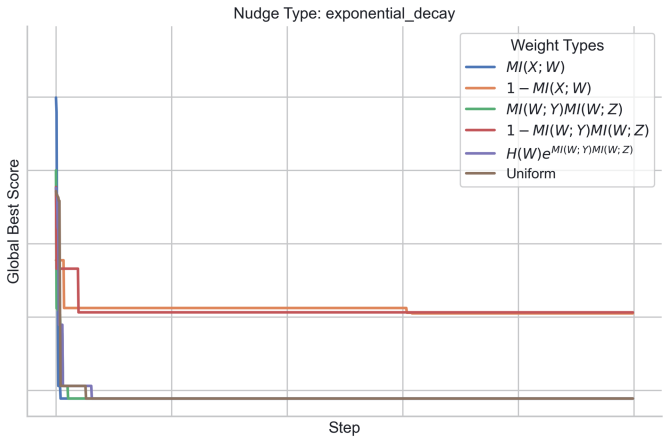


(c) Quick improvement: This plot shows a trajectory where the SA algorithm rapidly finds a highly synergistic triplet within the first few thousand steps and maintains the low score with minor fluctuations.

Figure B.4: Training trajectories of three different runs of the SA algorithm. These plots illustrate the performance and challenges faced by the SA in navigating the optimisation landscape. The variability in the trajectories indicates the algorithm's sensitivity to initial conditions and potential difficulties in escaping local minima.

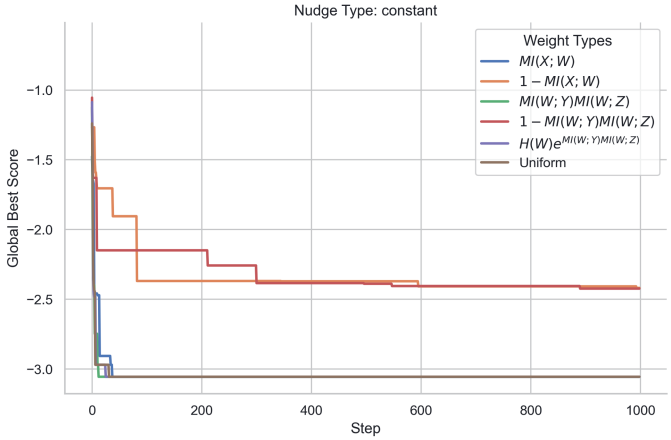


(a) Nudge Type: none

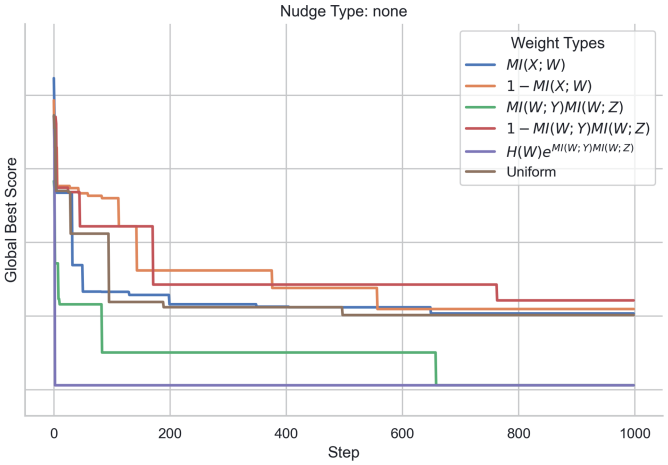


(b) Nudge Type: constant

Figure B.5: Convergence rates of the PSO algorithm with nudge types none and constant. The plots illustrate how the global best score evolves over time for various weight types used in the SA algorithm.



(a) Nudge Type: exponential decay



(b) Nudge Type: adaptive

Figure B.6: Convergence rates of the PSO algorithm with nudge types exponential decay and adaptive. The plots illustrate how the global best score evolves over time for various weight types used in the SA algorithm.

Features	RMSE
Mutual Information Sum, Mutual Information Spread, Minimum Mutual Information, Maximum Mutual Information, Variation of Information, Normalized Variation of Information	0.016931
Mutual Information Spread, Minimum Mutual Information, Maximum Mutual Information, Variation of Information, Normalized Variation of Information, Information Overlap	0.016965
Mutual Information Spread, Minimum Mutual Information, Maximum Mutual Information, Variation of Information, Information Overlap, Normalized Information Overlap	0.016986
Mutual Information Sum, Mutual Information Spread, Maximum Mutual Information, Variation of Information, Normalized Variation of Information, Information Overlap	0.016988
Mutual Information Sum, Mutual Information Spread, Minimum Mutual Information, Maximum Mutual Information, Variation of Information, Information Overlap	0.017115
Mutual Information Sum, Mutual Information Spread, Minimum Mutual Information, Maximum Mutual Information, Normalized Variation of Information, Information Overlap	0.017143
Mutual Information Spread, Minimum Mutual Information, Maximum Mutual Information, Normalized Variation of Information, Information Overlap, Normalized Information Overlap	0.017433
Mutual Information Sum, Mutual Information Spread, Minimum Mutual Information, Maximum Mutual Information, Variation of Information, Information Overlap, Normalized Information Overlap	0.017512
Mutual Information Spread, Minimum Mutual Information, Maximum Mutual Information, Variation of Information, Information Overlap, Normalized Information Overlap	0.017532
Mutual Information Sum, Mutual Information Spread, Maximum Mutual Information, Variation of Information, Information Overlap, Normalized Information Overlap	0.017534

Table B.1: Feature Combinations and respective RMSE Scores (Part 1) of the LightGBM model for predicting PID-based synergy scores in the UK Biobank NMR data. Each row gives the pairwise metrics used as training features and the RMSE of the model trained on them. Parts 1 to 4 report 42 distinct feature sets of one to seven metrics, ordered by increasing RMSE. Some sets were evaluated more than once and therefore appear on more than one row. The division into four parts is for page fit only. The lowest RMSE here is 0.0169, against 0.009 for the best model in Chapter 3, which used more features.

Features	RMSE
Mutual Information Sum, Minimum Mutual Information, Maximum Mutual Information, Variation of Information, Normalized Variation of Information, Information Overlap	0.017590
Mutual Information Sum, Minimum Mutual Information, Maximum Mutual Information, Variation of Information, Information Overlap, Normalized Information Overlap	0.017590
Mutual Information Sum, Mutual Information Spread, Minimum Mutual Information, Maximum Mutual Information, Variation of Information, Normalized Variation of Information, Information Overlap	0.017601
Mutual Information Sum, Mutual Information Spread, Minimum Mutual Information, Maximum Mutual Information, Variation of Information, Normalized Variation of Information, Information Overlap	0.017616
Minimum Mutual Information, Variation of Information, Normalized Variation of Information, Information Overlap, Normalized Information Overlap	0.017633
Mutual Information Sum, Minimum Mutual Information, Maximum Mutual Information, Normalized Variation of Information, Information Overlap, Normalized Information Overlap	0.017676
Minimum Mutual Information, Maximum Mutual Information, Normalized Variation of Information, Information Overlap, Normalized Information Overlap	0.017677
Mutual Information Sum, Minimum Mutual Information, Maximum Mutual Information, Variation of Information, Information Overlap, Normalized Information Overlap	0.017678
Mutual Information Spread, Minimum Mutual Information, Maximum Mutual Information, Variation of Information, Normalized Variation of Information, Information Overlap	0.017925
Mutual Information Sum, Mutual Information Spread, Minimum Mutual Information, Maximum Mutual Information, Variation of Information, Information Overlap, Normalized Information Overlap	0.017952

Table B.2: Feature Combinations and respective RMSE Scores (Part 2) of the LightGBM model for predicting PID-based synergy scores in the UK Biobank data.

Features	RMSE
Mutual Information Sum, Mutual Information Spread, Minimum Mutual Information, Maximum Mutual Information, Variation of Information, Normalized Variation of Information, Information Overlap	0.017992
Mutual Information Spread, Minimum Mutual Information, Maximum Mutual Information, Variation of Information, Information Overlap, Normalized Information Overlap	0.018010
Mutual Information Sum, Minimum Mutual Information, Maximum Mutual Information, Normalized Variation of Information, Information Overlap	0.053983
Mutual Information Spread, Information Overlap	0.054077
Mutual Information Sum, Mutual Information Spread	0.055605
Mutual Information Sum, Mutual Information Spread	0.055606
Mutual Information Spread, Variation of Information, Information Overlap	0.055612
Mutual Information Sum, Variation of Information, Information Overlap	0.059347
Mutual Information Sum, Avg_Corr	0.059360
Variation of Information, Avg_Corr	0.059373

Table B.3: Feature Combinations and respective RMSE Scores (Part 3) of the LightGBM model for predicting PID-based synergy scores in the UK Biobank data.

B

Features	RMSE
Maximum Mutual Information, Avg_Corr	0.059897
Mutual Information Spread, Avg_Corr	0.060684
Normalized Information Overlap, Avg_Corr	0.060902
Information Overlap, Avg_Corr	0.061232
Minimum Mutual Information, Normalized Information Overlap	0.061695
Minimum Mutual Information, Variation of Information, Normalized Information Overlap	0.062058
Mutual Information Sum, Minimum Mutual Information, Normalized Information Overlap	0.062059
Mutual Information Sum, Minimum Mutual Information, Normalized Information Overlap	0.062059
Minimum Mutual Information, Information Overlap	0.062973
Mutual Information Spread, Minimum Mutual Information	0.063003
Minimum Mutual Information, Maximum Mutual Information	0.063024
Mutual Information Spread, Minimum Mutual Information	0.063055
Mutual Information Spread, Maximum Mutual Information	0.063528
Mutual Information Sum	0.065725
Variation of Information	0.065725
Mutual Information Sum, Variation of Information	0.065725
Minimum Mutual Information	0.067377
Avg_Corr	0.067407
Maximum Mutual Information	0.067513
Mutual Information Spread	0.067699
Normalized Information Overlap	0.067820
Information Overlap	0.070113

Table B.4: Feature Combinations and respective RMSE Scores (Part 4) of the LightGBM model for predicting PID-based synergy scores in the UK Biobank data.

Variable Name	Description	Cluster/Grouping
HDL Cholesterol	Cholesterol transported by HDL, involved in reverse cholesterol transport.	Cholesterol Transport and Esterification
Cholesteryl Esters in HDL	Esterified cholesterol in HDL, stored for transport to the liver.	Cholesterol Transport and Esterification
Concentration of VLDL Particles	Number of VLDL particles transporting triglycerides and cholesterol.	Lipoprotein Particle Concentration and Size
Total Phospholipids	Total amount of phospholipids in all lipoprotein particles.	Phospholipids and Lipid Composition
Cholesteryl Esters in Very Large VLDL	Esterified cholesterol in large VLDL particles, linked to lipid transport.	Cholesterol Transport and Esterification
Total Lipids in Small HDL	Total lipid content in small HDL particles, involved in cholesterol efflux.	Apolipoproteins and Small HDL Function
Apolipoprotein A1	Main protein component of HDL, critical for HDL formation and function.	Apolipoproteins and Small HDL Function
Concentration of VLDL	Concentration of VLDL particles, related to lipid transport.	Lipoprotein Particle Concentration and Size
Free Cholesterol in VLDL	Free, unesterified cholesterol in VLDL particles.	Other Metabolites
Total Cholines	Choline-containing compounds, precursors for phospholipid synthesis.	Other Metabolites
Average Diameter for HDL	Average size of HDL particles, affects cholesterol transport efficiency.	Lipoprotein Particle Concentration and Size
Phospholipids in VLDL	Phospholipids specifically in VLDL, affecting its structure and function.	Phospholipids and Lipid Composition
Cholesteryl Esters in Medium HDL	Esterified cholesterol in medium HDL particles, contributing to cholesterol transport.	Cholesterol Transport and Esterification

Table B.6: Biological clustering of lipid metabolism variables in the ten most synergistic triplets as per the bias-corrected O-information metric. Each variable is categorized based on its biological function into one of five clusters: cholesterol transport and esterification, lipoprotein particle concentration and size, phospholipids and lipid composition, apolipoproteins and small HDL function, or other metabolites. These groupings provide insight into the functional roles of the variables and help interpret their interactions in lipid metabolism.

Table B.7: Summary of PID synergy statistics for different strategies (UKB NMR data). Each strategy grows the pairwise network edge by edge in the order given by one pairwise metric. Each row summarises the PID synergy scores of the cliques returned. The metrics are defined in Chapter 3, where the strategy labelled MI2 here is called MI.

Strategy	Mean	Median	Std Dev	IQR
MI_ent	0.0164	0.0161	0.0039	0.0051
Inverse_MI	0.0164	0.0161	0.0039	0.0051
CH	0.0164	0.0161	0.0039	0.0051
Inverse_MI_squared	0.0164	0.0161	0.0039	0.0051
Alternating MI	0.0768	0.0541	0.0723	0.1226
MI2	0.1381	0.1334	0.0577	0.0738

Table B.8: Proportion of PID Synergy Scores Above Ground Truth Percentiles for Different Strategies (UKB NMR Data)

Strategy	Above 50th	Above 75th	Above 90th	Above 95th
MI_ent	0.00%	0.00%	0.00%	0.00%
Inverse_MI	0.00%	0.00%	0.00%	0.00%
CH	0.00%	0.00%	0.00%	0.00%
Inverse_MI_squared	0.00%	0.00%	0.00%	0.00%
Alternating MI	50.63%	39.60%	14.34%	4.62%
MI2	95.86%	75.00%	30.00%	11.08%

Table B.9: Summary of PID Synergy Statistics for Different Strategies (YFS Data)

Strategy	Mean	Median	Std Dev	IQR
MI_ent	0.0304	0.0301	0.0050	0.0068
Inverse_MI	0.0304	0.0301	0.0050	0.0068
CH	0.0304	0.0301	0.0050	0.0068
Inverse_MI_squared	0.0304	0.0301	0.0050	0.0068
Alternating MI	0.1534	0.1470	0.0785	0.0985
MI2	0.1622	0.1526	0.0736	0.0928

Table B.10: Proportion of PID Synergy Scores Above Ground Truth Percentiles for Different Strategies (YFS Data)

Strategy	Above 50th	Above 75th	Above 90th	Above 95th
MI_ent	0.14%	0.00%	0.00%	0.00%
Inverse_MI	0.15%	0.00%	0.00%	0.00%
CH	0.14%	0.00%	0.00%	0.00%
Inverse_MI_squared	0.15%	0.00%	0.00%	0.00%
Alternating MI	90.62%	87.35%	79.03%	68.86%
MI2	96.40%	92.95%	84.30%	73.76%

Table B.11: Summary of Statistics for Different Strategies Using the O-information Measure (UKB NMR Data). Scores are negated O-information, so that higher values indicate greater synergy.

Strategy	Mean	Median	Std Dev	IQR
MI_ent	0.4904	0.4902	0.0252	0.0343
Inverse_MI	0.4904	0.4902	0.0252	0.0343
CH	0.4904	0.4902	0.0252	0.0343
Inverse_MI_squared	0.4904	0.4902	0.0252	0.0343
Alternating MI	-0.9126	-1.3427	0.8532	1.6449
MI2	-1.4117	-1.3430	0.2502	0.2890

Table B.12: Proportion of Scores Above Ground Truth Percentiles for Different Strategies Using the O-information Measure (UKB NMR Data)

Strategy	Above 50th	Above 75th	Above 90th	Above 95th
MI_ent	99.63%	79.74%	38.96%	17.33%
Inverse_MI	99.63%	79.75%	38.96%	17.34%
CH	99.63%	79.74%	38.96%	17.33%
Inverse_MI_squared	99.63%	79.75%	38.96%	17.34%
Alternating MI	6.30%	5.13%	2.56%	1.12%
MI2	0.00%	0.00%	0.00%	0.00%

Table B.13: Summary of Statistics for Different Strategies Using the O-information Measure (YFS Data)

Strategy	Mean	Median	Std Dev	IQR
MI_ent	0.2496	0.2491	0.0224	0.0300
Inverse_MI	0.2496	0.2491	0.0224	0.0300
CH	0.2496	0.2491	0.0224	0.0300
Inverse_MI_squared	0.2496	0.2491	0.0224	0.0300
Alternating MI	-0.2765	-0.2886	0.2725	0.2300
MI2	-0.3390	-0.2910	0.1886	0.2006

Table B.14: Proportion of Scores Above Ground Truth Percentiles for Different Strategies Using the O-information Measure (YFS Data)

Strategy	Above 50th	Above 75th	Above 90th	Above 95th
MI_ent	59.43%	29.55%	11.73%	5.52%
Inverse_MI	59.42%	29.56%	11.73%	5.52%
CH	59.43%	29.55%	11.73%	5.52%
Inverse_MI_squared	59.42%	29.56%	11.73%	5.52%
Alternating MI	1.37%	0.37%	0.10%	0.05%
MI2	0.00%	0.00%	0.00%	0.00%

B

Table B.15: Summary statistics for different SA configurations on the YFS data, discretised by quantile binning with the number of bins set by Sturges' rule. Synergy was computed via the O-information, and every configuration used geometric cooling with random initialisations across 100 runs. Only the mean and median best score are reported here. The standard deviation, interquartile range and the number of runs reaching the synergy threshold are given for the UKB NMR data in Table 3.1.

Configuration	**Mean**	**Median**
Inverse_MI_geometric	-0.4867	-0.4734
MI_ent_geometric	-0.4780	-0.4648
no_weights_geometric	-0.4827	-0.4635
MI_geometric	-0.5134	-0.5081
Inverse_MI_squared_geometric	-0.4862	-0.4674



SUPPLEMENTARY MATERIAL FOR CHAPTER 4

C.1. Extended methods: measurement of the symptoms and signs in LASA

The six symptoms included:

- Depressive symptoms were assessed with the Center for Epidemiologic Studies Depression Scale (CES-D), with scores of 16 and higher indicating presence of depressive symptoms [242]. Chapter 5 dichotomises the same scale at 20 rather than 16. It also sets the HADS anxiety subscale at 4 and mean arterial pressure at 93 mmHg, against the 8 and 136.5 mmHg used here. The thresholds in this appendix mark presence, because each symptom and sign enters the analysis only through a domain count. Chapter 5 instead models each variable in its own right, so the two sets of cut-points are not interchangeable.
- Anxiety symptoms were identified using the anxiety subscale of the Hospital Anxiety and Depression Scale (HADS), with scores of 8 or greater indicating presence of clinically relevant anxiety symptoms [172].
- Sleep problems (range 3-12) were derived from three categorical questions asking about problems falling asleep, waking through the night and waking up too early. Responses were summed, and sleep problems were present if participants scored positively on at least one of these three questions, that is, for a summed score above the scale minimum of 3 (no problems).
- Pain severity was measured using a subscale of the Nottingham Health Profile (range 5–10) and dichotomised as score of 5 (no pain) versus > 5 (pain present) [243].
- Hearing complaints were coded as present if any difficulty or inability to hear was reported.
- Memory complaints were determined by a yes/no response to the question: “Do you have complaints about your memory?”.

The five signs included:

- Mean arterial pressure (MAP) was calculated as the average of the diastolic and systolic blood pressure measured with the automatic Omron device (Omron HEM 815F). MAP was dichotomised using a threshold of 136.5 mmHg, the midpoint between normal values and a clearly hypertensive value of 170 mmHg. This cut-point is study-specific rather than guideline-defined. The European guidelines define hypertension as an office blood pressure of ≥ 140/90 mmHg [244].
- Grip strength was measured using a hand grip dynamometer (Takei TKK 5001, Takei Scientific Instruments Co. Ltd). Sex-specific thresholds (<27 kg for men, <16 kg for women) established by the EWGSOP2 consensus were used to identify individuals with low muscle strength [169].
- Body mass index (BMI) was calculated based on measured height and weight and dichotomised as obese (BMI≥30) versus non-obese (BMI<30) [245].

- The waist-hip ratio (WHR) was calculated as the waist circumference divided by hip circumference measured using standardised procedures. WHR was dichotomised using sex-specific cutoffs of ≥ 0.90 in men and ≥ 0.85 in women to indicate abdominal obesity [246].
- Cognitive functioning was measured with the Mini-Mental State Examination (MMSE), with scores of 23 or lower indicating impairment [247].

Table C.1: Sensitivity analyses repeating the partial information decomposition (PID) models using data from waves 1995/96 and 2021/22.

Domain	Lost Information	Unique Information	MI with Functional Limitations	Pure Synergy	Synergy Bounds (%)
Total sample					
Wave 1995/96: entropy=1.842, total MI=0.301					
Signs	19.34% (0.058 bits)	0%	1.50% (0.028 bits)	19.34% (0.058 bits)	19.34–29.74% (0.058 - 0.090 bits)
Symptoms	48.14% (0.145 bits)	0.37% (0.001 bits)	8.54% (0.157 bits)	37.37% (0.112 bits)	47.78 - 47.78% (0.144 - 0.144 bits)
Diseases	37.37% (0.112 bits)	0%	7.02% (0.129 bits)	37.37% (0.112 bits)	37.37 - 47.78% (0.112 - 0.144 bits)
Wave 2021/22: entropy=1.901, total MI=0.349					
Signs	28.80% (0.101 bits)	0%	2.79 % (0.053 bits)	28.80% (0.101 bits)	28.80 – 45.71% (0.101 - 0.159 bits)
Symptoms	40.12% (0.140 bits)	0%	6.41% (0.122 bits)	40.12% (0.140 bits)	40.12 – 57.03% (0.140 - 0.199 bits)
Diseases	46.48% (0.162 bits)	0%	7.89% (0.150 bits)	40.12% (0.140 bits)	46.48 – 57.03% (0.162 - 0.199 bits)
Multimorbidity subgroup					
Wave 1995/96: entropy=1.972, total MI=0.277					
Signs	28.73% (0.080 bits)	0%	1.16% (0.023 bits)	28.73% (0.080 bits)	28.73 - 40.21% (0.080 - 0.111 bits)
Symptoms	69.41% (0.192 bits)	28.88% (0.080 bits)	8.35% (0.165 bits)	28.73% (0.080 bits)	40.21% - 40.21% (0.111 - 0.111 bits)
Diseases	23.29% (0.064 bits)	0%	2.67% (0.053 bits)	23.29% (0.064 bits)	23.29% - 35.12% (0.064 - 0.097 bits)
Wave 2021/22: entropy=1.997, total MI=0.294					
Signs	39.83% (0.117 bits)	0%	4.29% (0.086 bits)	39.83% (0.117 bits)	39.83% - 58.91% (0.117 - 0.173 bits)

Domain	Lost Information	Unique Information	MI with Functional Limitations	Pure Synergy	Synergy Bounds (%)
Symptoms	52.76% (0.155 bits)	0%	6.06% (0.121 bits)	39.83% (0.117 bits)	52.76% - 58.91% (0.156 - 0.173 bits)
Diseases	23.75% (0.070 bits)	0%	3.34% (0.067 bits)	23.75% (0.070 bits)	23.75% - 42.84% (0.070 - 0.126 bits)

Mutual information (MI) quantifies how much variability in the outcome functional limitations is captured by information on symptoms, signs and diseases. Total mutual information captures the combined predictive power of all three domains.
Lost information = the information lost if a specific domain is excluded.
Unique information = the amount of predictive power that a single domain (signs, symptoms or diseases) contributes on its own, independent of the other domains. If unique information is zero, it means every insight a given domain offers about functional limitations can be gleaned equally from other domains or only arises in combination with them.
Pure Synergy = the amount of information gained only when that domain is paired with at least one other domain.
Synergy Lower Bound = The guaranteed, minimal extra information from combinations involving that domain—the information forgone if it never interacts with any other domain. Computed as the lost information minus the unique information.
Synergy Upper Bound = The maximal possible extra signal that could come from any combination of that domain with one, two, or all three of the other domains. It shows the ceiling on how much interaction-driven information that domain might contribute.
Percentages indicate the relative contribution of each domain in explaining functional limitations compared to the total entropy of the outcome, taken as 100%. Higher percentages indicate a larger contribution.

Table B. Sensitivity analyses repeating the Random Forest models using data from waves 1995/96 and 2021/22

	Diseases only Mean Cohen's κ (CI)	Diseases + Signs Mean Cohen's κ (CI)	Diseases + Signs + Symptoms Mean Cohen's κ (CI)
Wave 1995/96			
Total sample	0.302 (0.225–0.405)	0.304 (0.237–0.408)	0.443 (0.341–0.524)
Multimorbidity subgroup	0·166 (0·053–0·320)	0·169 (0·023–0·343)	0·370 (0·217–0·479)
Wave 2021/22			

	Diseases only Mean Cohen's κ (CI)	Diseases + Signs Mean Cohen's κ (CI)	Diseases + Signs + Symptoms Mean Cohen's κ (CI)
Total sample	0·282 (0·217–0·340)	0·325 (0·242–0·408)	0·525 (0·477–0·575)
Multimorbidity subgroup	0·241 (0·102–0·352)	0·291 (0·164–0·400)	0·510 (0·406–0·593)
	Symptoms only Mean Cohen's κ (CI)	Symptoms + Signs Mean Cohen's κ (CI)	Symptoms + Signs + Diseases Mean Cohen's κ (CI)
Wave 1995/96			
Total sample	0·415 (0·321–0·506)	0·414 (0·322–0·501)	0·443 (0·341–0·524)
Multimorbidity subgroup	0·369 (0·215–0·464)	0·381 (0·246–0·492)	0·370 (0·217–0·479)
Wave 2021/22			
Total sample	0·494 (0·439–0·560)	0·503 (0·448–0·560)	0·525 (0·477–0·575)
Multimorbidity subgroup	0·464 (0·402–0·525)	0·447 (0·371–0·521)	0·510 (0·406–0·593)
	Signs only Mean Cohen's κ (CI)	Signs + Symptoms Mean Cohen's κ (CI)	Signs + Symptoms + Diseases Mean Cohen's κ (CI)
Wave 1995/96			
Total sample	0·092 (0·053–0·142)	0·414 (0·322–0·501)	0·443 (0·341–0·524)
Multimorbidity subgroup	0·131 (0·050 – 0·216)	0·381 (0·246 – 0·492)	0·370 (0·217 – 0·479)
Wave 2021/22			
Total sample	0·077 (0·044 – 0·114)	0·503 (0·448 – 0·560)	0·525 (0·477 – 0·575)
Multimorbidity subgroup	0·059 (0·032 – 0·091)	0·447 (0·371 – 0·521)	0·510 (0·406 – 0·593)

In the original analyses, to increase the number of observations available for the partial information decomposition (PID) and Random Forest (RF) models, data from waves 1995/96 to 2021/22 were treated as one cross-sectional dataset, not accounting for the within-person dependence of observations given the repeated measurements. In these sensitivity analyses, we reran the models using data from Waves 1995/96 (earliest included data collection wave) and 2021/22 (latest included data collection wave) to examine if not accounting for this dependency in observations affected the findings. The RF estimates

remained similar, which suggests little bias. The PID estimates agreed in direction rather than in magnitude: pure synergy was higher for every domain in both single waves than in the original analyses. Symptoms provided 28.88% unique information in the 1995/96 multimorbidity subgroup and none in 2021/22, against 34.13% in the pooled analysis. That non-replication should temper the strength of the unique-information claim. The wider confidence intervals reflect the reduced number of data points and thus a smaller effective sample size.



SUPPLEMENTARY MATERIAL FOR CHAPTER 5

This appendix collects the supplementary material for Chapter 5. Section D.1 reports the permutation screening of joint mutual information by stratum, the practical constraints on the estimates, and the robustness of the BUST breadth and uniformity indices (Section 5.5.6) to alternative deviation-from-uniformity measures and redundancy definitions. Section D.4 sets the estimated synergies against additive and multiplicative baselines through residual maps for a synergistic and a non-synergistic triplet. Sections D.5 and D.6 document the merging of sparse variable states and the rule-based analysis of joint-state behaviour behind the constellation results. Sections D.7–D.9 report the simulation validation of the estimation pipeline (Section 5.5.8), the sensitivity of BUST scores to discretisation (Section 5.5.9), and bootstrap confidence intervals for the highest-ranked pairs.

D.1. Joint Mutual Information Significance in LASA

Table D.1: Number of source–source–target triplets with significant joint mutual information (FDR $q = 0.05$) by stratum.

Stratum	Significant / total	Percent significant	FDR critical p
AgeGroup_0	1799/2040	88.2%	0.0439
AgeGroup_1	1901/2040	93.2%	0.0459
AgeGroup_2	1957/2040	95.9%	0.0479
AgeGroup_3	1750/2040	85.8%	0.0419
Multimorbidity_0	1930/2040	94.6%	0.0439
Multimorbidity_1	1917/2040	94.0%	0.0439
Sex_0	1978/2040	97.0%	0.0479
Sex_1	1901/2040	93.2%	0.0459
Full population	2,005/2,040	98.3%	0.0459

D.2. Estimation caveats and scope of the PID analysis

Three practical constraints qualify the estimates reported here. First, the synergy estimates depend on how the variables are binned, and reliable estimation requires adequate expected cell counts, with most triplets carrying ≥ 5–10 observations per joint cell. Section D.8 reports how far the BUST scores and pair rankings move under alternative discretisation schemes. Second, PID values are estimated from finite samples, so they are subject to noise. We therefore screened every triplet by permutation testing and estimated the PID atoms with bias correction. The screen covered 2,040 triplets in each stratum, so the resulting p-values were adjusted for multiplicity by controlling the false discovery rate at $q = 0.05$ (Table D.1). Finally, we restrict our analysis to *pairs of predictors* (bivariate PID). Extending PID to three or more predictors is theoretically more complex, because the redundancy lattice grows rapidly with the number of sources and multivariate redundancy has no agreed definition [22]. It is also impractical for the present sample, where reaching a minimum three-way cell count of five already required merging sparse states (Section D.5). Four-way tables would be sparser still.

D.3. Robustness of the BUST metrics

D.3.1. Robustness to the choice of deviation-from-uniformity measure

The main BUST map (Section 5.5.6) uses entropy-based measures of breadth and uniformity: normalised Shannon entropy of synergy shares across targets for breadth (H_B), and normalised entropy of synergy fractions across strata for uniformity (H_U). To assess whether our conclusions depend on this particular deviation-from-uniformity metric, we repeated all calculations using two alternative families of indices applied to the *same* synergy-based probability distributions: (i) a Gini-based measure, and (ii) a Kolmogorov–Smirnov (KS)-based measure.

Gini-based indices. Let $\mathbf{p} = (p_1, \ldots, p_n)$ denote a non-negative weight vector with $\sum_i p_i = 1$, representing either synergy shares across targets or stabilised synergy fractions across strata (Section 5.5.6). Let $p_{(1)} \le \cdots \le p_{(n)}$ denote the components sorted in ascending order. We compute the standard Gini coefficient

$$G(\mathbf{p}) = \frac{2}{n} \frac{\sum_{i=1}^{n} i\, p_{(i)}}{\sum_{i=1}^{n} p_{(i)}} - \frac{n+1}{n},$$

which lies in $[0, (n-1)/n]$ for a probability vector. To place breadth and uniformity on a $[0,1]$ scale where larger values indicate greater evenness (rather than inequality), we first normalise by the maximal value $G_{\max} = (n-1)/n$ and define

$$G_{\text{norm}}(\mathbf{p}) = \frac{G(\mathbf{p})}{G_{\max}}, \qquad \tilde{E}_{\text{Gini}}(\mathbf{p}) = 1 - G_{\text{norm}}(\mathbf{p}) \in [0,1].$$

We then substitute $\tilde{E}_{\text{Gini}}$ in place of H_B or H_U and apply the same median-centring to $[-1,1]$ as in Chapter 5, yielding Gini-based breadth and uniformity scores B_{Gini} and U_{Gini}.

KS-based indices. As a second alternative, we quantify deviation from uniformity using a discrete KS distance. For a probability vector $\mathbf{p}$ (as above), we sort components in *descending* order, $p_{[1]} \ge \cdots \ge p_{[n]}$, and define their cumulative sums $F_{\mathbf{p}}(k) = \sum_{i=1}^{k} p_{[i]}$ for $k = 1, \ldots, n$. The corresponding uniform cumulative distribution is $F_{\text{unif}}(k) = k/n$. The (one-sample) KS distance from uniformity is

$$D(\mathbf{p}) = \max_{1 \le k \le n} \left| F_{\mathbf{p}}(k) - \frac{k}{n} \right|,$$

which attains its maximum possible value $D_{\max} = 1 - 1/n$ for a probability vector concentrated entirely on a single category. We normalise and invert this to obtain

$$D_{\text{norm}}(\mathbf{p}) = \frac{D(\mathbf{p})}{D_{\max}}, \qquad \tilde{E}_{\text{KS}}(\mathbf{p}) = 1 - D_{\text{norm}}(\mathbf{p}) \in [0,1],$$

and again substitute $\tilde{E}_{\text{KS}}$ for H_B or H_U, followed by the same median-centring to generate KS-based breadth and uniformity scores B_{KS} and U_{KS}.

Empirical agreement of entropy-, Gini-, and KS-based metrics. We computed entropy-, Gini-, and KS-based breadth and uniformity metrics for all source pairs with non-zero synergy in the full population ($N = 135$ pairs with complete data). The three breadth indices were highly correlated: pairwise Pearson correlations between H_B, $\tilde{E}_{\text{Gini}}$, and $\tilde{E}_{\text{KS}}$ ranged from $r = 0.93$ to $r = 0.97$, and Spearman rank correlations from $\rho = 0.95$ to $\rho = 0.99$. Uniformity indices showed even stronger agreement: pairwise Pearson correlations between H_U, $\tilde{E}_{\text{Gini}}$, and $\tilde{E}_{\text{KS}}$ were all ≥ 0.98, with Spearman correlations ≥ 0.97.

The three families of indices differ only in the functional form of the deviation-from-uniformity measure; they are applied to the same synergy-based probability distributions, so this check bears on the choice of measure rather than on the synergy estimates themselves. Because the three families nonetheless show near-perfect agreement in the resulting rankings, our substantive conclusions about which pairs are broad vs. narrow and uniform vs. specific, and about the overall structure of the BUST map and network, are robust to the particular choice of deviation-from-uniformity measure.

D.3.2. Robustness to the choice of redundancy definition

To verify that our findings are not driven by the choice of redundancy measure, we repeated the full PID estimation for all source–target triplets in the full population using three definitions: $I_{\min}$ (Williams–Beer; primary) [22], I_{mmi} (minimum mutual information), and BROJA [105]. Spearman rank correlations of total synergy strength across all 136 source pairs were $\rho = 0.998$ ($I_{\min}$ vs. I_{mmi}), $\rho = 0.989$ ($I_{\min}$ vs. BROJA), and $\rho = 0.986$ (I_{mmi} vs. BROJA). All three correlations had $p < 10^{-10}$. The four pairs with the highest total synergy under $I_{\min}$—Pain–Physical Activity, Alcohol Use–Grip Strength, Health vs. Peers–Physical Activity, and Health vs. Peers–Functional Limitations—all remained within the top five under I_{mmi} and BROJA. BROJA synergy estimates were systematically lower in magnitude (consistent with its tighter redundancy definition) but preserved the same relative ordering. These results confirm that the qualitative conclusions and pair rankings reported in Chapter 5 are robust to the choice of redundancy definition.

D.4. Prediction surfaces and interpretation

D.4.1. Comparing synergy with additive and multiplicative baselines

Additive and multiplicative models encode built-in constraints: additive models force the joint effect to be a sum of main effects; multiplicative or independent-action models become additive on a log scale and likewise contain no irreducible joint term. When the only informative component is jointly available from two features, both families show a systematic lack-of-fit even if the main effects are strong. Table D.2 summarises these constraints and what to expect in the residuals when there is synergy.

We constructed residual maps to compare triplets with and without synergy. For each triplet $(X, Z \to Y)$, in which X and Z denote the two sources and Y the target, we form a grid from the discrete levels of X and Z to get every possible combination of variables

in their joint state space. Let i index levels of X and j index levels of Z. Define $\mu_{ij} = \mathbb{E}[Y \mid X = i, Z = j]$ as the observed mean outcome for the combination $X = i$ and $Z = j$. Let $\bar{Y}_{i\cdot}$, $\bar{Y}_{\cdot j}$, and $\bar{Y}_{\cdot\cdot}$ be the count-weighted row, column, and grand means. The *additive* expectation and residual are

$$\widehat{\mu}_{ij}^{\text{add}} = \bar{Y}_{i\cdot} + \bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot}, \qquad R_{ij}^{\text{add}} = \mu_{ij} - \widehat{\mu}_{ij}^{\text{add}}.$$

Here $R_{ij}^{\text{add}} > 0$ indicates outcomes worse than the additive expectation for that $X = i$, $Z = j$ combination; $R_{ij}^{\text{add}} < 0$ indicates outcomes better than the additive expectation.
To assess a *multiplicative* (independent-action) baseline, we operate on the log scale, adding a small constant ε to all cell means before taking logarithms to avoid zeros:

$$\widehat{\log \mu}_{ij}^{\text{mult}} = \log \bar{Y}_{i\cdot} + \log \bar{Y}_{\cdot j} - \log \bar{Y}_{\cdot\cdot}, \qquad L_{ij}^{\text{mult}} = \log \mu_{ij} - \widehat{\log \mu}_{ij}^{\text{mult}},$$

so that $\exp\left(L_{ij}^{\text{mult}}\right)$ is the observed/expected ratio under multiplicativity (values > 1 denote worse-than-multiplicative outcomes).
For a set of representative triplets, we display heatmaps of the observed means μ_{ij}, the additive expectations $\widehat{\mu}_{ij}^{\text{add}}$, and the additive residuals R_{ij}^{add} to show deviations from model expectations. Multiplicative log-residual maps L_{ij}^{mult} are shown in Figures D.1 and D.3 and are interpreted analogously.

Table D.2: Synergy and built-in constraints in two common model families.

Model family	**Built-in mathematical constraint**	**Consequence when synergy is present**
Additive (e.g., linear regression, main-effects GLM)	Effect = $\beta_0 + \beta_1 X_1 + \beta_2 X_2$	Interaction term is forced to 0; the model either pushes synergy into the error term or distorts the β-coefficients.
Multiplicative / independent-action (e.g., log-transformed models)	Effect = $\alpha X_1 X_2 \longrightarrow$ additive on the log scale	Still no irreducible joint term; systematic lack-of-fit appears in residual plots or curvature.

D.4.2. Synergistic example (Sleep × WHR → Function): multiplicative view

Figure D.1 shows the Sleep × WHR → Function triplet under a multiplicative (independent-action) baseline on the log scale. The log-residual map $L_{ij}^{\text{mult}} = \log \mu_{ij} - \widehat{\log \mu}_{ij}^{\text{mult}}$ highlights a local pocket of *positive* residuals at poor sleep and moderate WHR; $\exp(L_{ij}^{\text{mult}}) > 1$ indicates that observed functional limitation exceeds multiplicative expectations. This pattern mirrors the additive residual pocket in the main figure and is

consistent with strictly positive PID synergy for Sleep–WHR with respect to Function. In short, both additive and multiplicative baselines leave structured errors in the region where PID assigns positive synergy.

D.4.3. Non-synergistic example (blood pressure (MAP) × Hearing → Function): additive and multiplicative views

Figures D.2 and D.3 display, respectively, the additive residuals R_{ij}^{add} and the multiplicative log-residuals L_{ij}^{mult} for the blood pressure–Hearing pair. Both maps are near-symmetric around zero with low amplitude and no coherent spatial structure. PID synergy for this triplet is negligible, in agreement with the flat residual surfaces. Here, main-effects models adequately capture the association with Function; adding an interaction would be unlikely to improve fit meaningfully.

D.4.4. Practical reading of residual maps

The sign conventions are those fixed when the residuals were defined. A positive additive residual, $R_{ij}^{\text{add}} > 0$, means a joint effect larger than the sum of the parts, and a negative residual the reverse. For multiplicative residuals, values $\exp(L_{ij}^{\text{mult}}) > 1$ (or $L_{ij}^{\text{mult}} > 0$) indicate observed/expected ratios above one under independence, whereas ratios below one indicate the opposite. Residual maps should be interpreted jointly with PID results: structured, localised departures in either residual surface typically coincide with positive synergy, while uniformly pale maps coincide with near-zero synergy. Colour scales are centred at zero in every map shown here. Cell means are weighted by counts to avoid artefacts from sparse bins.

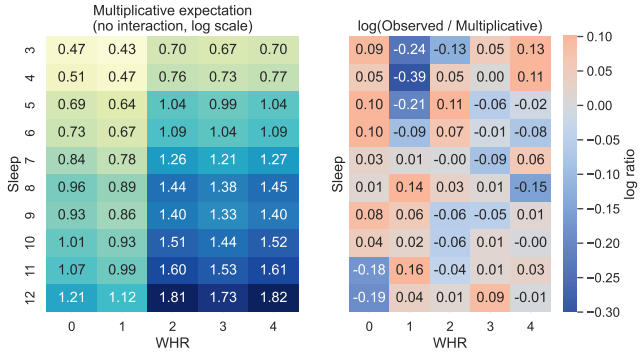


Figure D.1: Sleep × WHR → Function under multiplicative independence. Left: multiplicative expectation on the log scale. Right: log-residual L_{ij}^{mult}; values above zero indicate observed/expected > 1.

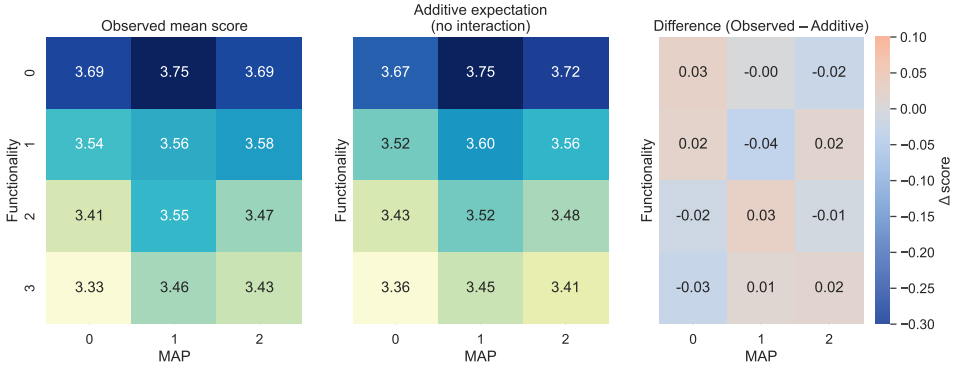


Figure D.2: Blood pressure (MAP) × Hearing → Function (non-synergistic). Left: observed means. Middle: additive expectation. Right: additive residuals showing minimal structure.

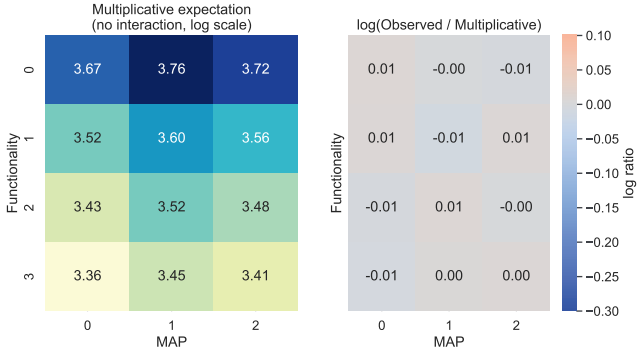


Figure D.3: Blood pressure (MAP) × Hearing → Function (non-synergistic), multiplicative view. Residuals on the log scale are near zero, consistent with negligible PID synergy.

D.5. Merging of variable states

Plug-in estimation of PID atoms is unreliable when a triplet's three-way contingency table contains near-empty cells. The clinically motivated discretisation was therefore followed by a deterministic merging step, which combined adjacent ordinal categories until no observed three-way joint state had fewer than five observations. Only neighbouring categories were merged, so each variable retains its clinical ordering but loses resolution. Thirteen variables required at least one such merge. Table D.3 records the resulting mapping, which was derived once on the full sample using the rule described in Chapter 5.

D.6. Rule-based joint-state patterns

To examine how specific constellations behaved with respect to individual outcomes, we evaluated the joint states of each constellation across all target variables not included among its members. All variables were discrete, and analyses were performed on the full population. For each constellation–target pair, we tabulated all observed joint states and computed the empirical mean of the target in each state. An ordinal regression model with the target as the response and the constellation variables as main-effect predictors served as an additive benchmark; expected values under this model were obtained by probability-weighted averaging across outcome categories.

Deviations were defined as the difference between the observed and additive-predicted means for each joint state. States with fewer than five observations were flagged as low-confidence. Rule-like configurations were those with absolute deviations in the upper 5% for that constellation–target combination. To assess whether these configurations contributed information beyond the main effects, we augmented the ordinal model with binary indicators for the flagged states and compared fit to the baseline model using the Akaike information criterion (AIC). Those configurations were themselves selected as the largest deviations from that baseline model, fitted to the same data. The resulting AIC differences are therefore descriptive rather than inferential, and are biased in favour of the augmented model.

Across constellations, the magnitude of the rule-like deviations varied. Their number was fixed by the definition. Rule-like states were the upper 5% of observed joint states, rounded up, so their number follows from the size of the state space rather than from the data. The Alcohol Use–grip strength pair had 12 joint states and one rule-like state per target. Deviations from additivity reached only 0.250, yet AIC improved for 14 of its 15 targets, by up to 57.4 units (Table D.4). The four-variable constellation of General Health, Pain, grip strength, and Cognition had 120 joint states and six rule-like states per target. Its deviations reached 1.110, and AIC fell for all 14 of its targets, by 2.6 to 30.4 units (Table D.5). The constellation of Health-vs-Peers, Physical Activity, and Sleep fell in between, with four rule-like states per target, deviations up to 0.536, and lower AIC for every target (Table D.6). The Sleep–Physical Activity pair had two rule-like states per target and deviated least, by at most 0.261. It improved fit for 13 of its 15 targets, most

Table D.3: Post-hoc merges applied after initial clinical discretisation to ensure a minimum observed three-way joint cell count of five. Only state changes are shown (states not listed map to themselves).

Variable	Post-hoc state mapping (old → new)
MMSE_Spell	0→2; 1→2
Blood pressure (MAP)	2→1
Anxiety	2→1; 3→1
Alcohol_Use	3→2; 4→2
Physical_Activity	1→2
Depression	0→1; 3→2
Health_vs_Peers	1→2
General_Health	4→3
Hearing	2→1
functionality_state	0→1
grip_category	4→3
Sleep	3→4
Pain	6→5

clearly for grip strength and Alcohol Use (Table D.7). Taken together, these tables indicate that synergistic constellations differ not only in the Breadth and Uniformity captured by BUST, but also in how far their joint states depart from an additive model.

Table D.4: Summary of joint-state deviations and AIC comparison for the Alcohol Use–grip strength constellation across all target variables. For each target, the table reports the number of observed joint states (n_{states}), the number of rule-like configurations (n_{rules}), the maximum absolute deviation between observed and additive-predicted means (max|Δ|), and the direction of that deviation. AIC is reported for the baseline and the rule-augmented ordinal model.

Target	n_{states}	n_{rules}	max|Δ|	Direction	AIC (baseline)	AIC (rule)	ΔAIC
Anxiety	12	1	0.040	positive	15,915.476	15,914.230	−1.247
Depression	12	1	0.037	positive	7998.854	7967.211	-31.644
Sleep	12	1	0.250	positive	52727.930	52720.505	-7.425
Cognition	12	1	0.049	negative	9778.249	9733.655	-44.594
Pain	12	1	0.164	negative	32288.510	32266.560	-21.951
Hearing	12	1	0.059	negative	24295.210	24295.126	-0.084
General_Health	12	1	0.088	negative	22360.130	22329.158	-30.972
Health_vs_Peers	12	1	0.066	positive	27455.668	27444.609	-11.059
Physical_Activity	12	1	0.161	positive	26829.965	26781.926	-48.039
BMI	12	1	0.055	negative	29837.861	29834.426	-3.435
Memory_Problems	12	1	0.052	negative	18048.117	18043.363	-4.755
Memory_Complaints	12	1	0.052	negative	18048.117	18043.363	-4.755
WHR_Obesity	12	1	0.013	negative	10365.228	10365.768	0.540
Blood pressure	12	1	0.035	negative	13981.366	13980.399	-0.967
functionality_state	12	1	0.191	negative	23665.826	23608.380	-57.446

Table D.5: Summary of joint-state deviations and AIC comparison for the General Health–Pain–grip strength–Cognition constellation across all target variables. For each target, the table reports the number of observed joint states (n_{states}), the number of rule-like configurations (n_{rules}), the maximum absolute deviation between observed and additive-predicted means (max|Δ|), the direction of this deviation, and the AIC values for the baseline and rule-augmented ordinal models.

Target	n_{states}	n_{rules}	max|Δ|	Direction	AIC (baseline)	AIC (rule)	ΔAIC
Anxiety	120	6	0.321	negative	14759.088	14747.285	-11.802
Depression	120	6	0.347	positive	6989.281	6973.826	-15.455
Sleep	120	6	1.110	negative	51736.208	51726.656	-9.553
Hearing	120	6	0.634	negative	24170.489	24158.966	-11.522
Health_vs_Peers	120	6	0.445	positive	24426.470	24409.246	-17.224
BMI	120	6	0.223	negative	29633.362	29613.307	-20.056
Alcohol_Use	120	6	0.365	positive	28954.214	28928.156	-26.058
Memory_Problems	120	6	0.389	negative	17898.648	17878.392	-20.256
Memory_Complaints	120	6	0.389	negative	17898.648	17878.392	-20.256
grip_category	120	6	0.531	negative	36287.994	36257.636	-30.358
WHR_Obesity	120	6	0.139	negative	10402.478	10398.841	-3.637
Blood pressure	120	6	0.124	negative	14072.473	14069.857	-2.616
functionality_state	120	6	0.641	negative	20818.320	20798.862	-19.458

Table D.6: Summary of joint-state deviations and AIC comparison for the Health-vs-Peers–Physical Activity–Sleep constellation across all target variables. For each target, the table reports the number of observed joint states (n_{states}), the number of rule-like configurations (n_{rules}), the maximum absolute deviation between observed and additive-predicted means (max|Δ|), the direction of this deviation, and the AIC values for the baseline and rule-augmented ordinal models.

Target	n_{states}	n_{rules}	max|Δ|	Direction	AIC (baseline)	AIC (rule)	ΔAIC
Anxiety	63	4	0.105	negative	14621.916	14617.814	-4.101
Depression	63	4	0.133	positive	7130.283	7123.223	-7.060
Cognition	63	4	0.169	positive	9436.955	9419.692	-17.263
Pain	63	4	0.499	positive	29859.282	29838.100	-21.181
Hearing	63	4	0.526	negative	24187.219	24159.503	-27.716
General_Health	63	4	0.113	positive	21187.059	21180.274	-6.785
BMI	63	4	0.122	negative	29642.914	29630.769	-12.145
Alcohol_Use	63	4	0.213	positive	28971.610	28953.016	-18.594
Memory_Problems	63	4	0.263	negative	17911.274	17901.771	-9.503
Memory_Complaints	63	4	0.263	negative	17911.274	17901.771	-9.503
grip_category	63	4	0.536	negative	36382.812	36339.400	-43.412
WHR_Obesity	63	4	0.084	negative	10402.737	10399.146	-3.590
Blood pressure	63	4	0.049	negative	14073.040	14071.042	-1.997
functionality_state	63	4	0.265	negative	20817.028	20808.873	-8.155

Table D.7: Summary of joint-state deviations and AIC comparison for the Sleep–Physical Activity constellation across all target variables. For each target, the table reports the number of observed joint states (n_{states}), the number of rule-like configurations (n_{rules}), the maximum absolute deviation between observed and additive-predicted means (max|Δ|), the direction of this maximum deviation, and the AIC values for the baseline and rule-augmented ordinal models.

Target	n_{states}	n_{rules}	max|Δ|	Direction	AIC (baseline)	AIC (rule)	ΔAIC
Anxiety	21	2	0.054	positive	14788.512	14785.504	-3.008
Depression	21	2	0.049	negative	7273.374	7271.163	-2.211
Cognition	21	2	0.046	positive	9443.915	9440.324	-3.591
Pain	21	2	0.198	positive	30714.896	30712.994	-1.903
Hearing	21	2	0.167	negative	24189.537	24185.939	-3.598
General_Health	21	2	0.069	positive	21199.229	21197.192	-2.037
Health_vs_Peers	21	2	0.064	positive	27182.606	27183.270	0.664
BMI	21	2	0.069	negative	29646.440	29643.477	-2.963
Alcohol_Use	21	2	0.115	positive	28998.317	28971.297	-27.020
Memory_Problems	21	2	0.108	negative	17921.059	17905.550	-15.509
Memory_Complaints	21	2	0.108	negative	17921.059	17905.550	-15.509
grip_category	21	2	0.261	negative	36105.356	36076.138	-29.217
WHR_Obesity	21	2	0.024	negative	10404.314	10405.087	0.773
Blood pressure	21	2	0.036	negative	14074.111	14067.868	-6.243
functionality_state	21	2	0.089	negative	20841.871	20840.387	-1.483

D.7. Simulation validation of the estimation pipeline

The three simulation experiments summarised in Methods Section 5.5.8 were designed to establish that the QE bias-corrected pipeline recovers known PID atoms and BUST scores under conditions matched to the LASA data; their full results are reported below.

Cardinality sweep. Figure D.4 shows synergy RMSE as a function of variable cardinality (K = 2–5) for each of the four target functions (modular copy, XOR, minimum, and redundancy-dominated) under $I_{\min}$. RMSE increases with K for all target functions except the redundancy-dominated function, whose true synergy is zero regardless of cardinality and is recovered with negligible error. Fig D.5 shows the interaction between cardinality and sample size: at $K \leq 4$, synergy RMSE remains below 0.02 bits across all tested sample sizes (N = 500–2,636); at $K = 5$ with $N = 500$, RMSE rises to approximately 0.04 bits for the minimum function, the most challenging distribution tested.

LASA-matched cardinality profiles. Fig D.6 reports RMSE for all four PID atoms across four LASA-matched cardinality profiles ($3 \times 4 \to 5$, $2 \times 2 \to 7$, $3 \times 3 \to 3$, $5 \times 7 \to 5$) under three information regimes (function-of-one, mixed, and synergy-rich). All atoms are recovered with RMSE < 0.02 bits at the largest tested sample size for most profile–regime combinations. The highest-cardinality profile ($5 \times 7 \to 5$) exhibits moderately elevated RMSE for the synergy atom under the synergy-rich regime (≈ 0.08 bits), consistent with the cardinality sweep results.

BUST score recovery. Fig D.7 shows estimated vs. true BUST scores (Breadth, Synergy strength, Total information) for three synthetic source pairs with distinct breadth profiles. The Uniformity dimension was not simulated: it is defined by the consistency of synergy across population strata rather than within a single joint distribution, so it has no counterpart in these single-distribution ground truths, and the recovery results reported here therefore bear on B, S and T only. Synergy strength and Total information fall close to the identity line across all pairs and sample sizes. Breadth exhibits a systematic downward bias for the broad-synergy pair (true $B \approx 0.96$, estimated $B \approx 0.78$), consistent with estimation noise deflating normalised entropy; the narrow-synergy pair ($B \approx 0$) is recovered accurately. The relative ranking of pairs by all three BUST components is preserved across all conditions.

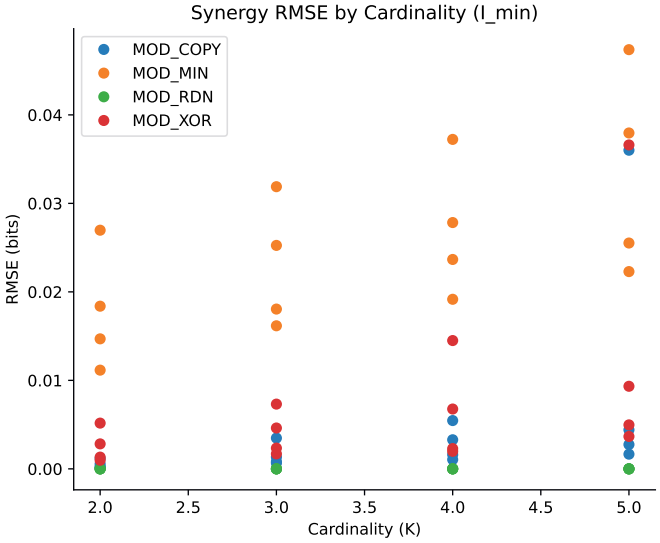


Figure D.4: Synergy RMSE as a function of variable cardinality (K) for four deterministic target functions under $I_{\min}$, pooled across sample sizes and source-coupling conditions. Each point represents one condition (sample size × coupling); vertical spread shows variation across these conditions.

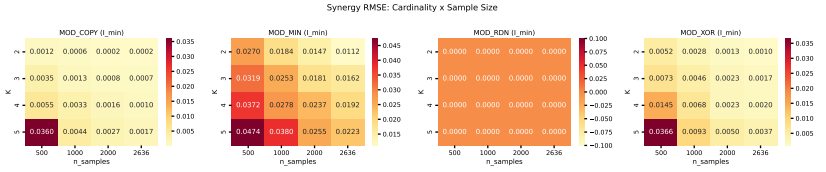


Figure D.5: Joint effect of variable cardinality and sample size on synergy RMSE ($I_{\min}$, independent sources). Each cell shows the RMSE (bits) for one distribution–sample-size combination.

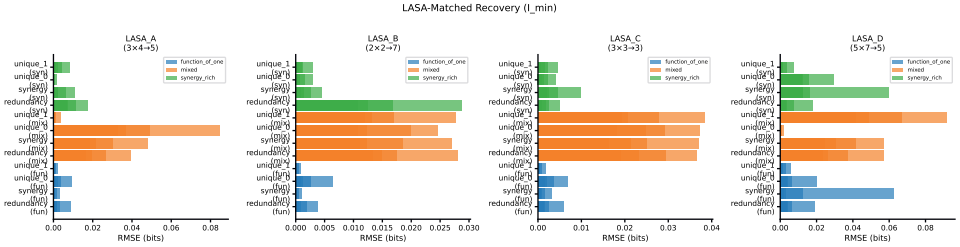


Figure D.6: PID atom recovery under four LASA-matched cardinality profiles and three information regimes ($I_{\min}$, $N = 2{,}636$). Horizontal bars show RMSE (bits) for each atom–regime combination. Lower-cardinality profiles (LASA_A–LASA_D denote the four profiles in the order given in the text; LASA_B: $2 \times 2 \to 7$; LASA_C: $3 \times 3 \to 3$) achieve the lowest RMSE; the highest-cardinality profile (LASA_D: $5 \times 7 \to 5$) shows moderately elevated RMSE for synergy under the synergy-rich regime.

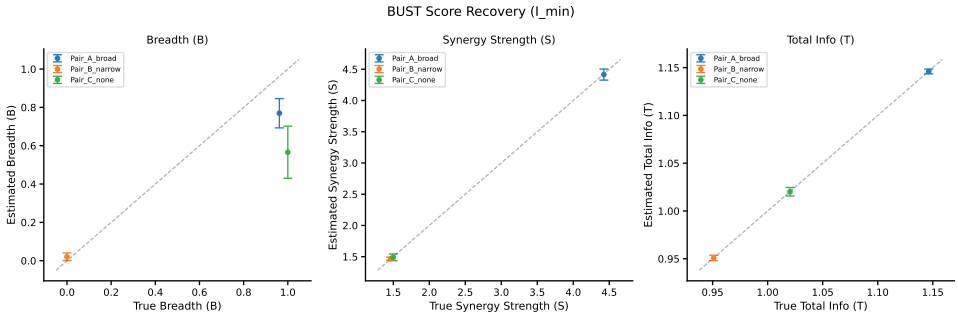


Figure D.7: BUST score recovery: estimated vs. true values for Breadth (B), Synergy strength (S), and Total information (T) across three synthetic source pairs with known breadth profiles (broad, narrow, none). Points show means across replicates; error bars show ± 1 SD. The dashed line is the identity. Synergy strength and Total information are recovered accurately; Breadth shows a conservative downward bias for the broad-synergy pair.

D.8. Sensitivity to discretisation choices

To assess the robustness of BUST scores to discretisation, we re-ran the full PID and BUST pipeline under four alternative binning strategies, applied to all continuous and ordinal variables with more than two levels, and compared them with the clinical scheme used in the primary analysis (five schemes in total):

1. **Clinical** (primary analysis): cut-points based on clinical guidelines and LASA documentation, with post-hoc sparse-state merging (see Section 5.5.5).
2. **Equal-frequency tertiles**: each variable split into three bins of approximately equal size.
3. **Equal-frequency quartiles**: each variable split into four equal-frequency bins.
4. **Equal-width tertiles**: each variable split into three bins of equal range.
5. **Coarsened binary**: each variable dichotomised at its median.

Binary variables (e.g., memory complaints, anxiety, WHR obesity) retained their original coding under all schemes. For each scheme, PID atoms were estimated with QE bias correction and target-shuffle baseline subtraction, and BUST composite scores were computed exactly as in the primary analysis.

Rank correlations. Table D.8 reports pairwise Spearman rank correlations of composite BUST scores across schemes. Among the four non-binary schemes, correlations ranged from 0.83 to 0.96, indicating that the relative ordering of source pairs by BUST score is largely preserved regardless of whether clinical, equal-frequency, or equal-width bins are used. The equal-frequency tertile and quartile schemes are nearly interchangeable (ρ = 0.96). The coarsened binary scheme is poorly correlated with all others (ρ = 0.06–0.18; all $p > 0.1$), reflecting the severe information loss incurred when multi-level variables are collapsed to two categories.

Table D.8: Pairwise Spearman rank correlations of composite BUST scores across five discretisation schemes. Bold values indicate $p < 0.001$.

	Clinical	Freq-3	Freq-4	Width-3	Binary
Clinical	1.00	**0.84**	**0.83**	**0.85**	0.18
EF-3		1.00	**0.96**	**0.90**	0.06
EF-4			1.00	**0.86**	0.14
EW-3				1.00	0.10
Binary					1.00

Top-pair overlap. The Jaccard index for the top-10 ranked pairs between equal-frequency tertiles and quartiles was 0.82, indicating that the highest-ranked BUST edges are robust to granularity within equal-frequency schemes. Between clinical and equal-frequency schemes, top-10 Jaccard indices ranged from 0.18 to 0.25—three or four shared pairs out of ten. Top-10 membership is therefore sensitive to where the cut-points fall. Clinical cut-points and data-driven quantiles partition variable space differently. What the schemes preserve is the ordering of the full set of pairs rather than the identity of the leading ones ($\rho = 0.83$–0.85, Table D.8). The source pair Physical Activity–Memory Complaints appeared in the top-10 under all four non-binary schemes. The binary scheme shared no top-10 pairs with the clinical scheme (Jaccard = 0.00).

BUST quadrant stability. Quadrant assignment was only moderately stable. Pairwise BUST quadrant agreement (broad vs. narrow classification) among the four non-binary schemes ranged from 59% to 84% (mean 72%), so at the lower end it was little better than the 50% expected by chance for a two-class assignment. Including the binary scheme reduced overall agreement to 27%, driven by the binary scheme's tendency to assign high breadth scores to all pairs owing to the collapse of multi-level variables. The poor agreement of the coarsened binary scheme is expected. By the data processing inequality [69], collapsing multi-level variables to two categories can only reduce mutual information. A uniform reduction would leave the ranking intact, so the reordering seen here indicates that pairs lose unequal amounts of information under binary coding. The resulting $2 \times 2 \times 2$ joint triplet states provide far less resolution for non-additive patterns to manifest than the richer joint state spaces available under multi-level discretisations.

Visual comparison of BUST dimensions. The four non-binary schemes yield visually similar Breadth–Synergy landscapes, whereas the coarsened binary scheme does not. Because each binning scheme was applied to the full population only (without stratification), the Uniformity dimension cannot be estimated and is omitted from the visual comparison. Fig D.8 displays the three remaining BUST dimensions—Breadth (B), Synergy strength (S), and Total information (T) - for each scheme, using an encoding that mirrors the main BUST map (Fig 5.2): the x-axis shows median-centred Breadth, the y-axis shows Synergy strength in bits, point size encodes Total information, and point colour encodes the Synergy fraction (S/T). The four non-binary schemes produce visually similar point clouds, with the same pairs occupying high-synergy, moderate-breadth positions across panels. The coarsened binary scheme shows a markedly different structure: most pairs are pushed to high breadth (because binary variables yield near-uniform synergy shares across targets) and low synergy strength, consistent with the information loss from collapsing multi-level variables.
Fig D.9 presents the pairwise Spearman rank correlations of composite BUST scores as a heatmap, and Fig D.10 shows the Jaccard overlap of the top-10 ranked pairs for each scheme comparison. Robustness depends on how finely the ranking is read. The ordering of source pairs is preserved across the four non-binary schemes ($\rho = 0.83$–0.96). The

identity of the pairs at the head of that ordering is more sensitive: the top-10 Jaccard index between clinical and equal-frequency schemes is 0.18–0.25, and quadrant agreement among the same four schemes ranges from 59% to 84%.

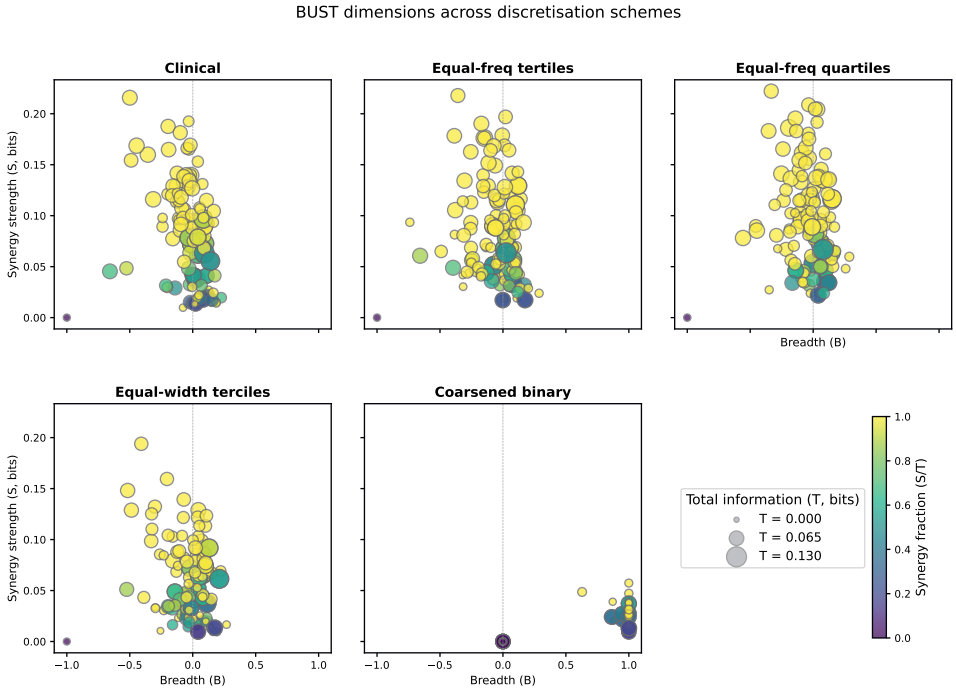


Figure D.8: Breadth–Synergy scatter for all source pairs under five discretisation schemes. The x-axis shows median-centred Breadth (B); the y-axis shows Synergy strength (S, bits); point size encodes Total information (T, bits); point colour encodes Synergy fraction (S/T). The Uniformity dimension is omitted because each scheme was evaluated on the full population only. The four non-binary schemes produce similar point clouds, whereas the coarsened binary scheme shows a distinctly different pattern with compressed synergy magnitudes and inflated breadth.

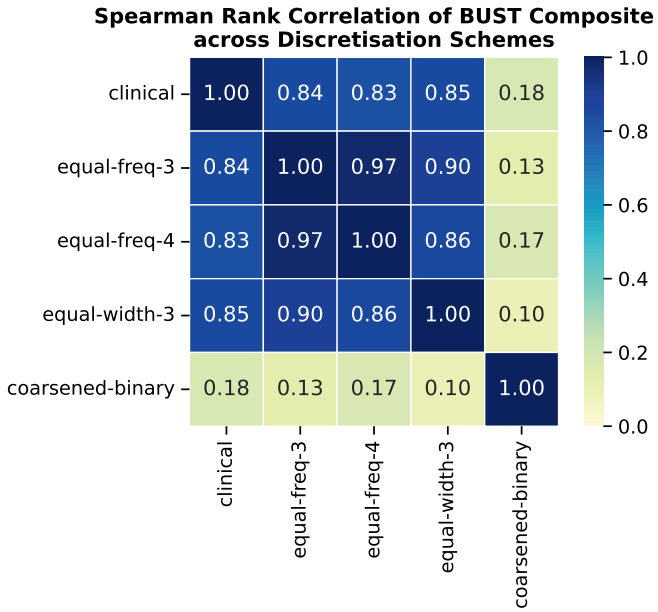


Figure D.9: Spearman rank correlation matrix of composite BUST scores across five discretisation schemes. Non-binary schemes are highly correlated ($\rho = 0.83$–0.96), while the coarsened binary scheme is poorly correlated with all others ($\rho = 0.06$–0.18).

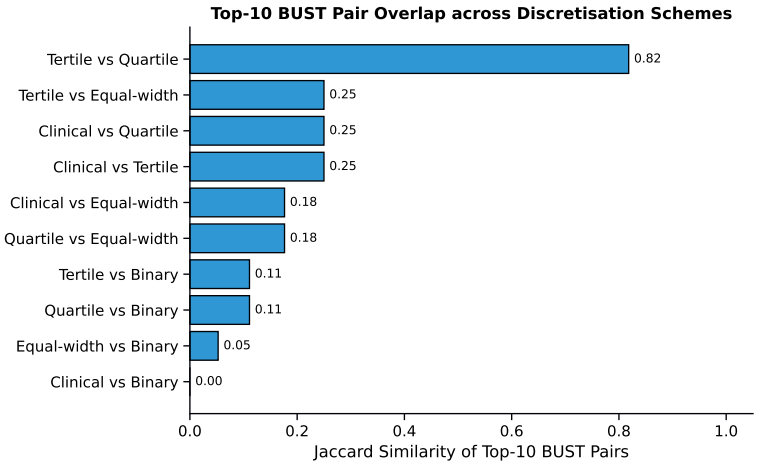


Figure D.10: Jaccard similarity of the top-10 ranked BUST pairs across all pairwise scheme comparisons. The equal-frequency tertile and quartile schemes share nine of their ten top-ranked pairs (Jaccard = 0.82). Comparisons involving the coarsened binary scheme show near-zero overlap.

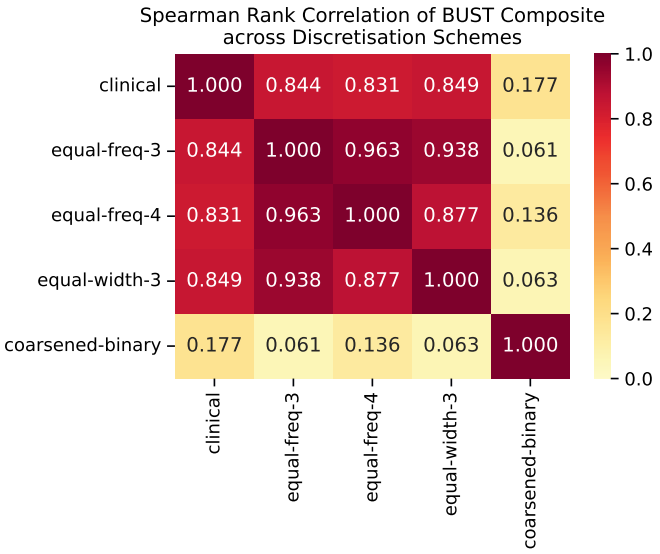


Figure D.11: Spearman rank correlations of composite BUST scores between each pair of discretisation schemes, displayed as an annotated heatmap.

D.9. Bootstrap confidence intervals for synergy estimates

To assess the sampling variability of synergy estimates for the highest-ranked BUST pairs, we performed a non-parametric bootstrap with 1,000 replicates. In each replicate, the full-population discretised dataset (N = 14,716 observations) was resampled with replacement, and QE-corrected PID atoms were re-estimated for the 148 (pair, target) combinations associated with the top-10 composite BUST pairs. Percentile-based 95% confidence intervals were computed for each combination and for the per-pair mean synergy across targets.

All ten pairs have CIs for mean synergy that exclude zero (Fig D.12, ordered by point estimate). The fraction of individual targets whose synergy CI excludes zero ranges from 73% (Memory Complaints–Physical Activity, Memory Problems–Physical Activity) to 93% (Physical Activity–Sleep, Pain–Grip Strength). Synergy is therefore statistically distinguishable from zero for most targets rather than for isolated cases. Detectability is not magnitude. Breadth summarises the distribution of synergy magnitude across targets, not the number of targets with non-zero synergy.

Fig D.13 presents target-specific synergy CIs for the top-5 pairs, showing which targets contribute most to each pair's aggregate synergy. For example, the pain–physical activity pair derives its strongest synergy from functional limitations (≈ 0.12 bits), whereas its synergy with respect to most other targets is an order of magnitude smaller.

Fig D.14 characterises the precision of individual estimates. The median CI width across all 148 (pair, target) combinations was 0.005 bits, and CI width scaled approximately linearly with effect size, indicating approximately constant relative precision rather than a loss of precision at larger effect sizes; the median width is small relative to the between-pair differences in synergy on which the BUST ranking depends, although no formal precision criterion was set in advance.

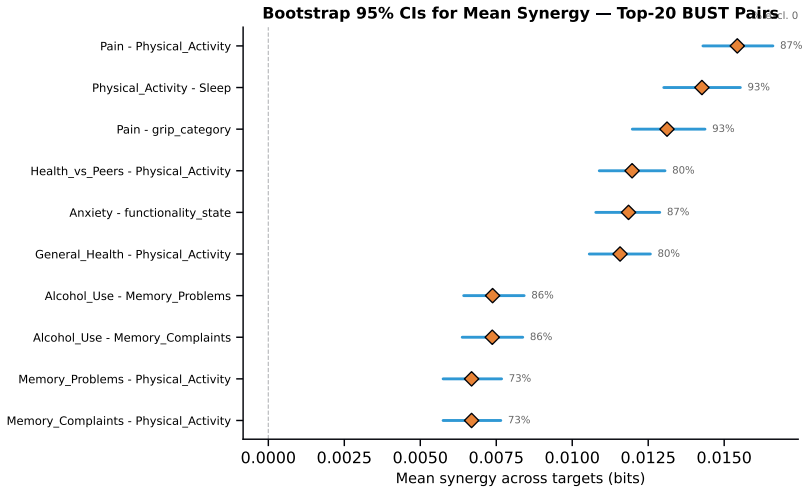


Figure D.12: **Bootstrap 95% confidence intervals for the mean synergy of the top-10 BUST pairs across targets.** Diamond markers show the point estimate (mean synergy across all targets); horizontal bars show percentile-based 95% CIs from 1,000 bootstrap replicates. The percentage to the right of each bar indicates the fraction of individual targets whose target-specific synergy CI excludes zero. All ten pairs have aggregate CIs that exclude zero.

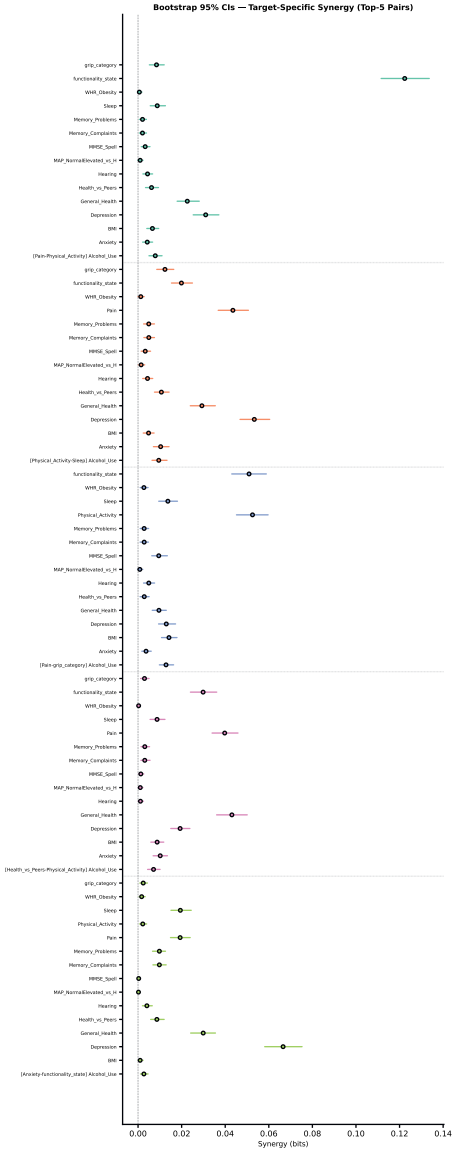


Figure D.13: **Target-specific bootstrap 95% CIs for synergy in the top-5 BUST pairs.** Each panel corresponds to one source pair; rows within panels show individual targets. Horizontal bars are percentile-based 95% CIs (1,000 replicates). The variation in synergy magnitude across targets illustrates the breadth dimension of the BUST representation.

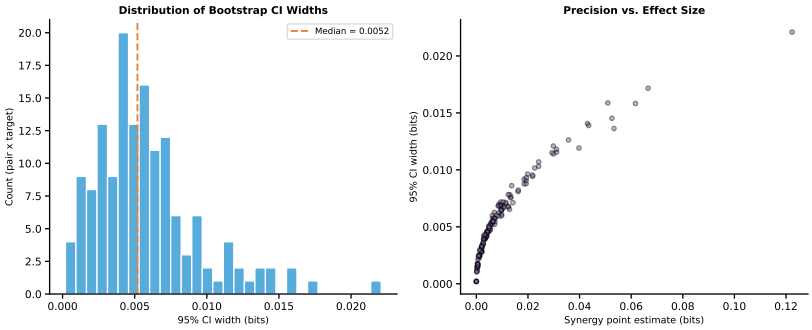


Figure D.14: **Distribution of bootstrap CI widths.** Left: histogram of 95% CI widths across all 148 (pair, target) combinations (median = 0.005 bits). Right: CI width plotted against the synergy point estimate, showing an approximately linear relationship between precision and effect size.



SUPPLEMENTARY MATERIAL FOR CHAPTER 6

This appendix comprises the following sections: Appendix E.1 describes the multi-dataset benchmark design and participating datasets. Appendix E.2 provides detailed descriptions and hyperparameter settings for all 13 embedding methods. Appendix E.3 discusses the information-bottleneck interpretation of the ACIB embedding and computational scaling properties. Appendix E.4 presents calibration statistics for the synthetic Bayesian Network ensemble. Appendix E.5 reports omnibus and post-hoc statistical tests for the embedding benchmark. Appendix E.6 reports additional embedding-fidelity diagnostics and symptom-level PID examples. Appendix E.7 examines the prevalence of non-linear dependencies in clinical datasets. Appendix E.8 provides cross-cohort comparisons of information-theoretic PHQ-9 networks. Appendix E.10 presents an illustrative three-variable example contrasting how partial correlation, conditional mutual information, and PID reveal progressively richer dependence structure. Appendix E.11 demonstrates that the reported decompositions are robust to the ACIB embedding hyperparameters and justifies embedding only a bounded remainder context.

E.1. Multi-Dataset Benchmark

This section describes the assembly and preprocessing of the 83 real-world datasets used to evaluate embedding fidelity across diverse data structures and PID measures (Section 6.2.6).

E.1.1. Benchmark dataset assembly and preprocessing

We assembled 83 datasets from publicly available repositories spanning clinical, epidemiological, and psychometric domains, including NHANES [248], SHARE [249], ELSA [250], HRS [251], PROMIS [252], OpenPsychometrics, the UCI Machine Learning Repository, OSF, BRFSS [253], MEPS [254], the Household Pulse Survey [255], ICPSR, Zenodo, Figshare, and Mendeley Data. Datasets span depression, anxiety, PTSD, stress, personality, eating disorders, physical functioning, chronic conditions, sleep, and other clinical domains, with 5–56 variables per dataset and sample sizes ranging from 32 to 445,000 observations. For each dataset, we selected only substantive item-level or symptom-level variables, excluding identifiers, demographic covariates, composite scores, and diagnostic labels. Source-specific missing-value codes (e.g., 7/77/777 and 9/99/999 in NHANES; negative codes in SHARE) were recoded to missing. For NHANES SAS transport files, the XPT format encodes zero as a tiny floating-point number ($\approx 5.4 \times 10^{-79}$); these values were rounded to zero before further processing.

Columns with more than 30% missing values were dropped, followed by rows with more than 50% missing values across the remaining columns. Remaining missing values were imputed using k-nearest neighbours ($k = 5$) via scikit-learn's `KNNImputer`. For four datasets where this filtering left fewer than three columns or thirty rows (`nhanes_phys_func`, `covidistress_28`, `ncsr_depression_18`, `nhanes_alcohol`), listwise deletion (complete-case analysis) was used instead, retaining all variables and only rows with no missing values. After imputation, any genuinely constant columns

(zero variance) were removed.

For each dataset, two association matrices were estimated. Partial correlations were computed from an ℓ_1-regularised precision matrix obtained via `GraphicalLassoCV` with 5-fold cross-validation, which estimates partial correlations as scaled off-diagonal entries of the regularised precision matrix Σ^{-1}, yielding a sparse estimate of conditional dependencies. For mutual information estimation, all imputed values were first rounded to the nearest integer to recover ordinal response levels. Variables with five or fewer unique integer values were retained at their original levels; variables with more than five unique values were discretised into eight equal-frequency bins. For partial information decomposition, all variables were further reduced to at most k discrete states by iteratively merging adjacent categories with the smallest combined frequency, where k was the largest integer satisfying $k^{(N_{\text{sources}}+1)} \leq N/5$, ensuring adequate coverage of the joint state space. Pairwise mutual information was estimated using the discrete estimator from the `npeet` library, with diagonal entries set to the Shannon entropy of each variable.

Table E.1 lists all 83 datasets, their domain, number of selected variables, sample size, source repository, and citation.

Table E.1: **Benchmark datasets.** The 83 datasets used in the multi-dataset embedding benchmark. *Items* indicates the number of substantive variables selected before preprocessing; N is the raw sample size before row-level missingness filtering. easySHARE waves (6a–6g) are drawn from the same longitudinal study but are treated as separate networks because each wave has an independent cross-section of respondents.

Dataset	Domain	Items	N	Source	Citation
CSWS Items	Self-worth	35	680	OSF	[256]
CSWS Subscales	Self-worth	7	680	OSF	[256]
SHARE W7	Depression	17	77,000	SHARE	[249]
ELSA W6	Depression	17	10,000	UK Data	[250]
HRS W13	Sleep/pain/health	18	21,000	HRS	[251]
easySHARE 2004 (W1)	Summary health	8	30,000	SHARE	[257, 258]
easySHARE 2007 (W2)	Summary health	9	37,000	SHARE	[257, 258]
easySHARE 2011 (W4)	Summary health	8	58,000	SHARE	[257, 258]
easySHARE 2013 (W5)	Summary health	8	66,000	SHARE	[257, 258]
easySHARE 2015 (W6)	Summary health	9	68,000	SHARE	[257, 258]
easySHARE 2017 (W7)	Summary health	6	77,000	SHARE	[257, 258]
easySHARE 2020 (W8)	Summary health	9	47,000	SHARE	[257, 258]
NHANES PHQ-9	Depression	9	5,500	CDC	[184, 248]
NHANES Sleep	Sleep	10	6,200	CDC	[248]
NHANES Phys. Func.	Disability	24	8,400	CDC	[248]
NHANES Health Status	General health	8	8,400	CDC	[248]
NHANES Alcohol	Alcohol	9	5,500	CDC	[248]
NHANES Blood Pressure	Cardiovascular	10	6,200	CDC	[248]
NHANES Med. Cond.	Chronic conditions	17	8,900	CDC	[248]
NHANES Smoking	Smoking	10	6,700	CDC	[248]
NHANES Phys. Activity	Activity	16	5,900	CDC	[248]
NHANES Hospital Util.	Healthcare	9	9,300	CDC	[248]
DASS-42	Dep/Anx/Stress	42	39,800	OpenPsych	[259]
TMAS	Trait anxiety	50	5,400	OpenPsych	[260]

Continued on next page

Dataset	Domain	Items	*N*	Source	Citation
Big Five (IPIP)	Personality	50	19,700	OpenPsych	[261]
RSE	Self-esteem	10	48,000	OpenPsych	[262]
ECR	Attachment	36	51,500	OpenPsych	[263]
HSNS+DD	Narcissism/dep	10	54,000	OpenPsych	[264, 265]
KIMS	Mindfulness	39	601	OpenPsych	[266]
SCS	Sexual compulsiv.	12	3,400	OpenPsych	[267]
Dermatology	Skin disease	34	366	UCI	[268, 269]
Primary Tumor	Cancer features	17	339	UCI	[270]
Early-Stage Diabetes	Diabetes symptoms	16	520	UCI	[271, 272]
Thyroid Cancer Recur.	Cancer clinical	16	383	UCI	[273]
Autism Screen. Adult	ASD screening	21	704	UCI	[274, 275]
Thoracic Surgery	Surgical symptoms	16	470	UCI	[276, 277]
Mammographic Mass	Breast imaging	5	961	UCI	[278, 279]
Chronic Kidney Disease	Kidney signs/labs	24	400	UCI	[280]
MI Complications	Cardiac symptoms	111	1,700	UCI	[281, 282]
Glioma Grading	Cancer mutations	25	839	UCI	[283]
Cervical Cancer Risk	STD/risk factors	36	858	UCI	[284]
CDC Diabetes Indic.	Health indicators	21	253,700	UCI/CDC	[285]
AIDS Clinical Trials	HIV clinical	25	2,139	UCI	[286, 287]
Mesothelioma	Cancer symptoms	34	324	UCI	[288]
NPHA Healthy Aging	Health self-report	15	714	UCI	[289]
HCV Egyptian Patients	Liver symptoms	28	1,385	UCI	[290]
Xinxiang PHQ-9	Depression	9	24,292	Zenodo	[181, 182]
Xinxiang GAD-7	Anxiety	7	24,292	Zenodo	[181, 182]
Xinxiang ISI	Insomnia	7	24,292	Zenodo	[181, 182]
Xinxiang PSS	Stress	10	24,292	Zenodo	[181, 182]
PHQ+GAD+ESS Mexico	Dep/Anx/Sleep	33	783	Figshare	[291]
MHP Students	Anx/Stress/Dep	26	2,000	Figshare	[292]
Bangladesh MH	Multi-scale dep	45	502	Mendeley	[293]
PHQ-9 Students	Depression	9	682	Mendeley	[294]
Colombia Well-Being	Stress/Anx/Dep	26	3,000	Mendeley	[295, 296]
McNally 2014	PTSD	17	344	OSF	[297]
ED Network	Eating disorders	22	245	OSF	[298]
IRI Empathy	Empathy	28	1,973	OSF	[183]
CPS Chinese	Purpose in life	12	598	OSF	[299, 300]
COVIDiSTRESS	Stress/Loneliness	28	173,000	OSF	[301]
PROMIS Anxiety	Anxiety	56	817	Harvard DV	[252]
PROMIS Depression	Depression	56	811	Harvard DV	[252]
PROMIS Anger	Anger	56	918	Harvard DV	[252]
PROMIS Fatigue (Exp)	Fatigue	56	821	Harvard DV	[252]
PROMIS Fatigue (Imp)	Fatigue	56	820	Harvard DV	[252]
PROMIS Pain Interf.	Pain	56	866	Harvard DV	[252]
PROMIS Pain Quality	Pain	56	862	Harvard DV	[252]
PROMIS Pain Behavior	Pain	56	859	Harvard DV	[252]
PROMIS PhysFun A	Physical function	56	812	Harvard DV	[252]
PROMIS PhysFun B	Physical function	56	814	Harvard DV	[252]
PROMIS Social Perf	Social function	56	864	Harvard DV	[252]
PROMIS Social Sat	Social function	56	851	Harvard DV	[252]
PROMIS Alcohol	Alcohol	56	903	Harvard DV	[252]
PROMIS Profile 29	Multi-domain	29	4,500	Harvard DV	[302]

Continued on next page

Dataset	Domain	Items	N	Source	Citation
BRFSS 2022	Health indicators	25	445,000	CDC	[253]
PHQ+GAD+ISI+PSS comb.	All 4 scales	33	24,292	Zenodo	[181, 182]
NCS-R Depression	MDE symptoms	18	2,300	ICPSR	[303, 304]
MIDUS 2 Phys. Symp.	Somatic symptoms	10	4,000	ICPSR	[305]
MIDUS 2 Chronic Cond.	Multimorbidity	20	4,000	ICPSR	[305]
MIDUS 2 Health Limits	Functional limits	10	4,000	ICPSR	[305]
MEPS 2022 SF-12	Health status	17	11,300	AHRQ	[254]
MEPS 2022 Conditions	Chronic conditions	17	22,300	AHRQ	[254]
HPS Dec 2024	PHQ-2+GAD-2+Disab.	14	9,400	Census	[255]

Per-dataset variable selections, exclusion criteria, and recoding rules are available in the project's data documentation repository.

E.2. Embedding Methods: Descriptions and Hyperparameters

To compress the multivariate remainder $\mathbf{Z}_{i\to k} = \mathbf{X} \setminus \{X_i, X_k\}$ into a single discrete embedding variable $E_{i\to k}$ for two-source PID computation, we benchmarked 13 embedding methods spanning five families. Table E.2 summarises the method families, software implementations, and key hyperparameters used throughout the benchmark. All methods were applied with a global random seed of 42 for reproducibility. Below, we describe each method and its hyperparameter settings.

E.2.1. Information-theoretic methods

Agglomerative Conditional Information Bottleneck (ACIB). ACIB is a two-phase, deterministic, target-directed compression algorithm that maps the joint states of the multivariate remainder $\mathbf{Z}_{i\to k}$ into a low-cardinality discrete variable $E_{i\to k}$ while preserving as much information as possible about the target X_k. In the first phase, a k-CIB (Conditional Information Bottleneck) initialisation is performed. The unique joint states of $\mathbf{Z}_{i\to k}$ are assigned to $K_{\max} = 12$ initial clusters using a farthest-first seeding procedure based on the Jensen–Shannon (JS) divergence of the conditional target distributions $P(X_k \mid \mathbf{Z}_{i\to k} = z, X_i = s)$ across source states s. Cluster assignments are then refined by iteratively reassigning each joint state to the cluster whose prototype minimises the Kullback–Leibler divergence, with conditional distributions smoothed via Laplace smoothing ($\alpha = 0.5$) to regularise sparse cells. This alternating assignment–update loop runs for up to 40 iterations or until assignments stabilise.

In the second phase, a greedy agglomerative merging procedure reduces the number of clusters below $K_{\max}$. At each step, the pair of clusters whose merger incurs the smallest JS divergence loss is identified and merged. After each merge, the relative information loss is evaluated as $|I(X_i; X_k \mid \mathbf{Z}) - I(X_i; X_k \mid E)| \,/\, I(X_i; X_k \mid \mathbf{Z})$, where $I(\cdot;\cdot \mid \cdot)$ denotes conditional mutual information estimated via plug-in discrete estimators [npeet; 306]. Merging continues as long as the relative loss remains below a tolerance of 5%, and stops at a minimum of $K_{\min} = 2$ clusters. The result is a deterministic mapping $f_{i\to k}$ from joint remainder states to a compact discrete code whose cardinality adapts to the complexity

Table E.2: Summary of the 13 embedding methods, grouped by family. All methods share a global random seed of 42. "K" denotes the number of clusters or output cardinality; "k" denotes the number of selected features; "B" denotes the number of quantile bins per dimension.

Method	Family	Package/Library	Key Hyperparameters
ACIB	Info-theoretic	Custom (npeet)	$K_{\max} = 12$, $K_{\min} = 2$, $\alpha = 0.5$, loss tol. = 5%, max_iter = 40
Greedy CMI	Info-theoretic	Custom (npeet)	max_vars = 4, loss tol. = 5%
JMI	MI-based selection	scikit-feature	$k = 3$
CMIM	MI-based selection	scikit-feature	$k = 3$
ReliefF	MI-based selection	skrebate	$k = 3$, n_neighbours = 100
k-modes	Clustering	kmodes	$K = 6$, init = Huang, n_init = 5
Spectral Hamming	Clustering	scikit-learn	$K = 6$, $\gamma = 5.0$, labels = k-means
Agglom. Hamming	Clustering	scikit-learn	$K = 6$, linkage = average
MCA + k-means	Dim. red. + clust.	prince, scikit-learn	n_comp = 2, $K = 6$, n_init = 10
SVD + k-means	Dim. red. + clust.	scikit-learn	n_comp = 2, $K = 6$, n_init = 10
NMF + k-means	Dim. red. + clust.	scikit-learn	n_comp = 2, $K = 6$, init = nndsvda, n_init = 10
SIR + binning	Supervised SDR	sliced	n_dir = 2, $B = 4$
PLS + binning	Supervised SDR	scikit-learn	n_comp = 2, $B = 4$

of each edge.

Greedy CMI subset selection. Greedy CMI is a forward feature-selection wrapper that selects a small subset of the remainder variables whose joint state preserves the conditional mutual information $I(X_i; X_k \mid \mathbf{Z})$. Starting from an empty set, it greedily adds the variable that minimises the relative information loss at each step, stopping when the loss falls below 5% or when a maximum of 4 variables have been selected. The embedding is the joint categorical code of the selected variables. This method directly targets information preservation without intermediate dimensionality reduction.

E.2.2. MI-based feature selection methods

Joint Mutual Information (JMI). JMI [202] is a filter-based feature selection criterion that scores each candidate feature by its joint mutual information with the target, accounting for redundancy among already-selected features. We use the implementation from the scikit-feature library [307] with $k = 3$ features selected. The embedding is the joint categorical code of the three selected remainder variables.

Conditional Mutual Information Maximisation (CMIM). CMIM [201] selects features by maximising the minimum conditional mutual information between each candidate feature and the target, given every previously selected feature. This criterion is more conservative than JMI, as it penalises redundancy more aggressively. We use the scikit-feature implementation with $k = 3$ selected features, and the embedding is again the joint categorical code.

ReliefF. ReliefF [204] is an instance-based feature weighting algorithm that evaluates features by their ability to distinguish between instances from different classes and instances from the same class, using nearest-neighbour distances. We use the skrebate implementation [205] with 100 neighbours and $k = 3$ top-ranked features. The embedding is the joint categorical code of the selected variables.

E.2.3. Clustering methods

k-modes. The k-modes algorithm [207] is a categorical analogue of k-means that uses the Hamming distance and a frequency-based mode update rule. We use the kmodes Python package with Huang initialisation, $n_init = 5$ random restarts, and $K = 6$ clusters. Cluster labels serve directly as the discrete embedding.

Spectral clustering on Hamming affinity. Spectral clustering [208] is applied to a Hamming-based affinity matrix. The pairwise Hamming distance matrix is converted to an affinity matrix via a Gaussian kernel $A_{ij} = \exp(-\gamma \cdot d_H(z_i, z_j))$ with bandwidth $\gamma = 5.0$. Spectral clustering with $K = 6$ clusters and k-means label assignment is then performed on the Laplacian eigenvectors using scikit-learn [119]. To manage memory, rows are

subsampled to 5,000 when the dataset is larger; remaining rows are assigned to the cluster of their nearest subsampled neighbour by Hamming distance.

Agglomerative clustering on Hamming distance. Agglomerative (hierarchical) clustering with average linkage is applied to the pairwise Hamming distance matrix using scikit-learn, with $K = 6$ clusters. As with spectral clustering, datasets exceeding 5,000 rows are subsampled and out-of-sample rows are assigned via nearest-neighbour lookup.

E.2.4. Dimensionality reduction followed by clustering

MCA + k-means. Multiple Correspondence Analysis (MCA) is a dimensionality reduction technique for categorical data that generalises PCA to indicator matrices [194]. We extract the first two MCA components using the prince Python library [195], then apply k-means clustering ($K = 6$, n_init = 10) to the two-dimensional scores. Cluster labels form the discrete embedding.

Truncated SVD + k-means. The remainder variables are one-hot encoded and projected onto two components via truncated singular value decomposition (SVD) using scikit-learn. The two-dimensional projections are then clustered with k-means ($K = 6$, n_init = 10). This approach treats categorical data as sparse binary features before linear dimensionality reduction.

NMF + k-means. Non-negative Matrix Factorisation (NMF) decomposes the one-hot encoded remainder matrix into two non-negative low-rank factors [196]. We extract two components using the `nndsvda` initialisation in scikit-learn, then apply k-means ($K = 6$, n_init = 10) to the coefficient matrix. The non-negativity constraint can produce parts-based representations that may capture interpretable response patterns.

E.2.5. Supervised sufficient dimension reduction followed by binning

Sliced Inverse Regression (SIR) + binning. SIR [197] estimates the central dimension-reduction subspace by exploiting the inverse regression of predictors on the response. Using the sliced Python package [198], we project the one-hot encoded remainder onto two SIR directions with respect to the target X_k. The two-dimensional projections are discretised via a $B \times B = 4 \times 4$ quantile grid, yielding up to 16 discrete categories. Because SIR uses the target for projection, this method produces a target-directed embedding.

Partial Least Squares (PLS) + binning. PLS regression [199] finds latent components that maximise the covariance between the (one-hot encoded) remainder and the target variable. We extract two PLS components using scikit-learn's `PLSRegression` and discretise the scores with a 4×4 quantile grid, producing up to 16 categories. Like SIR, PLS is target-directed, making the embedding tailored to information about X_k.

E.2.6. Computational considerations

Two practical decisions affect the computational footprint of the embedding benchmark. First, for the two distance-based methods—spectral clustering on Hamming affinity and agglomerative clustering on Hamming distance—the pairwise distance matrix requires $O(n^2)$ memory. For datasets exceeding 5,000 rows, we subsample to 5,000 observations (drawn without replacement using the global random seed) and assign out-of-sample observations to the cluster of their nearest subsampled neighbour. This cap ensures that the $n \times n$ distance matrix fits in RAM while retaining the full dataset size for all other methods.

Second, all methods that involve random initialisation (e.g., k-means, k-modes, spectral label assignment) use a fixed random seed of 42 to ensure reproducibility across runs and computing environments. The ACIB and Greedy CMI methods are deterministic given the data, so the seed affects only the subsampling step for distance-based methods and the initialisation of downstream clustering routines.

Amalgamation of the multivariate PID lattice. This appendix gives the full antichain classification rule summarised in Section 6.2.5. Each multivariate atom corresponds to an antichain: a collection of source subsets, none of which contains another. We first label each subset according to whether it contains only X_i (label "A"), only non-focal sources (label "B"), or both X_i and at least one other source (label "AB"). We then apply a conservative classification rule: if any A or B label is present in the antichain, all AB labels are discarded, and the atom is assigned based on the remaining labels:

- source-unique if only A labels remain;
- remainder-unique if only B labels remain;
- redundancy if both A and B labels remain;
- synergy only if the antichain contained exclusively AB labels (i.e., no pure A or B subset existed).

This rule is conservative with respect to synergy: an atom is classified as synergistic only when every subset in its antichain necessarily spans both the focal source and the remainder group, with no subset attributable to either alone. When a pure-A or pure-B subset coexists with an AB subset, the atom is attributed to the group that can account for it without invoking cross-group interaction. The grouping preserves the two-source consistency relations: the amalgamated source-unique plus redundancy equals $I(X_i; X_t)$, and the amalgamated remainder-unique plus redundancy equals $I(\{X_j : j \neq i\}; X_t)$.

Software. PID was computed using the `dit` Python library [166], with the I_{mmi} (minimum mutual information) redundancy measure as the primary PID specification. As robustness checks we additionally computed decompositions with the $I_{\min}$ (Williams–Beer) [22], $I_{\pm}$ (Finn–Lizier) [73], and $I_{\wedge}$ (Gács–Körner) [190] measures. For $N \geq 5$ sources,

$I_\pm$ and $I_\wedge$ were computed via the fast Möbius transform using precomputed lattice data from the `algebraicPID` package [191]. Mutual information was estimated using plug-in frequency-based estimators via `npeet` [306]; conditional mutual information for the PHQ-9 analyses was computed using `pyitlib`. Rank-based partial correlations for the PHQ-9 analyses were computed using the `pingouin` package [308].

E.2.7. Normalisation and effect-size scaling

Because information-theoretic measures are expressed in bits and are bounded above by the uncertainty of the target, we report normalised effect sizes to facilitate comparisons across targets and across cohorts.

For conditional mutual information (CMI), we report the target-normalised quantity

$$\mathrm{nCMI}_{X\to Y} = \frac{I(X;Y \mid \mathbf{Z})}{H(Y)} \in [0,1], \tag{E.1}$$

where $\mathbf{Z}$ denotes the remaining symptoms.

For embedding-based PID atoms about a given target Y, we report shares of the target entropy,

$$\pi^{(\cdot)}_{X\to Y} = \frac{\mathrm{Atom}^{(\cdot)}(X, E_{X\to Y}; Y)}{H(Y)} \in [0,1], \tag{E.2}$$

where $\mathrm{Atom}^{(\cdot)}$ denotes source-unique, remainder-unique, redundant, or synergistic information from the two-source decomposition on $(X, E_{X\to Y}; Y)$, and $E_{X\to Y}$ is the embedding of the remaining symptoms. Both quantities lie in $[0,1]$ and read as the fraction of the target's marginal uncertainty explained via a given information channel.

PHQ-9 items are ordinal and empirically skewed toward low-severity response categories, which reduces their effective entropy and constrains the attainable magnitude of mutual information and related quantities; this motivates reporting normalised effect sizes rather than raw bits. The nCMI normalisation (Eq. E.1) is conservative because $H(Y) \geq H(Y \mid \mathbf{Z})$, and it allows nCMI to be interpreted as the fraction of the target's marginal uncertainty associated with the source after conditioning on the rest of the symptom set. The PID shares (Eq. E.2) read as "the fraction of $H(Y)$ explained via a given information channel" and are reported alongside unnormalised totals in bits when needed.

Visualisation pruning. Statistical significance alone can yield dense networks in large samples (especially in UK Biobank). To obtain readable network figures and to enable direct cross-cohort comparison, we apply a two-stage rule: (i) edge screening by CMI significance, followed by (ii) rank-based pruning for visualisation by retaining, for each target, the incoming edges with the largest normalised effect size $\mathrm{nCMI}_{X\to Y}$ (Eq. E.1). The number of retained edges per target is specified in the figure captions.

E.3. Embedding Background and Computational Scaling

Relation to information bottleneck and redundancy bottleneck. In the main analyses with real PHQ-9 data, we use the Agglomerative Conditional Information Bottleneck

(ACIB) embedding, which is target-directed: it iteratively merges states of the multivariate remainder to minimise information loss about X_k under the fixed-cardinality constraint. ACIB is estimated separately for each (X_i, X_k) pair using $\mathbf{X} \setminus \{X_i, X_k\}$ as inputs and X_k as the supervised target. ACIB can be viewed as an instance of the information bottleneck principle [309]: it constructs a compressed discrete representation of the multivariate remainder $E_{i\to k} = f_{i\to k}(\mathbf{X} \setminus \{X_i, X_k\})$ that preserves information about the target X_k. More precisely, ACIB solves a discrete variant of the information bottleneck: it seeks to minimise $I(Z_{i\to k}; E_{i\to k})$ (compression) subject to the constraint $I(E_{i\to k}; X_k) \geq (1-\delta) \cdot I(Z_{i\to k}; X_k)$, where δ is a loss tolerance set to 5% in our analyses. The greedy merging phase traverses the IB rate–distortion curve from high cardinality (many clusters) to low cardinality, stopping when further compression would violate the information-preservation constraint. The 5% loss tolerance corresponds to operating near the "knee" of the IB curve, the regime where additional compression incurs disproportionately large information loss about the target. This connection provides principled guidance for the tolerance hyperparameter: it should be set small enough to preserve target-relevant structure while allowing sufficient compression for tractable PID computation. Recent theoretical work has further shown that a principled notion of PID redundancy can itself be formulated as a bottleneck trade-off via the *redundancy bottleneck*, yielding redundancy–compression curves and efficient optimisation procedures [187]. Whereas redundancy-bottleneck formulations provide a decision-theoretic grounding for redundancy, we use bottleneck-style compression as a practical approximation step that enables scalable two-source PID (including synergy) in high-dimensional symptom systems. Because $E_{i\to k}$ is learned without access to the focal source X_i, a theoretical failure mode arises when the multivariate remainder is informative about X_k primarily through its interaction with X_i (i.e., interaction-only motifs); in such settings, any lossy compression of the remainder can attenuate synergy estimates. In the present application, symptoms exhibit broad marginal dependence and do not resemble near-deterministic XOR-like constructions in which marginal associations vanish, reducing the plausibility of this edge case for PHQ-9 networks.

Computational scaling. A full n-source PID requires bookkeeping over a lattice whose number of atoms grows according to the Dedekind numbers, making exact multivariate decompositions infeasible beyond small n. By contrast, our approximation replaces the multivariate remainder by a low-cardinality discrete embedding and then computes a two-source PID on $(X_i, E_{i\to k}; X_k)$. For fixed item cardinality K and embedding size $|E_{i\to k}| = K$, the PID step depends only on a K^3 contingency table, while the dominant cost lies in constructing $E_{i\to k}$ for each ordered pair (i,k). This yields a scalable $O(p^2)$ pipeline in the number of symptom pairs, with the per-pair embedding cost determined by the number of observed joint states of the remainder.

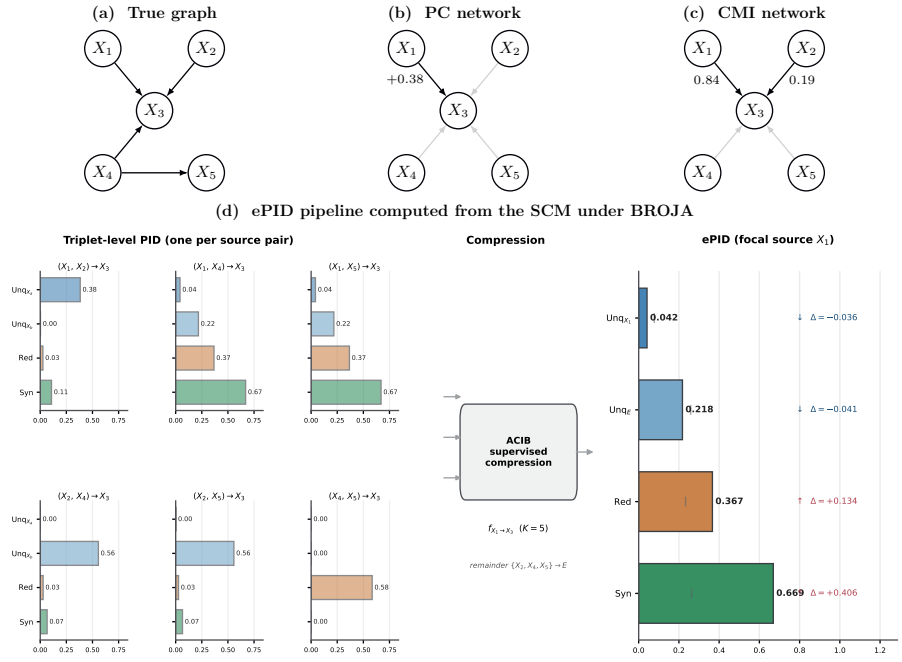


Figure E.1: **Methods comparison under BROJA.** Companion to Figure 6.1, identical structural model and pipeline, with the bottom-row two-source PIDs computed under the BROJA redundancy measure [105] instead of $I_{\min}$ [22]. The top row (panels **a–c**: true graph, PC, CMI) is measure-independent. On the homogeneous source pair (X_1, X_2), both source variables live in the h-channel of X_3 and the two measures agree to several decimals. On heterogeneous source pairs (X_a, X_b) where one source lives in the h-channel and the other in the W-channel of X_3, $I_{\min}$ allocates information across redundancy and synergy while BROJA assigns the same information to pure uniques: a known divergence between the two measures that motivates BROJA's operationalist definition. BROJA = Bertschinger–Rauh–Olbrich–Jost–Ay; PC = partial correlation; CMI = conditional mutual information.

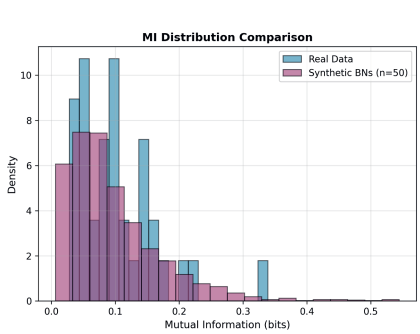
E.3.1. Toy SCM: methods comparison under BROJA

Figure E.1 recomputes the bottom-row ePID pipeline of Figure 6.1 under the BROJA redundancy measure [105] on the identical SCM. The top-row panels (true graph, PC, CMI) are measure-independent and therefore reproduced unchanged from that figure. The redundancy lattice underlying the type-1 amalgamation, alongside the truth-versus-ePID recovery for the focal source X_1, is shown in Figure 6.3.

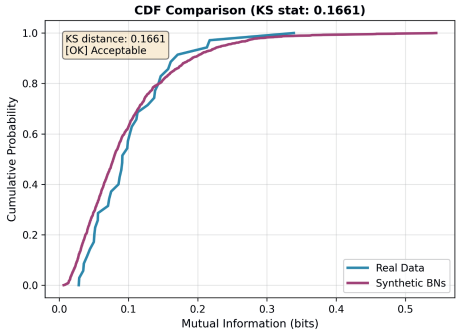
E.4. Bayesian Network Calibration Statistics

The following figures report calibration diagnostics for the ensemble of 50 synthetic Bayesian networks used in the embedding benchmark (Section 6.2.5). We compare the mutual information distributions of the synthetic networks against the empirical UK

Biobank data to verify that the synthetic data preserve the statistical properties relevant to PID computation.

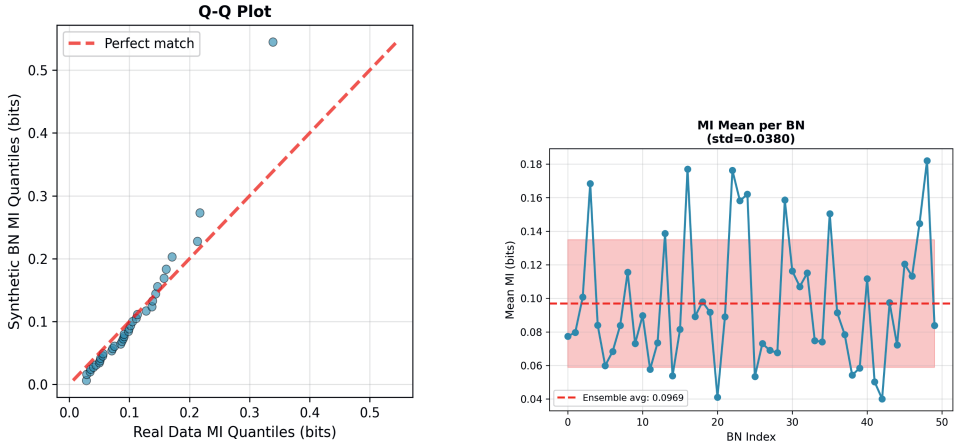


(a) MI Distribution Comparison



(b) CDF Comparison (KS stat: 0.166)

Figure E.2: **Mutual information distribution comparison between real data and synthetic Bayesian network ensemble.** (a) Histogram overlay showing density distributions of MI values. (b) Cumulative distribution functions (CDFs) with Kolmogorov–Smirnov (KS) statistic quantifying the distributional match. The ensemble of 50 synthetic BNs closely matches the real-data distribution, validating the calibration procedure.



(a) Q-Q Plot

(b) Ensemble Variability

Figure E.3: **Detailed validation of Bayesian network ensemble calibration.** (a) Quantile–quantile (Q–Q) plot comparing MI quantiles between real data and synthetic BNs. Points close to the diagonal indicate good distributional match. (b) Mean MI per network across the ensemble, showing consistent calibration with low variability (std = 0.0382). The red dashed line indicates the ensemble average, with shaded region showing ±1 standard deviation.

E

E.5. Multi-Dataset Benchmark: Full Results and Statistical Tests

This appendix reports the full multi-dataset embedding benchmark whose headline is summarised in Section 6.3.1: per-measure fidelity and method rankings, the $I_{\pm}$ exception, and the omnibus and post-hoc statistical tests.

E.5.1. Multi-dataset embedding benchmark

To assess embedding fidelity beyond the synthetic benchmark, we evaluated 13 embedding methods across 83 real-world datasets, 5 PID measures, and $N \in \{3,4,5\}$ sources (Section 6.2.6). Figure E.5 summarises the results across all methods; Table E.3 reports key statistics for the recommended method (ACIB) at $N = 5$.

Embedding fidelity is measure-dependent. At $N = 5$, mean relative error (RE) for ACIB ranged from 19.4% ($I_{\wedge}$) to 115.1% ($I_{\pm}$) (Table E.3), with corresponding mean absolute errors (MAE) of 0.057–0.247 bits. To put these in scale, typical ground-truth atom magnitudes are 0.05–0.20 bits, so an RE of 45% reflects MAE on the same order as a single atom rather than near-total signal loss; for comparison, a trivial "predict-zero" baseline yields RE > 130% for $I_{\min}$. $I_{\wedge}$ (Gács–Körner) was best preserved because its decomposition is sparsest, with most atoms exactly zero and the few non-zero atoms recovered with

<13% per-atom RE. $I_{\min}$ and I_{mmi} showed moderate distortion (RE ≈ 45%; AE ≈ 0.10 bits), driven almost entirely by the remainder-unique and synergy atoms (Section E.5.1); the redundancy and source-unique atoms were recovered with <3% RE in absolute terms. I_{rr} showed substantial distortion (RE = 70.9%), and $I_{\pm}$ errors exceeded the total signal (RE > 100%) due to a signed-atom cancellation failure analysed in Section E.5.1, which persists even when the dominant remainder-unique error component is removed.

ACIB is the recommended default for non-$I_{\pm}$ measures. ACIB delivered the lowest approximation error on the largest share of datasets across non-$I_{\pm}$ measures, and its advantage grew with the number of sources. To make this claim concrete we ranked the 13 embedding methods on each dataset by their per-atom absolute error and counted, for every measure×N combination, how often each method ranked first. Across all 15 measure×N combinations ACIB ranked first in 8 (all $N \geq 4$ except $I_{\pm}$); pooling per-dataset comparisons, ACIB ranked first in 44.3% of 1,141 cases overall, and in 55.7% once $I_{\pm}$ is excluded. At $N = 5$, ACIB ranked first on 84.4% of datasets for $I_{\wedge}$ (95% CI: [78.4, 91.1]%), 77.9% for $I_{\min}$ [71.1, 85.0], and 75.3% for I_{mmi} [68.6, 82.9]; the next-best methods (CMIM: 10.5%, ReliefF: 7.4%) trailed substantially. ACIB's advantage is not that it is dramatically better at any single N, but that it degrades most slowly as N grows: its error growth slope was 0.041 [95% CI: 0.035, 0.046], versus 0.044 [0.039, 0.049] for SIR and PLS, and 0.046 [0.040, 0.053] for ReliefF. At $N = 3$, where the embedding compresses only two sources, methods were statistically indistinguishable (Friedman $\chi^2 = 12.0$, $p = 0.45$). Figure E.4 summarises method ordering on the synthetic benchmark by per-pair Pearson correlation between approximated and ground-truth synergy; the family-level separation is visible at a glance, and the win-rate ordering reported above is preserved.

Number of sources matters more than method choice. Approximation error roughly tripled from $N = 3$ to $N = 5$ regardless of method. At $N = 3$, methods were indistinguishable (as noted above; Nemenyi post-hoc: 0 significant pairs for all measures). At $N = 4$ and $N = 5$, method differences became highly significant (Friedman $p < 10^{-22}$). The Nemenyi post-hoc test, which compares mean ranks across all $\binom{13}{2} = 78$ method pairs while controlling the family-wise error rate, identified 14–37 significant pairs at $N = 4$ [310]. This pattern—visible as convergent lines at $N = 3$ separating sharply at higher N in Figure E.5—indicates that the fundamental compression challenge of higher-dimensional remainder sets dominates the choice of compression algorithm. Full statistical test results are provided in the following subsections.

The $I_{\pm}$ exception

The high relative-error figures for $I_{\pm}$ reflect a structural mismatch between $I_{\pm}$'s information-flow semantics and the compression that any ePID pipeline performs. $I_{\pm}$ decomposes mutual information *pointwise*, atom by atom over individual outcomes, and the resulting signed atoms encode local information flow that cancels in structured ways across the joint distribution [73]. At $N = 5$, 77.7% of ground-truth Unq_{S_1} values were

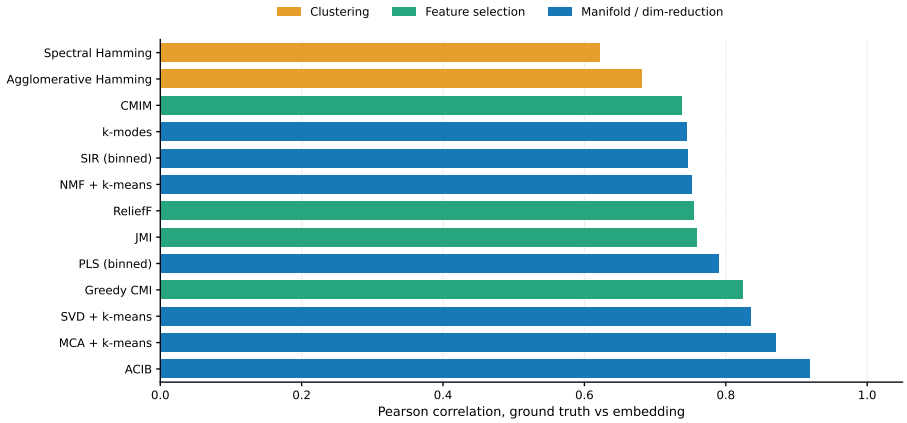


Figure E.4: **Ranking of 13 embedding methods by ground-truth synergy correlation.** Pearson correlation r between embedding-approximated and ground-truth synergy across the synthetic Bayesian network benchmark (50 networks × 9 targets × 3 source counts $N \in \{3,4,5\}$, $n = 5{,}377$ source–target configurations; type-1 amalgamation throughout). Bars are coloured by method family: *manifold / dim-reduction* (blue), *feature selection* (green), *clustering* (orange). ACIB ranks first by this criterion; manifold-family methods occupy the top six positions. ACIB = Agglomerative Conditional Information Bottleneck; JMI = Joint Mutual Information; CMIM = Conditional Mutual Information Maximisation; SIR = Sliced Inverse Regression; MCA = Multiple Correspondence Analysis; PLS = Partial Least Squares; SVD = Singular Value Decomposition; NMF = Non-negative Matrix Factorisation.

negative (compared to 0.7% for $I_{\min}$). Embedding methods are designed to compress the joint distribution of the remainder into a single discrete summary; this compression is well-suited to atoms that average over the joint distribution, but it aggregates over precisely the local outcome-level structure that $I_{\pm}$ relies on, so cancellations no longer line up after compression. The mismatch is therefore conceptual, not a tunable embedding choice.

Practical recommendation. Based on the benchmark, we recommend ACIB for analyses using $I_{\min}$, I_{mmi}, I_{rr}, or $I_{\wedge}$ (best at $N \geq 4$); for $I_{\pm}$, ReliefF ($N \leq 4$) or SVD+k-means ($N = 5$) achieved the lowest absolute errors in our runs, but the structural mismatch means embedding-based estimates of $I_{\pm}$ should be interpreted as approximate compositional profiles rather than exact quantities. At $N = 3$, method choice is immaterial. The empirical PHQ-9 analyses in this chapter use I_{mmi}, for which ACIB is the best-validated embedding.

E.5.2. Friedman omnibus test

Table E.4 reports the Friedman test statistic and p-value for each combination of PID measure and number of sources. The Friedman test assesses whether the ranking of embedding methods differs significantly across datasets.

Table E.3: **ACIB embedding fidelity by PID measure at $N = 5$ sources.** Mean absolute error (MAE, bits) is the per-atom absolute error averaged across the four PID atoms (source-unique, remainder-unique, redundancy, synergy) and across datasets. Mean relative error (RE) is the same quantity divided by the magnitude of the corresponding ground-truth atom. Win rate is the proportion of per-dataset contests (across 77–83 datasets) in which ACIB achieved the lowest error among the 13 candidate embeddings, with bootstrap 95% confidence intervals. For I_{rr}, 12 of the 13 embeddings were available (greedy CMI lacked an I_{rr} ground truth).

Measure	MAE (bits)	RE (%)	Win rate (%)	95% CI
$I_\wedge$	0.057	19.4	84.4	[78.4, 91.1]
I_{min}	0.105	44.7	77.9	[71.1, 85.0]
I_{mmi}	0.106	45.2	75.3	[68.6, 82.9]
I_{rr}	0.146	70.9	64.9	[56.2, 73.3]
$I_\pm$	0.247	115.1	2.6	[0.0, 4.4]

Table E.4: **Friedman test results by PID measure and number of sources.** χ^2 statistic and p-value testing whether method rankings differ across datasets. At $N = 3$, no significant differences were detected for any measure.

	$N = 3$		$N = 4$		$N = 5$	
Measure	χ^2	p	χ^2	p	χ^2	p
I_{min}	12.0	0.45	131.4	9×10^{-23}	145.9	1×10^{-25}
I_{mmi}	12.0	0.45	128.6	3×10^{-22}	131.1	1×10^{-22}
I_{rr}	12.0	0.45	134.9	2×10^{-23}	37.6	9×10^{-5}
$I_\pm$	12.0	0.45	252.7	8×10^{-48}	237.7	1×10^{-44}
$I_\wedge$	12.0	0.45	181.8	4×10^{-33}	238.3	8×10^{-45}

E.5.3. Nemenyi post-hoc tests

Table E.5 reports the number of significantly different method pairs (out of 78 possible) identified by the Nemenyi post-hoc test following significant Friedman results.

Table E.5: **Number of significantly different method pairs (Nemenyi post-hoc).** Out of 78 possible pairwise comparisons between 13 methods. Zero pairs at $N = 3$ confirms that all methods perform equivalently when compressing only two sources.

Measure	$N = 3$	$N = 4$	$N = 5$
I_{min}	0	15	22
I_{mmi}	0	14	19
I_{rr}	0	15	6
$I_\pm$	0	37	36
$I_\wedge$	0	23	35

E.5.4. Supplementary benchmark figures

Figure E.5 reports embedding approximation error across all 13 methods, 5 PID measures, and $N \in \{3,4,5\}$ sources.

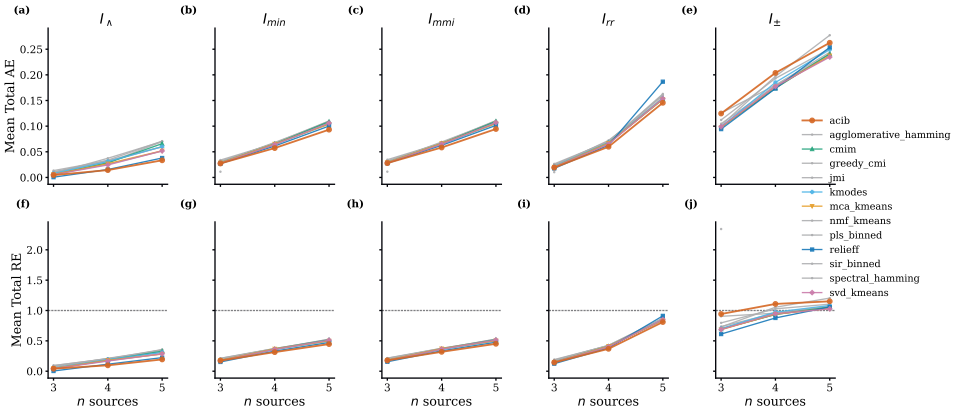


Figure E.5: **Embedding approximation error across PID measures and number of sources (all 13 methods).** Top row: mean absolute error (bits); bottom row: mean relative error (%). Each column corresponds to one PID measure, ordered by increasing difficulty. ACIB (bold orange) is the lowest-error method at $N \geq 4$ for all measures except $I_\pm$. The dashed line at RE = 100% marks the threshold where approximation error equals the total signal. I_{rr} at $N = 5$ is based on 23 datasets; all others on 77–83 datasets.

Figure E.6 reports the relative-error analogue of Figure 6.4. Normalising each absolute error by the magnitude of the corresponding ground-truth atom is informative because absolute errors are bounded above by atom magnitude: methods that appear comparable in bits can diverge sharply once normalised. The relative-error ranking is consistent with the absolute-error ranking, with ACIB dominant across all four atoms; the gap between manifold and clustering families widens slightly under relative-error scaling.

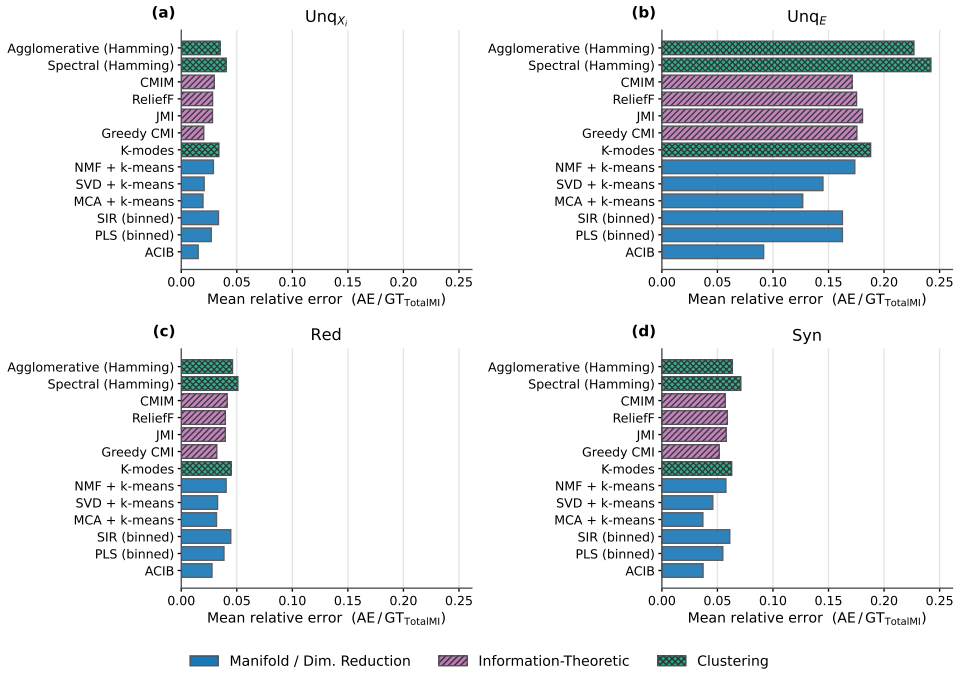


Figure E.6: **Mean relative error by embedding method and PID atom.** Per-atom error normalised by the magnitude of the corresponding ground-truth atom, complementing the absolute-error view in Figure 6.4. Panels, methods, and family colour coding are identical to that figure. Relative-error normalisation makes small absolute errors visible when ground-truth atoms are small in magnitude; the ranking of methods within each panel is preserved against the absolute-error view. ACIB = Agglomerative Conditional Information Bottleneck.

E.5.5. Atom-level error profile across PID measures

For non-$I_\pm$ measures, embedding error in the multi-dataset benchmark is concentrated in the atoms that involve the compressed remainder: Unq$_E$ (remainder-unique) and Syn (synergy) carry most of the approximation error, while Unq$_{S_1}$ (focal-source unique) and Red (redundancy) are well-preserved (Fig. E.7). For $I_\pm$, all four atoms exhibit high error, consistent with the disruption of signed cancellation across the full decomposition discussed in Chapter 6.

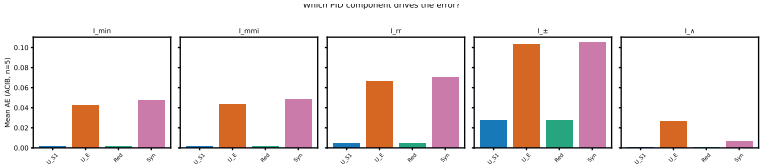


Figure E.7: **Atom-level error profiles by PID measure (ACIB, $N = 5$).** For non-$I_\pm$ measures, error is concentrated in the remainder-dependent atoms (Unq_E, Syn). For $I_\pm$, all atoms show high error due to disruption of signed cancellation structure.

E.6. Additional Fidelity Diagnostics and Symptom-Level Examples

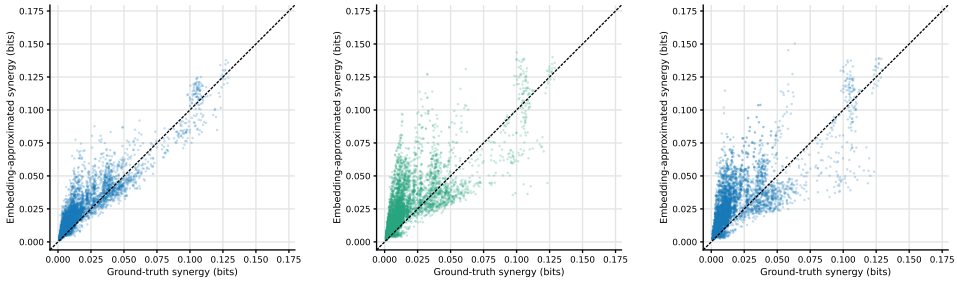
This appendix collects three diagnostic items referenced from Chapter 6: per-pair synergy fidelity for representative methods (Section E.6.1), the scaling of synergy approximation error with the number of embedded sources (Section E.6.2), and target-centred PID examples for two representative PHQ-9 symptoms in the Xinxiang student sample (Section E.6.3).

E.6.1. Per-pair synergy fidelity for representative methods

Figure E.8 reports per-pair scatter for one representative method from each family tier, illustrating the spread visible in the aggregate atom errors of Figure 6.4 and in the 13-method ranking of Figure E.4. ACIB (manifold/dim-reduction) aligns closely with the identity line ($r = 0.98$); JMI (feature selection) shows moderate scatter ($r = 0.41$); k-modes (clustering) aligns reasonably well ($r = 0.90$), illustrating the upper end of clustering performance. Hamming-based clustering methods (collapse to $r \approx 0$; see Figure E.4) are not shown in this trio.

E.6.2. Synergy approximation error scaling with the number of sources

For ACIB on the synthetic Bayesian network ensemble, mean absolute error in the synergy atom increased smoothly from 0.0008 bits at $N = 3$ to 0.0056 bits at $N = 5$, remaining below 0.006 bits throughout (Fig. E.9). A sliced inverse regression (SIR) embedding with binning showed slightly higher but remarkably stable synergy error across N ($\approx$ 0.0045–0.0048 bits), whereas linear SVD+k-means and clustering methods had both larger baseline errors and steeper increases with N.



(a) ACIB ($r = 0.98$) (b) JMI ($r = 0.41$) (c) k-modes ($r = 0.90$)

Figure E.8: **Approximated versus ground-truth synergy for one representative method per family tier.** Per-pair scatter of embedding-approximated synergy (y-axis) against ground-truth synergy (x-axis), in bits, across 5,377 (G, X_t, N) configurations from the synthetic Bayesian network benchmark. Each panel shows one representative method per family. **(a)** ACIB, manifold/dim-reduction family, Pearson $r = 0.98$. **(b)** JMI, feature-selection family, $r = 0.41$. **(c)** k-modes, clustering family, $r = 0.90$. Dashed diagonal is the identity line. The visible density clustering near ground-truth synergy ≈ 0.10 reflects multiple synthetic networks producing similar synergy values. Methods within each family vary in fidelity; Figure E.4 ranks all 13 methods. The corresponding scaling of synergy approximation error with the number of embedded sources is reported in Section E.6.2. ACIB = Agglomerative Conditional Information Bottleneck; JMI = Joint Mutual Information.

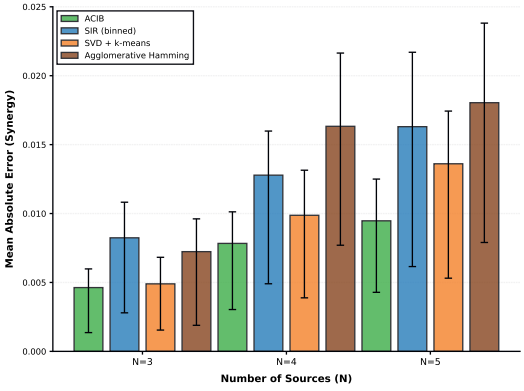
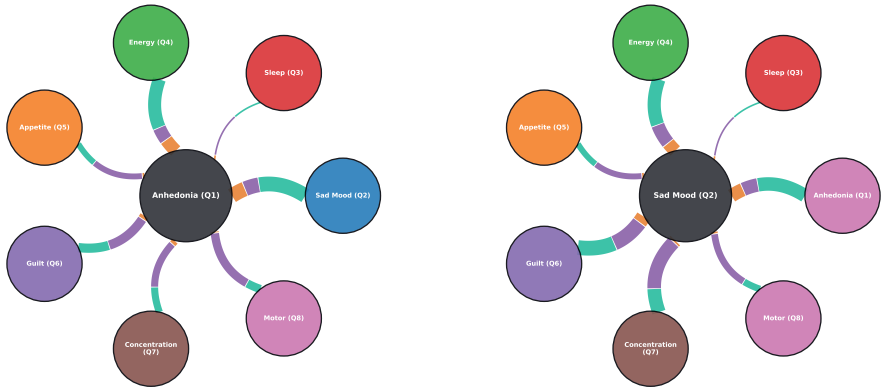


Figure E.9: **Synergy approximation error as a function of the number of embedded sources** ($N = 3, 4, 5$)**.** ACIB error increases sub-linearly from 0.0008 bits at $N = 3$ to 0.0056 bits at $N = 5$.

E.6.3. Symptom-level PID examples in the Xinxiang student sample

Figure E.10 visualises the embedding-based PID for two representative targets in the Xinxiang student sample (anhedonia, sad mood) as target-centred radial graphs. The central dark grey node represents the target; surrounding coloured nodes represent source variables. Each edge is segmented into four PID components (teal: redundancy; blue: unique source; purple: unique embedding; orange: synergy), with edge thickness scaled to total mutual information. These examples illustrate how strongly connected affective

symptoms (e.g., anhedonia and sad mood) combine substantial unique contributions with non-negligible synergy, whereas somatic symptoms such as sleep and motor disturbance more often contribute via synergy or redundancy rather than purely unique channels.



(a) Anhedonia as target

(b) Sad mood as target

Figure E.10: **Target-centred PID radial graphs for two representative targets in the Xinxiang student sample.** Each edge is segmented into four PID atoms (teal: redundancy; blue: unique source; purple: unique embedding; orange: synergy); edge thickness scales with total mutual information.

E.7. Dependency Structure in Clinical Datasets

To assess whether non-linear conditional dependencies are prevalent in the types of datasets used for symptom network analysis, we compared three unconditional bivariate dependency measures across 80 clinical and psychometric datasets (32,889 variable pairs in total).

For each dataset, we computed (i) absolute Pearson correlation on the imputed continuous data, (ii) absolute Spearman correlation on rank-transformed data, and (iii) mutual information on discretised data (8 equal-frequency bins). All three measures are unconditional and bivariate, ensuring direct comparability without confounding by conditioning structure. Statistical significance was assessed using permutation tests (1,000 permutations per dataset, columns shuffled independently) with Benjamini–Hochberg FDR correction at $\alpha = 0.05$, applied uniformly across all three measures.

Figure E.11 summarises the results. At the level of statistical significance, Pearson and Spearman correlation agreed almost perfectly (panel a): fewer than 1% of significant pairs were detected by one but not the other, confirming that monotonic-but-non-linear dependencies are negligible in these data. Mutual information detected fewer significant pairs than either correlation measure (panel b), consistent with discretisation-induced

information loss rather than with MI capturing additional non-linear structure. Across all 80 datasets, only 193 of 30,467 MI-significant pairs (0.6%) were detected by MI but not by either correlation measure (panel c), indicating that non-linear dependencies invisible to correlation are rare in clinical and psychometric datasets with ordinal response scales. These results suggest that the linearity assumption underlying partial correlation networks is largely met in practice for ordinal symptom items. The primary motivation for information-theoretic methods in this context is therefore not the detection of non-linear pairwise associations, but the decomposition of multivariate dependence into redundant, unique, and synergistic components—a distinction that correlation-based methods cannot provide regardless of functional form.

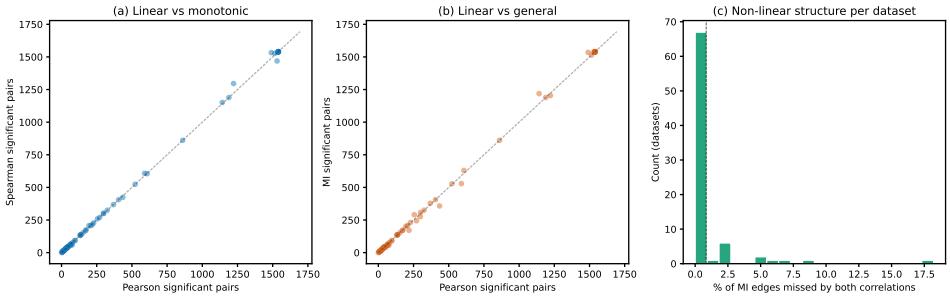


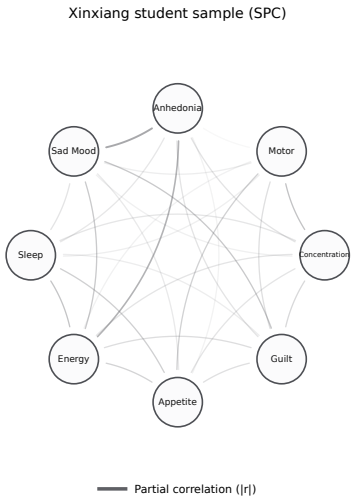
Figure E.11: **Comparison of bivariate dependency measures across 80 clinical and psychometric datasets.** (a) Number of significant pairs detected by Pearson vs. Spearman correlation per dataset; near-perfect agreement indicates negligible non-linear monotonic structure. (b) Number of significant pairs detected by Pearson correlation vs. mutual information; MI generally detects fewer pairs, consistent with discretisation-induced information loss. (c) Distribution of the percentage of MI-significant pairs not detected by either correlation measure; the vast majority of datasets show fewer than 2.5% non-linear-only dependencies.

E.8. Cross-cohort Comparison of Information-Theoretic PHQ-9 Networks

Figure E.15 quantifies cross-dataset agreement in dependence strength and PID composition. Edge-wise differences in TMI are systematically positive (reflecting larger absolute dependence in UK Biobank), whereas synergy-proportion differences are concentrated near zero, and the synergy proportions show substantial cross-dataset agreement (Pearson $r = 0.663$; Spearman $\rho = 0.777$).

E.8.1. Edge definition and comparability across cohorts

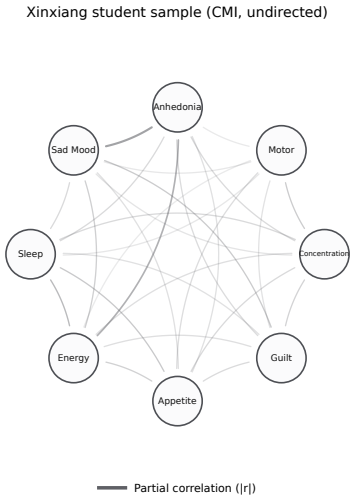
We compared information-theoretic symptom networks estimated separately in two datasets using the same eight PHQ-9 symptom variables. For each symptom pair (X, Y), we estimated the conditional mutual information (CMI) $I(X;Y \mid \mathbf{Z})$, where $\mathbf{Z}$ denotes the remaining six symptoms. The "significant edge" sets reported below correspond to the cohort-specific CMI screening procedure exported in the significant-edge files. Although



(a) Pearson Partial Correlation Network



(b) Spearman Partial Correlation Network

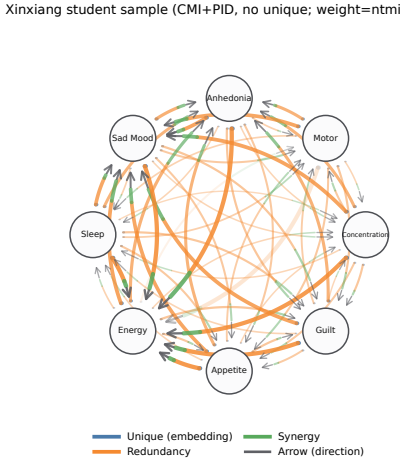


(c) Conditional Mutual Information Network

Figure E.12: **Xinxiang student sample PHQ-9 networks under three association measures.** Three measures that progressively relax modelling assumptions of linearity and monotonicity, each leading to a similar network structure: (a) Pearson partial correlation, (b) Spearman partial correlation, (c) conditional mutual information.

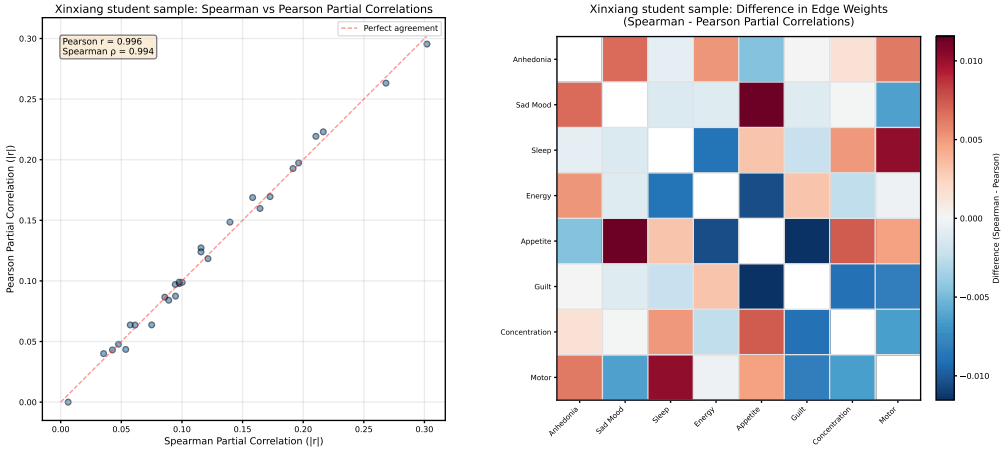


(a) Full ePID view (all four atoms)



(b) No-remainder-unique view (redundancy and synergy only)

Figure E.13: **Two ePID network views in the Xinxiang student sample.** Directed edges represent ordered source→target relations. **(a) Full ePID view:** each edge is partitioned into remainder-unique (blue), redundancy (orange), and synergy (green) atoms from the two-source PID of $(X_i, E_{i\to k}; X_k)$; the terminal black segment indicates direction. **(b) No-remainder-unique view:** the remainder-unique atom is removed and the remaining redundancy and synergy atoms are rescaled, highlighting dependence that involves the focal source beyond remainder-only predictability. Edge thickness encodes target-normalised total information $I(X_i, E_{i\to k}; X_k)/H(X_k)$.



(a) Edge-wise scatter comparison

(b) Edge-wise difference heatmap

Figure E.14: **Agreement between partial Pearson and partial Spearman correlation networks in the Xinxiang student sample.** Partial Pearson correlations (PPC) and Spearman partial correlations (SPC) were computed for each symptom pair while conditioning on all remaining PHQ-9 symptoms (suicidality excluded). **(a)** Edge-wise comparison of absolute partial correlations ($|r|$) shows near-identical weighting across pairs (Pearson $r = 0.997$; Spearman $\rho = 0.995$). **(b)** Edge-wise differences (SPC−PPC) are small across the matrix, indicating that rank-based partial correlation and linear partial correlation yield highly similar conditional-association structures in this setting.

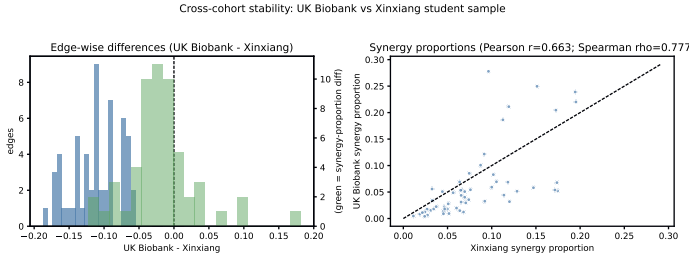


Figure E.15: **Cross-dataset comparison of information-theoretic dependence strength and decomposition stability (UK Biobank vs the Xinxiang student sample).** For each directed source→target symptom pair (eight PHQ-9 symptoms; suicidality excluded), we compare total mutual information (TMI) and the *synergy proportion* (synergy/TMI) derived from the embedding-based PID framework. Edge-wise differences (UK Biobank − Xinxiang) show systematically larger dependence strengths in UK Biobank, whereas synergy-proportion differences are concentrated near zero. Synergy proportions show substantial cross-dataset agreement (Pearson $r = 0.663$; Spearman $\rho = 0.777$), indicating that the *composition* of dependence is more stable across datasets than the overall dependence magnitude.

CMI is symmetric in X and Y given the same conditioning set, we represent edges as directed ($X \to Y$) for two reasons: (i) to align with the directional partial information decomposition (PID) that is defined for a designated target variable, and (ii) to permit visualisation of asymmetries in the PID atoms between the two orientations. Accordingly, edge direction in these plots should not be interpreted as causal direction.
Because the cohorts differed in marginal response distributions, their marginal entropies $H(X)$ also differed across items. Since CMI in bits is upper-bounded by target uncertainty (e.g., $I(X;Y \mid \mathbf{Z}) \leq H(Y)$), absolute thresholds expressed in bits are not directly comparable across cohorts when $H(Y)$ differs. To improve cross-cohort comparability, we therefore implemented an entropy-normalised effect size,

$$\mathrm{nCMI}_{X\to Y} = \frac{I(X;Y \mid \mathbf{Z})}{H(Y)}, \tag{E.3}$$

which can be interpreted as the fraction of the target's marginal uncertainty associated with the source after conditioning on the other symptoms. This normalisation is conservative because $H(Y)$ upper-bounds $H(Y \mid \mathbf{Z})$.

E.8.2. Do cohorts yield the same significant CMI edges?

At the level of statistical significance alone (i.e., without additional pruning), the UK Biobank cohort exhibited statistically significant CMI for all 56 directed symptom pairs among eight nodes, whereas the Xinxiang student sample exhibited 54 significant directed edges. Thus, the Xinxiang student sample's significant edge set was a strict subset of UK Biobank's. Table E.6 lists the 2 directed edges that were significant in UK Biobank but not in the Xinxiang student sample; equivalently, the Xinxiang student sample contained all remaining directed edges.

Table E.6: Directed CMI edges that were statistically significant in UK Biobank but not in the Xinxiang student sample (2 edges). UK Biobank contained all 56 directed symptom pairs; the Xinxiang student sample contained the complementary set of 54 edges.

Source	Target
Energy	Motor
Motor	Energy

E.8.3. Pruning rules for visualisation and "clinical relevance"

Statistical significance alone can yield dense graphs (notably in UK Biobank). To obtain readable figures while maintaining transparency, we evaluated several simple pruning rules applied *after* significance filtering. Table E.7 summarizes the number of retained edges in each cohort and the overlap between cohorts under: (i) an absolute cutoff on raw CMI (CMI≥ 0.05 bits), and (ii) rank-based rules computed on nCMI (top-10 edges overall, top quartile [top 25%], and top-3 incoming edges per target). The absolute CMI cutoff retained more edges in the Xinxiang student sample than in UK Biobank, consistent with the datasets' differing entropy scales. In contrast, rank-based rules on nCMI produced more comparable densities and larger overlaps, thereby supporting interpretable cross-dataset comparisons.

Table E.7: Overlap of statistically significant CMI edges across cohorts under alternative pruning rules. Rank-based rules use entropy-normalized CMI, nCMI = $I(X;Y \mid \mathbf{Z})/H(Y)$. The Jaccard index is $|E_\cap|/|E_\cup|$ for the directed edge sets.

Pruning rule	$\lvert E_{\text{Xinxiang}}\rvert$	$\lvert E_{\text{Biobank}}\rvert$	$\lvert E_\cap\rvert$	$\lvert E_\cup\rvert$	Jaccard
All significant (FDR)	54	56	54	56	0.96
CMI ≥ 0.05 bits	10	6	6	10	0.60
Top 10 by nCMI	10	10	6	14	0.43
Top 25% by nCMI	14	14	9	19	0.47
Top 3 per target by nCMI	24	24	17	31	0.55

E.8.4. Depression-status contrast: all significant edges

Figure E.16 reproduces the depression-status contrast of Figure 6.7 over all 56 significant directed edges rather than the top-three-per-target subset. The redundancy-to-unique/synergy shift is larger on the full edge set (mean shares: redundancy 0.26 → 0.11, unique 0.60 → 0.66, synergy 0.14 → 0.23), consistent with the pruned display in Chapter 6 but computed from the complete significant-edge network.

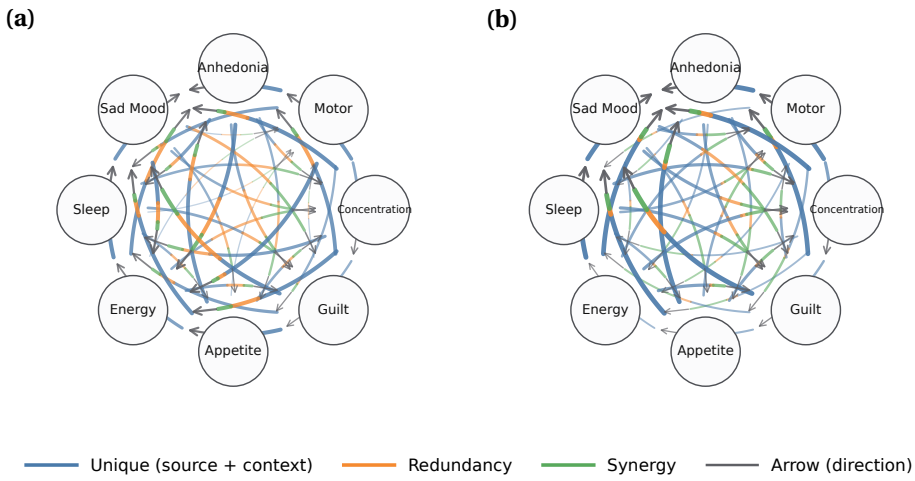


Figure E.16: **ePID networks in non-depressed versus depressed UK Biobank respondents (all significant edges).** Companion to Fig. 6.7, showing all 56 significant directed edges rather than the top-three-per-target subset. **(a)** non-depressed (PHQ-9 total < 10); **(b)** depressed (PHQ-9 total ≥ 10); equal-N matched at n = 8,879 per arm, five-seed average, eight PHQ-9 symptoms (suicidality excluded). Segment colours (unique = blue, redundancy = orange, synergy = green) and the shared total-mutual-information thickness scale are as in Fig. 6.7. Over all 56 edges the redundancy-to-unique/synergy shift is larger than for the pruned display (mean shares: redundancy 0.26 → 0.11, unique 0.60 → 0.66, synergy 0.14 → 0.23). As in the main figure, the absolute synergy share is N-dependent and the panels should be read as a matched-N contrast, not an absolute level.

E.9. IRI cross-subscale synergy and redundancy patterns

Figure E.17 provides a visual summary of the cross-subscale triplet PID analysis whose headline percentages appear in Section 6.3.4. The three panels report (a) synergy-dominance by source pair for Perspective-Taking targets, (b) redundancy ranking across source pairs pooled across targets, and (c) synergy-dominance rate per target subscale. Percentages match those quoted in Chapter 6 under the same informative-pair filter ($I(X_i, X_j; X_k) \geq 0.01$ bits) and the same cross-subscale criterion (one item from each of three distinct subscales as X_1, X_2, target).

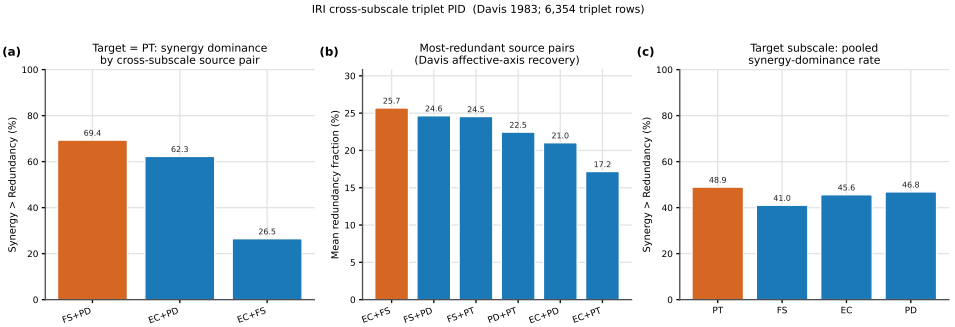


Figure E.17: **IRI cross-subscale synergy and redundancy patterns.** Each panel is computed over informative triplets (total mutual information $I(X_i, X_j; X_k) \geq 0.01$ bits) in which the source pair and target item belong to three distinct IRI subscales. **(a)** Synergy-dominance rate (Synergy > Redundancy) for triplets targeting a Perspective-Taking item, by cross-subscale source pair; FS+PD tops the ranking at 69.4%, consistent with the claim in Chapter 6 that mature cognitive empathy is informed synergistically by the affective extremes. **(b)** Mean redundancy fraction per source pair, pooled across all target subscales; EC+FS tops the ranking at 25.7%, and all three pairs containing Fantasy rank above all three pairs without it. **(c)** Synergy-dominance rate per target subscale, pooled across all cross-subscale source pairs; PT ranks highest at 48.9%, ahead of PD (46.8%), EC (45.6%) and FS (41.0%), consistent with developmental accounts in which cognitive empathy emerges later than its affective counterpart and comes to integrate it [211]. PT = Perspective Taking; FS = Fantasy Scale; EC = Empathic Concern; PD = Personal Distress.

E.10. Illustrative example: stress, sleep, and concentration

To illustrate how PC, CMI, and PID provide complementary information, we construct a hypothetical three-variable system: stress (X_2), sleep problems (X_3), and concentration (X_1), each on a five-level ordinal scale ($\{0,1,2,3,4\}$). Concentration closely tracks sleep quality under typical conditions ($X_1 = X_3$ when $X_3 \in \{0,1,2,3\}$), but when sleep is

extremely poor ($X_3 = 4$), concentration depends non-monotonically on stress[1]:

$$X_1 = \begin{cases} X_3, & \text{if } X_3 \in \{0,1,2,3\} \text{ (normal sleep)}, \\ 0, & \text{if } X_3 = 4,\ X_2 = 0 \text{ (very poor sleep, no stress)}, \\ 4, & \text{if } X_3 = 4,\ X_2 = 1 \text{ (very poor sleep, mild stress)}, \\ 3, & \text{if } X_3 = 4,\ X_2 = 2 \text{ (very poor sleep, moderate stress)}, \\ 2, & \text{if } X_3 = 4,\ X_2 = 3 \text{ (very poor sleep, high stress)}, \\ 1, & \text{if } X_3 = 4,\ X_2 = 4 \text{ (very poor sleep, extreme stress)}. \end{cases}$$

We simulate 10,000 samples and estimate the network using partial correlations, CMI, and PID (Figure E.18). The three methods reveal progressively richer structure:

- **Partial correlation:** Detects only the X_1–X_3 edge. The stress–concentration link is missed because it is non-monotonic and emerges through a synergistic interaction with sleep.
- **CMI:** Detects both the X_1–X_2 and X_1–X_3 edges, revealing that stress influences concentration conditionally—its effect appears only when sleep is very poor.
- **PID** (using $I_{\min}$, a redundancy measure formalised in Section 6.2.5)**:** Decomposes the joint influence into 71.5% unique to sleep, 21.1% synergistic between stress and sleep, 7.2% redundant, and 0.1% unique to stress. The synergy fraction quantifies precisely the interaction that PC misses and CMI detects only indirectly.

The deterministic structure of this example exaggerates the effect relative to empirical data. Nonetheless, the decomposition illustrates a general principle: when synergy is high, predictive models that accommodate interactions (e.g., random forests, which achieved perfect accuracy here) substantially outperform additive models (adjusted $R^2 = 0.45$ for linear regression). The information-theoretic decomposition thus provides principled guidance on when interaction terms are needed.

E.11. Robustness of the ACIB embedding to its hyperparameters

E.11.1. Question

The embedding step at the heart of ePID (the agglomerative conditional information bottleneck, ACIB) has three free hyperparameters: the conditional-mutual-information *loss tolerance* (default 5%), the embedding cardinality cap $K_{\max}$ (set to 12 in the main analyses), and a Dirichlet *smoothing* constant α (default 0.5). A reviewer will reasonably

[1] Note that this equation can be rewritten with use of the modulo operator, but we omit this for ease of explanation.

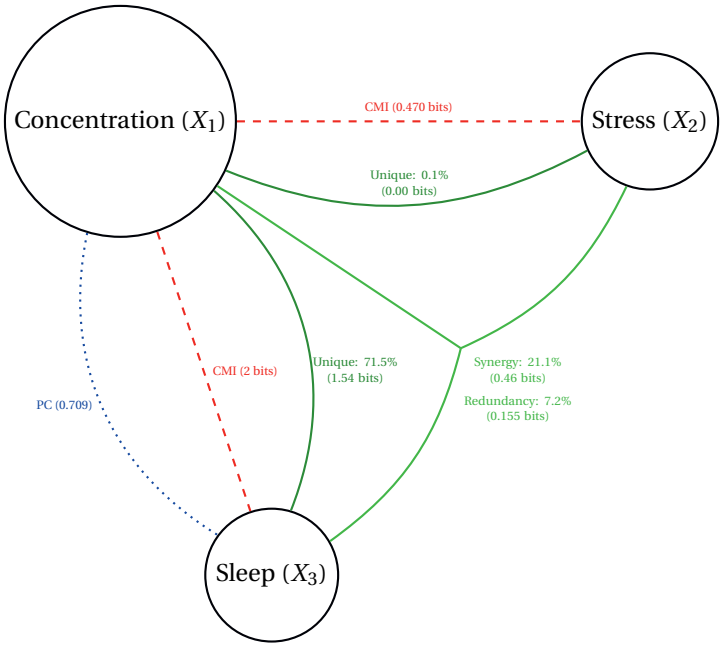


Figure E.18: **The same three-variable system as seen by three methods.** PC (blue, dotted) only recovers the linear X_1–X_3 edge; CMI (red, dashed) additionally detects the conditional X_1–X_2 dependence; PID (green, solid) further splits the joint information (2.16 bits total) into unique, redundant, and synergistic atoms, making explicit the synergy that PC misses and CMI detects only indirectly. This example is constructed for methodological demonstration and does not represent an empirically identified mechanism.

ask whether the decompositions we report are robust to these settings, or artefacts of a particular tuning. This appendix answers that with two complementary sensitivity analyses, and in doing so also justifies the choice to embed only $n_{\text{sources}} = 5$ remainder symptoms rather than the full network.

E.11.2. What we did

Fidelity against ground truth (synthetic). On the synthetic networks used to validate the embedding (which have known, exact ground-truth PID atoms), we re-ran ACIB across a 3×3 grid of loss tolerance $\in \{2\%, 5\%, 10\%\}$ and $K_{\max} \in \{6, 10, 14\}$, plus a one-dimensional slice over the smoothing constant $\alpha \in \{0.1, 0.5, 1.0\}$. For each setting we recorded the total absolute error of the recovered two-source atoms against ground truth (measure I_{mmi}, $N = 5$ sources), the realised embedding cardinality, and the retained conditional mutual information (the information-bottleneck "knee").

Stability of the empirical signal (cohorts). On the two empirical instruments (the Xinxiang student sample and the IRI empathy network) we recomputed the per-edge synergy

fraction across the same grid. We additionally contrasted the $n_{\text{sources}} = 5$ embedding protocol against a *full-context* embedding that folds all remaining items into a single variable.

E.11.3. What we found

(i) Fidelity depends only on $K_{\max}$, and the default is in the stable region. The approximation error is *insensitive* to the loss tolerance: it is identical to four decimal places across 2%, 5%, and 10% at every $K_{\max}$, and to the smoothing constant α. The only consequential knob is $K_{\max}$: error falls monotonically as more clusters are allowed, with clearly diminishing returns (Table E.8, Fig. E.19). The information-bottleneck retention curve plateaus at a realised cardinality of roughly ten (Fig. E.20). The cap used in the main analyses, $K_{\max} = 12$, therefore sits above the realised cardinality and within this stable region: it is statistically indistinguishable from the neighbouring grid values $K_{\max} = 10$ and $K_{\max} = 14$, which differ from one another by only 0.011 bits.

Table E.8: Mean total absolute error of the recovered atoms against ground truth ($N = 5$, I_{mmi}, $\alpha = 0.5$). Rows (loss tolerance) are identical; columns ($K_{\max}$) carry all the variation.

loss tolerance	$K_{\max} = 6$	$K_{\max} = 10$	$K_{\max} = 14$
2%	0.111	0.059	0.048
5%	0.111	0.059	0.048
10%	0.111	0.059	0.048

(ii) The empirical synergy contrast is not a hyperparameter artefact. Across the whole grid the per-edge synergy fraction is stable: 6.1 to 6.8% in the Xinxiang student sample and 25.5 to 26.4% in IRI. The headline contrast, namely depression networks redundancy-dominated and empathy networks synergy-dominated, therefore holds regardless of the embedding settings.

(iii) Full-context embedding saturates, validating $n_{\text{sources}} = 5$. Folding *all* remaining items into one embedding collapses the decomposition for the 28-item IRI network: synergy falls from 0.256 under $n_{\text{sources}} = 5$ to 0.044 under full context (Fig. E.21). For the 9-item PHQ-9 the change is modest (0.061 to 0.092, no collapse). This is the expected saturation: when too many items are conditioned on at once, the residual conditional dependence approaches zero and the atoms degenerate into numerical noise. Limiting the embedding to a small remainder context avoids this regime.

E.11.4. What we conclude

The reported decompositions are robust to the ACIB hyperparameters. The loss tolerance and the smoothing constant have no measurable effect on fidelity; the only meaningful knob, $K_{\max}$, sits in the stable region at its main-analysis value of 12, above the realised

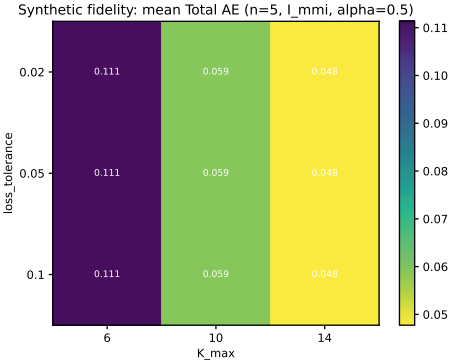


Figure E.19: Fidelity surface (mean total absolute error). $K_{\max}$ is the only active knob.

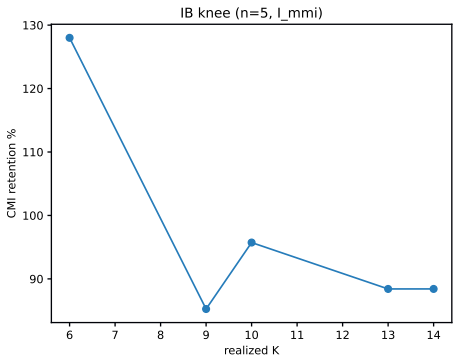


Figure E.20: Information-bottleneck retention plateaus near a realised cardinality of ten.

embedding cardinality. The empirical synergy estimates, and the redundancy-versus-synergy contrast between instruments, are stable across the grid. Finally, the saturation of the full-context embedding for large networks independently justifies the decision to embed only a bounded remainder context.

E

E.11.5. Caveats

- The synthetic sweep at $N = 5$ was run for I_{mmi}; the other measures ($I_{\min}$, $I_{\wedge}$) are covered at $N = 3, 4$ and in the full multi-dataset benchmark.
- The cohort sensitivity used a modest sample of edges per configuration: it is a robustness and saturation *demonstration*, not a precise re-estimate of the headline synergy levels (which are reported in Chapter 6 on the full samples).

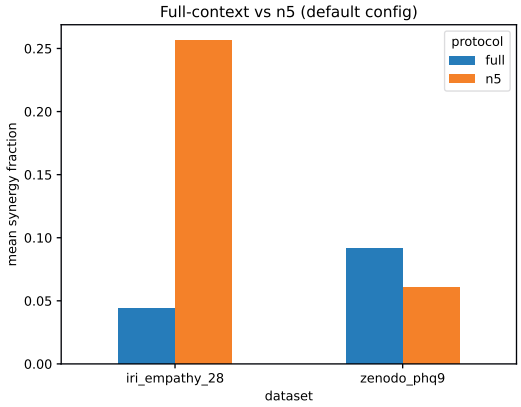


Figure E.21: Full-context versus $n_{\text{sources}} = 5$ synergy fraction. The 28-item IRI network saturates under full context (synergy collapses); the 9-item PHQ-9 does not.

Bibliography

[1] N. D. Jewson. The Disappearance of the Sick-Man from Medical Cosmology, 1770-1870. *Sociology*, 10(2):225–244, 5 1976. ISSN 0038-0385. doi: 10.1177/003803857601000202.

[2] A. Surís, R. Holliday, and C. North. The Evolution of the Classification of Psychiatric Disorders. *Behavioral Sciences*, 6(1):5, 2016.

[3] S. Heckers and K. S. Kendler. The evolution of Kraepelin's nosological principles. *World Psychiatry*, 19(3):381–388, 2020.

[4] A. T. Robinson. Pathology—The Beginnings of Laboratory Medicine. *Laboratory Medicine*, 52(2):e34–e35, 2021.

[5] M. Wieczorkowska. Biomedical model, reductionism and their consequences for body perception. *Acta Universitatis Lodziensis. Folia Sociologica*, (90):19–35, 2024. doi: 10.18778/0208-600x.90.02.

[6] M. Flores, G. Glusman, K. Brogaard, N. D. Price, and L. Hood. P4 medicine: how systems medicine will transform the healthcare sector and society. *Personalized Medicine*, 10(6):565–576, 2013.

[7] S. Naylor and J. Y. Chen. Unraveling human complexity and disease with systems biology and personalized medicine. *Personalized Medicine*, 7(3):275–289, 2010.

[8] K. Barnett, S. W. Mercer, M. Norbury, G. Watt, S. Wyke, and B. Guthrie. Epidemiology of multimorbidity and implications for health care, research, and medical education: A cross-sectional study. *The Lancet*, 380(9836):37–43, 7 2012. ISSN 1474547X. doi: 10.1016/S0140-6736(12)60240-2.

[9] A. Marengoni, S. Angleman, R. Melis, F. Mangialasche, A. Karp, A. Garmen, B. Meinow, and L. Fratiglioni. Aging with multimorbidity: A systematic review of the literature. *Ageing Research Reviews*, 10(4):430–439, 9 2011. ISSN 15681637. doi: 10.1016/j.arr.2011.03.003.

[10] Volksgezondheid en Zorg Info. Chronische aandoeningen en multimorbiditeit — Multimorbiditeit naar leeftijd en geslacht, 2024. URL https://www.vzinfo.nl/chronische-aandoeningen-en-multimorbiditeit/leeftijd-en-geslacht/multimorbiditeit. Data: Nivel Zorgregistraties Eerste Lijn, peildatum 1 januari 2024.

[11] M. G. M. O. Rikkert and G. M. E. E. G. Peeters. The multicausality of declines in wellbeing and cognition in older adults with multimorbidity. *The Lancet Healthy Longevity*, 2(11):e681–e682, 11 2021. ISSN 26667568. doi: 10.1016/S2666-7568(21)00254-3.

[12] L. E. Stirland, L. González-Saavedra, D. S. Mullin, C. W. Ritchie, G. Muniz-Terrera, and T. C. Russ. Measuring multimorbidity beyond counting diseases: systematic review of community and population studies and guide to index choice. *BMJ*, page m160, 2 2020. ISSN 1756-1833. doi: 10.1136/bmj.m160.

[13] R. J. Melis, S. M. Gijzel, and M. G. Olde Rikkert. Moving beyond multimorbidity as a simple count of diseases. *Journal of Evaluation in Clinical Practice*, 23(1):216–218, 2 2017. ISSN 13561294. doi: 10.1111/jep.12693.

[14] R. van Seben, L. A. Reichardt, D. R. Essink, B. C. van Munster, J. A. Bosch, and B. M. Buurman. "I Feel Worn Out, as if I Neglected Myself": Older Patients' Perspectives on Post-hospital Symptoms After Acute Hospitalization. *The Gerontologist*, 59(2):315–326, 3 2019. ISSN 0016-9013. doi: 10.1093/geront/gnx192.

[15] M. G. M. O. Rikkert, R. J. F. Melis, A. A. Cohen, G. M. E. E. . Geeske, and . Peeters. Why illness is more important than disease in old age. *Age and Ageing*, 51(1):1–6, 1 2022. ISSN 0002-0729. doi: 10.1093/AGEING/AFAB267.

[16] S. T. Skou, F. S. Mair, M. Fortin, B. Guthrie, B. P. Nunes, J. J. Miranda, C. M. Boyd, S. Pati, S. Mtenga, and S. M. Smith. Multimorbidity. *Nature Reviews Disease Primers*, 8(1):48, 7 2022. ISSN 2056-676X. doi: 10.1038/s41572-022-00376-4.

[17] C. M. Boyd, J. Darer, C. Boult, L. P. Fried, L. Boult, and A. W. Wu. Clinical Practice Guidelines and Quality of Care for Older Patients With Multiple Comorbid Diseases: Implications for Pay for Performance. *JAMA*, 294(6):716, 8 2005. ISSN 0098-7484. doi: 10.1001/jama.294.6.716.

[18] N. D. Shippee, N. D. Shah, C. R. May, F. S. Mair, and V. M. Montori. Cumulative complexity: a functional, patient-centered model of patient complexity can improve research and practice. *Journal of Clinical Epidemiology*, 65(10):1041–1051, 10 2012. ISSN 08954356. doi: 10.1016/j.jclinepi.2012.05.005.

[19] R. L. Maher, J. Hanlon, and E. R. Hajjar. Clinical consequences of polypharmacy in elderly. *Expert Opinion on Drug Safety*, 13(1):57–65, 1 2014. ISSN 1474-0338. doi: 10.1517/14740338.2013.827660.

[20] C. M. Boyd and D. M. Kent. Evidence-Based Medicine and the Hard Problem of Multimorbidity. *Journal of General Internal Medicine 2013 29:4*, 29(4):552–553, 1 2014. ISSN 1525-1497. doi: 10.1007/S11606-013-2658-Z.

[21] T. R. Fried. Health Outcome Prioritization as a Tool for Decision Making Among Older Persons With Multiple Chronic Conditions. *Archives of Internal Medicine*, 171(20):1856, 11 2011. ISSN 0003-9926. doi: 10.1001/archinternmed.2011.424.

[22] P. L. Williams and R. D. Beer. Nonnegative Decomposition of Multivariate Information. *arXiv preprint arXiv:1004.2515*, 4 2010.

[23] M. E. Tinetti, J. Esterson, R. Ferris, P. Posner, and C. S. Blaum. Patient Priority-Directed Decision Making and Care for Older Adults with Multiple Chronic Conditions. *Clinics in Geriatric Medicine*, 32(2):261–275, 2016. doi: 10.1016/j.cger.2016.01.012.

[24] A. S. Bierman, B. T. Burke, L. N. Comfort, M. Gerstein, N. M. Mueller, and C. A. Umscheid. From precision medicine to precision care: Choosing and using precision medicine in the context of multimorbidity. *Cambridge Prisms: Precision Medicine*, 1:e19, 2023. doi: 10.1017/pcm.2023.8.

[25] F. R. Steele. Personalized Medicine: Something Old, Something New. *Personalized Medicine*, 6(1):1–5, 1 2009. ISSN 1741-0541. doi: 10.2217/17410541.6.1.1.

[26] M. Huisman, J. Poppelaars, M. van der Horst, A. T. F. Beekman, J. Brug, T. G. van Tilburg, and D. J. H. Deeg. Cohort profile: the longitudinal aging study amsterdam. *Int J Epidemiol*, 40(4):868–76, 2011. doi: 10.1093/ije/dyq219.

[27] C. Sudlow, J. Gallacher, N. Allen, V. Beral, P. Burton, J. Danesh, P. Downey, P. Elliott, J. Green, M. Landray, B. Liu, P. Matthews, G. Ong, J. Pell, A. Silman, A. Young, T. Sprosen, T. Peakman, and R. Collins. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine*, 12(3):e1001779, 3 2015. ISSN 1549-1676. doi: 10.1371/JOURNAL.PMED.1001779.

[28] O. T. Raitakari, M. Juonala, T. Rönnemaa, L. Keltikangas-Järvinen, L. Räsänen, M. Pietikäinen, N. Hutri-Kähönen, L. Taittonen, E. Jokinen, J. Marniemi, A. Jula, R. Telama, M. Kähönen, T. Lehtimäki, H. K. Åkerblom, and J. S. Viikari. Cohort Profile: The Cardiovascular Risk in Young Finns Study. *International Journal of Epidemiology*, 37(6):1220–1226, 12 2008. ISSN 0300-5771. doi: 10.1093/IJE/DYM225.

[29] D. Borsboom and A. O. Cramer. Network Analysis: An Integrative Approach to the Structure of Psychopathology. *Annual Review of Clinical Psychology*, 9(1):91–121, 3 2013. ISSN 1548-5943. doi: 10.1146/annurev-clinpsy-050212-185608.

[30] S. Epskamp, D. Borsboom, and E. I. Fried. Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, 50(1):195–212, 2 2018. ISSN 15543528. doi: 10.3758/s13428-017-0862-1.

[31] M. Fralick, E. M. Macdonald, T. Gomes, T. Antoniou, S. Hollands, M. M. Mamdani, and D. N. Juurlink. Co-trimoxazole and sudden death in patients receiving inhibitors of renin-angiotensin system: population based study. *BMJ*, 349:g6196, 2014. doi: 10.1136/bmj.g6196.

[32] F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J. G. Young, and G. Petri. Networks beyond pairwise interactions: Structure and dynamics. *Physics Reports*, 874:1–92, 8 2020. ISSN 03701573. doi: 10.1016/j.physrep.2020.05.004.

[33] A. O. Cramer, L. J. Waldorp, H. L. Van Der Maas, and D. Borsboom. Comorbidity: A network perspective. *Behavioral and Brain Sciences*, 33(2-3):137–150, 2010. ISSN 1469-1825. doi: 10.1017/S0140525X09991567.

[34] A. L. Huntley, R. Johnson, S. Purdy, J. M. Valderas, and C. Salisbury. Measures of Multimorbidity and Morbidity Burden for Use in Primary Care and Community Settings: A Systematic Review and Guide. *The Annals of Family Medicine*, 10(2): 134–141, 3 2012. ISSN 1544-1709. doi: 10.1370/AFM.1363.

[35] K.-I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A.-L. Barabási. The human disease network. *Proceedings of the National Academy of Sciences*, 104(21): 8685–8690, 5 2007. ISSN 0027-8424. doi: 10.1073/pnas.0701361104.

[36] S. Epskamp and E. I. Fried. A Tutorial on Regularized Partial Correlation Networks. *Psychological Methods*, 23(4), 2018. ISSN 1082989X. doi: 10.1037/met0000167.

[37] C. Giusti, R. Ghrist, and D. S. Bassett. Two's company, three (or more) is a simplex: Algebraic-topological tools for understanding higher-order structure in neural data. *Journal of Computational Neuroscience*, 41(1):1–14, 8 2016. ISSN 0929-5313. doi: 10.1007/s10827-016-0608-6.

[38] C. E. Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 10 1948. ISSN 0005-8580. doi: 10.1002/j.1538-7305.1948.tb00917.x.

[39] A. Kraskov, H. Stögbauer, and P. Grassberger. Estimating mutual information. *Physical Review E*, 69(6):066138, 6 2004. ISSN 1539-3755. doi: 10.1103/PhysRevE.69.066138.

[40] J. T. Lizier, N. Bertschinger, J. Jost, and M. Wibral. Information decomposition of target effects from multi-source interactions: Perspectives on previous, current and futurework. *Entropy*, 20(4):307, 4 2018. ISSN 10994300. doi: 10.3390/E20040307.

[41] N. Timme, W. Alford, B. Flecker, and J. M. Beggs. Synergy, redundancy, and multivariate information measures: An experimentalist's perspective. *Journal of Computational Neuroscience*, 36(2):119–140, 4 2014. ISSN 15736873. doi: 10.1007/S10827-013-0458-4.

[42] F. Rosas, P. A. M. Mediano, M. Gastpar, and H. J. Jensen. Quantifying High-order Interdependencies via Multivariate Extensions of the Mutual Information. *Physical Review E*, 100(3):032305, 2 2019. ISSN 2470-0045. doi: 10.1103/PhysRevE.100.032305.

[43] R. A. A. Ince. The partial entropy decomposition: Decomposing multivariate entropy and mutual information via pointwise common surprisal, 2017.

[44] M. Harder, C. Salge, and D. Polani. Bivariate measure of redundant information. *Physical Review E*, 87(1):012130, 1 2013. ISSN 1539-3755. doi: 10.1103/PhysRevE.87.012130.

[45] V. Griffith and T. Ho. Quantifying Redundant Information in Predicting a Target Random Variable. *Entropy*, 17(7):4644–4653, 7 2015. ISSN 1099-4300. doi: 10.3390/e17074644.

[46] D. Chicharro and S. Panzeri. Synergy and Redundancy in Dual Decompositions of Mutual Information Gain and Information Loss. *Entropy*, 19(2):71, 2 2017. ISSN 1099-4300. doi: 10.3390/e19020071.

[47] P. A. M. Mediano, F. E. Rosas, A. I. Luppi, R. L. Carhart-Harris, D. Bor, A. K. Seth, and A. B. Barrett. Toward a unified taxonomy of information dynamics via Integrated Information Decomposition. *Proceedings of the National Academy of Sciences*, 122 (39), 9 2025. ISSN 0027-8424. doi: 10.1073/pnas.2423297122.

[48] C. Hourican, G. Peeters, R. J. Melis, S. L. Wezeman, T. M. Gill, M. G. Olde Rikkert, and R. Quax. Understanding multimorbidity requires sign-disease networks and higher-order interactions, a perspective. *Frontiers in Systems Biology*, 3:1155599, 6 2023. ISSN 2674-0702. doi: 10.3389/fsysb.2023.1155599.

[49] H. Nguyen, G. Manolova, C. Daskalopoulou, S. Vitoratou, M. Prince, and A. M. Prina. Prevalence of multimorbidity in community settings: A systematic review and meta-analysis of observational studies. *Journal of Comorbidity*, 9:2235042X1987093, 1 2019. ISSN 2235-042X. doi: 10.1177/2235042X19870934.

[50] M. Fortin, G. Bravo, C. Hudon, A. Vanasse, and L. Lapointe. Prevalence of Multimorbidity Among Adults Seen in Family Practice. *The Annals of Family Medicine*, 3 (3):223–228, 5 2005. ISSN 1544-1709. doi: 10.1370/AFM.272.

[51] C. Salisbury, L. Johnson, S. Purdy, J. M. Valderas, and A. A. Montgomery. Epidemiology and impact of multimorbidity in primary care: a retrospective cohort study. *British Journal of General Practice*, 61(582):e12–e21, 1 2011. ISSN 0960-1643. doi: 10.3399/BJGP11X548929.

[52] M. G. M. Olde Rikkert, A. S. Rigaud, R. J. Hoeyweghen, and J. Graaf. Geriatric syndromes: medical misnomer or progress in geriatrics? *The Netherlands journal of medicine*, 61(3):83–7, 3 2003. ISSN 0300-2977.

[53] A. L. Barabási, N. Gulbahce, and J. Loscalzo. Network medicine: a network-based approach to human disease. *Nature Reviews Genetics 2011 12:1*, 12(1):56–68, 12 2010. ISSN 1471-0064. doi: 10.1038/nrg2918.

[54] I. Jones, F. Cocker, M. Jose, M. Charleston, and A. L. Neil. Methods of analysing patterns of multimorbidity using network analysis: a scoping review. *Journal of Public Health (Germany)*, 2022. ISSN 16132238. doi: 10.1007/S10389-021-01685-W.

[55] N. Haug, J. Sorger, T. Gisinger, M. Gyimesi, A. Kautzky-Willer, S. Thurner, and P. Klimek. Decompression of multimorbidity along the disease trajectories of diabetes mellitus patients. *Frontiers in Physiology*, 11, 1 2021. ISSN 1664-042X. doi: 10.3389/fphys.2020.612604.

[56] J. Rafferty, A. Watkins, J. Lyons, R. A. Lyons, A. Akbari, N. Peek, F. Jalali-najafabadi, T. Ba Dhafari, A. Pate, G. P. Martin, and R. Bailey. Ranking sets of morbidities using hypergraph centrality. *Journal of Biomedical Informatics*, 122, 10 2021. ISSN 15320464. doi: 10.1016/J.JBI.2021.103916.

[57] T. Tripp-Reimer, J. K. Williams, S. E. Gardner, B. Rakel, K. Herr, A. M. McCarthy, L. L. Hand, S. Gilbertson-White, and C. Cherwin. An integrated model of multimorbidity and symptom science. *Nursing Outlook*, 68(4):430–439, 7 2020. ISSN 00296554. doi: 10.1016/j.outlook.2020.03.003.

[58] J. Liang and J. Han. Stochastic Boolean networks: An efficient approach to modeling gene regulatory networks. *BMC Systems Biology*, 6(1):113, 8 2012. ISSN 17520509. doi: 10.1186/1752-0509-6-113.

[59] T. Tsuchiya. Risk of performing multiple logistic regression analysis without considering multiplicity: an overview for clinicians and practitioners. *Journal of Medical Statistics and Informatics*, 2(1):7, 9 2014. doi: 10.7243/2053-7662-2-7.

[60] C. Marx, F. d. P. Calmon, and B. Ustun. Predictive Multiplicity in Classification. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, volume 119 of *PMLR*, pages 6765–6774, 2020.

[61] P. J. Lucas, L. C. Van Der Gaag, and A. Abu-Hanna. Bayesian networks in biomedicine and health-care. *Artificial Intelligence in Medicine*, 30(3):201–214, 3 2004. ISSN 0933-3657. doi: 10.1016/J.ARTMED.2003.11.001.

[62] M. Lappenschaar, A. Hommersom, P. J. Lucas, J. Lagro, S. Visscher, J. C. Korevaar, and F. G. Schellevis. Multilevel temporal bayesian networks can model longitudinal change in multimorbidity. *Journal of Clinical Epidemiology*, 66(12):1405–1416, 12 2013. ISSN 08954356. doi: 10.1016/j.jclinepi.2013.06.018.

[63] M. Lappenschaar, A. Hommersom, J. Lagro, and P. J. Lucas. Understanding the co-occurrence of diseases using structure learning. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7885 LNAI:135–144, 2013. ISSN 16113349. doi: 10.1007/978-3-642-38326-7_21.

[64] W. Song, H. Gong, Q. Wang, L. Zhang, L. Qiu, X. Hu, H. Han, Y. Li, R. Li, and Y. Li. Using bayesian networks with max-min hill-climbing algorithm to detect factors related to multimorbidity. *Frontiers in Cardiovascular Medicine*, 9, 8 2022. ISSN 2297-055X. doi: 10.3389/fcvm.2022.984883.

[65] T. G. Willadsen, A. Bebe, R. Køster-Rasmussen, D. E. Jarbøl, A. D. Guassora, F. B. Waldorff, S. Reventlow, and N. d. F. Olivarius. The role of diseases, risk factors and symptoms in the definition of multimorbidity – a systematic review. *Scandinavian Journal of Primary Health Care*, 34(2):112–121, 4 2016. ISSN 0281-3432. doi: 10.3109/02813432.2016.1153242.

[66] A. J. Yarnall, A. A. Sayer, A. Clegg, K. Rockwood, S. Parker, and J. Hindle. New horizons in multimorbidity in older adults. *Age and Ageing*, 46(6):882–888, 11 2017. ISSN 0002-0729. doi: 10.1093/ageing/afx150.

[67] L. E. Griffith, A. Gruneir, K. A. Fisher, K. Nicholson, D. Panjwani, C. Patterson, M. Markle-Reid, J. Ploeg, A. S. Bierman, D. B. Hogan, and R. Upshur. Key factors to consider when measuring multimorbidity: Results from an expert panel and online survey. *Journal of Comorbidity*, 8(1):2235042X1879530, 1 2018. ISSN 2235-042X. doi: 10.1177/2235042X18795306.

[68] J. A. L. Kempen, H. J. Schers, I. Philp, M. G. M. Olde Rikkert, and R. J. F. Melis. Predictive validity of a two-step tool to map frailty in primary care. *BMC Medicine*, 13(1):287, 12 2015. ISSN 1741-7015. doi: 10.1186/s12916-015-0519-9.

[69] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 2005. ISBN 9780471241959. doi: 10.1002/047174882X.

[70] J. B. Kinney and G. S. Atwal. Equitability, mutual information, and the maximal information coefficient. *Proceedings of the National Academy of Sciences of the United States of America*, 111(9):3354–3359, 3 2014. ISSN 00278424. doi: 10.1073/pnas.1309933111.

[71] R. G. James, N. Barnett, and J. P. Crutchfield. Information Flows? A Critique of Transfer Entropies. *Physical Review Letters*, 116(23):238701, 6 2016. ISSN 10797114. doi: 10.1103/PhysRevLett.116.238701.

[72] E. Olbrich, N. Bertschinger, and J. Rauh. Information Decomposition and Synergy. *Entropy 2015, Vol. 17, Pages 3501-3517*, 17(5):3501–3517, 5 2015. ISSN 1099-4300. doi: 10.3390/E17053501.

[73] C. Finn and J. T. Lizier. Pointwise Partial Information Decomposition Using the Specificity and Ambiguity Lattices. *Entropy 2018, Vol. 20, Page 297*, 20(4):297, 4 2018. ISSN 10994300. doi: 10.3390/E20040297.

[74] F. A. N. Santos, P. K. B. Tewarie, P. Baudot, A. Luchicchi, D. Barros De Souza, G. Girier, A. P. Milan, T. Broeders, E. G. Z. Centeno, R. Cofre, F. E. Rosas, D. Carone, J. Kennedy, C. J. Stam, A. Hillebrand, M. Desroches, S. Rodrigues, M. Schoonheim, L. Douw, and R. Quax. Emergence of High-Order Functional Hubs in the Human Brain. bioRxiv preprint, 2023. URL https://doi.org/10.1101/2023.02.10.528083.

[75] S. Stramaglia, T. Scagliarini, B. C. Daniels, and D. Marinazzo. Quantifying Dynamical High-Order Interdependencies From the O-Information: An Application to Neural Spiking Dynamics. *Frontiers in Physiology*, 11:1784, 1 2021. ISSN 1664042X. doi: 10.3389/fphys.2020.595736.

[76] R. J. Tallarida. Quantitative methods for assessing drug synergism. *Genes & cancer*, 2(11):1003–1008, 11 2011. ISSN 1947-6027. doi: 10.1177/1947601912440575.

[77] J. Wildenhain, M. Spitzer, S. Dolma, N. Jarvik, R. White, M. Roy, E. Griffiths, D. S. Bellows, G. D. Wright, and M. Tyers. Prediction of Synergism from Chemical-Genetic Interactions by Machine Learning. *Cell systems*, 1(6):383–395, 12 2015. ISSN 2405-4712. doi: 10.1016/J.CELS.2015.12.003.

[78] L. Rønneberg, A. Cremaschi, R. Hanes, J. M. Enserink, and M. Zucknick. bayesynergy: flexible Bayesian modelling of synergistic interaction effects in in vitro drug combination experiments. *Briefings in Bioinformatics*, 22(6):1–12, 11 2021. ISSN 14774054. doi: 10.1093/BIB/BBAB251.

[79] A. B. Owen. Sobol' Indices and Shapley Value. *https://doi.org/10.1137/130936233*, 2 (1):245–251, 6 2014. ISSN 21662525. doi: 10.1137/130936233.

[80] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra. Grad-CAM: Visual Explanations From Deep Networks via Gradient-Based Localization, 10 2017. URL http://gradcam.cloudcv.org.

[81] G. Montavon, A. Binder, S. Lapuschkin, W. Samek, and K. R. Müller. Layer-Wise Relevance Propagation: An Overview. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11700 LNCS:193–209, 2019. ISSN 16113349. doi: 10.1007/978-3-030-28954-6_10.

[82] A. Gevaert and Y. Saeys. PDD-SHAP: Fast Approximations for Shapley Values using Functional Decomposition. arXiv preprint arXiv:2208.12595, 2022.

[83] D. D. Lundstrom, T. Huang, and M. Razaviyayn. A Rigorous Study of Integrated Gradients Method and Extensions to Internal Neuron Attributions, 6 2022. ISSN 2640-3498. URL https://proceedings.mlr.press/v162/lundstrom22a.html.

[84] G. Briganti, M. Scutari, and R. J. McNally. A tutorial on bayesian networks for psychopathology researchers. *Psychological Methods*, 2022. ISSN 1082989X. doi: 10.1037/met0000479.

[85] D. Borsboom. A network theory of mental disorders. *World Psychiatry*, 16(1):5–13, 2 2017. ISSN 1723-8617. doi: 10.1002/wps.20375.

[86] J. G. Richens, C. M. Lee, and S. Johri. Improving the accuracy of medical diagnosis with causal machine learning. *Nature Communications*, 11(1), 12 2020. ISSN 20411723. doi: 10.1038/S41467-020-17419-7.

[87] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques - Daphne Koller, Nir Friedman - Google Books*. MIT press., 2009.

[88] R. Quax, O. Har-Shemesh, and P. M. Sloot. Quantifying Synergistic Information Using Intermediate Stochastic Variables. *Entropy 2017, Vol. 19, Page 85*, 19(2):85, 2 2017. ISSN 10994300. doi: 10.3390/E19020085.

[89] F. Eberhardt and R. Scheines. Interventions and Causal Inference. *Philosophy of Science*, 74(5):981–995, 12 2007. ISSN 0031-8248. doi: 10.1086/525638.

[90] G. Rose. Sick individuals and sick populations. *International Journal of Epidemiology*, 30:427–432, 11 2001. doi: 10.1093/oso/9780195180848.003.0003.

[91] J. A. Ferreira and A. H. Zwinderman. On the benjamini–hochberg method. *The Annals of Statistics*, 34(4), 8 2006. ISSN 0090-5364. doi: 10.1214/009053606000000425.

[92] L. K. King, L. March, and A. Anandacoomarasamy. Obesity & osteoarthritis. *Indian J Med Res*, 138:185–193, 2013.

[93] V. Bernhardt and T. G. Babb. Weight loss reduces dyspnea on exertion in obese women. *Respiratory Physiology & Neurobiology*, 204:86–92, 12 2014. ISSN 15699048. doi: 10.1016/j.resp.2014.09.004.

[94] V. Bernhardt, D. M. Bhammar, R. Marines-Price, and T. G. Babb. Weight loss reduces dyspnea on exertion and unpleasantness of dyspnea in obese men. *Respiratory Physiology & Neurobiology*, 261:55–61, 3 2019. ISSN 15699048. doi: 10.1016/j.resp.2019.01.007.

[95] A. Hayen, M. Herigstad, and K. T. Pattinson. Understanding dyspnea as a complex individual experience. *Maturitas*, 76(1):45–50, 9 2013. ISSN 03785122. doi: 10.1016/j.maturitas.2013.06.005.

[96] V. Griffith and C. Koch. Quantifying Synergistic Mutual Information. In M. Prokopenko, editor, *Guided Self-Organization: Inception*, volume 9 of *Emergence, Complexity and Computation*, pages 159–190. Springer, Berlin, Heidelberg, 2014. doi: 10.1007/978-3-642-53734-9_6.

[97] C. Finn and J. Lizier. Generalised measures of multivariate information content. *Entropy*, 22(2):216, 2 2020. ISSN 1099-4300. doi: 10.3390/e22020216.

[98] E. Fried. The network approach to psychopathology: pitfalls, challenges, and future directions, 6 2017. URL https://psych-networks.com/challenges-to-the-network-approach/.

[99] C. Hourican, J. Li, P. P. Mishra, T. Lehtimäki, B. H. Mishra, M. Kähönen, O. T. Raitakari, R. Laaksonen, L. Keltikangas-Järvinen, M. Juonala, and R. Quax. Efficient Search Algorithms for Identifying Synergistic Associations in High-Dimensional Datasets. *Entropy*, 26(11):968, 11 2024. ISSN 1099-4300. doi: 10.3390/e26110968.

[100] K. Clauw, S. Stramaglia, and D. Marinazzo. Higher-order mutual information reveals synergistic sub-networks for multi-neuron importance. arXiv preprint arXiv:2211.00416, 2022. URL https://arxiv.org/abs/2211.00416v2.

[101] L. Faes, G. Mijatovic, L. Sparacino, R. Pernice, Y. Antonacci, A. Porta, and S. Stramaglia. Quantifying High-Order Interactions in Cardiovascular and Cerebrovascular Networks. *2022 12th Conference of the European Study Group on Cardiovascular Oscillations, ESGCO 2022*, 2022. doi: 10.1109/ESGCO55423.2022.9931385.

[102] M. Gatica, R. Cofré, P. A. Mediano, F. E. Rosas, P. Orio, I. Diez, S. P. Swinnen, and J. M. Cortes. High-Order Interdependencies in the Aging Brain. *Brain Connectivity*, 11(9):734–744, 11 2021. ISSN 2158-0014. doi: 10.1089/brain.2020.0982.

[103] D. Marinazzo, J. Van Roozendaal, F. E. Rosas, M. Stella, R. Comolatti, N. Colenbier, S. Stramaglia, and Y. Rosseel. An information-theoretic approach to build hypergraphs in psychometrics. *Behavior Research Methods*, 56(7):8057–8079, 2024. doi: 10.3758/s13428-024-02471-8.

[104] P. Perrone and N. Ay. Hierarchical Quantification of Synergy in Channels. *Frontiers Robotics AI*, 2(JAN), 12 2015. ISSN 22969144. doi: 10.3389/frobt.2015.00035.

[105] N. Bertschinger, J. Rauh, E. Olbrich, J. Jost, and N. Ay. Quantifying Unique Information. *Entropy*, 16(4):2161–2183, 2014. doi: 10.3390/e16042161.

[106] A. Makkeh, D. Theis, and R. Vicente. Bivariate Partial Information Decomposition: The Optimization Perspective. *Entropy*, 19(10):530, 10 2017. ISSN 1099-4300. doi: 10.3390/e19100530.

[107] P. Wollstadt, J. Lizier, R. Vicente, C. Finn, M. Martinez-Zarzuela, P. Mediano, L. Novelli, and M. Wibral. IDTxl: The Information Dynamics Toolkit xl: a Python package for the efficient analysis of multivariate information dynamics in networks. *Journal of Open Source Software*, 4(34):1081, 2 2019. ISSN 2475-9066. doi: 10.21105/joss.01081.

[108] Y. Antonacci, L. Astolfi, G. Nollo, and L. Faes. Information Transfer in Linear Multivariate Processes Assessed through Penalized Regression Techniques: Validation and Application to Physiological Networks. *Entropy*, 22(7):732, 7 2020. ISSN 1099-4300. doi: 10.3390/e22070732.

[109] T. Scagliarini, D. Marinazzo, Y. Guo, S. Stramaglia, and F. E. Rosas. Quantifying high-order interdependencies on individual patterns via the local O-information: Theory and applications to music analysis. *Physical Review Research*, 4(1):013184, 3 2022. ISSN 2643-1564. doi: 10.1103/PhysRevResearch.4.013184.

[110] L. Faes, S. Member, G. Mijatovic, Y. Antonacci, R. Pernice, C. Barà, L. Sparacino, M. Sammartino, A. Porta, D. Marinazzo, and S. Stramaglia. A New Framework for the Time-and Frequency-Domain Assessment of High-Order Interactions in Networks of Random Processes. *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, 70:2022, 11 2022. ISSN 1053-587X. doi: 10.1109/TSP.2022.3221892.

[111] T. Scagliarini, D. Nuzzi, Y. Antonacci, L. Faes, F. E. Rosas, D. Marinazzo, and S. Stramaglia. Gradients of O-information: low-order descriptors of high-order dependencies. *Physical Review Research*, 5(1), 7 2022. ISSN 2643-1564. doi: 10.1103/physrevresearch.5.013025.

[112] F. E. Rosas, P. A. M Mediano, A. I. Luppi, T. F. Varley, J. T. Lizier, S. Stramaglia, H. J. Jensen, and D. Marinazzo. Disentangling high-order mechanisms and high-order behaviours in complex systems. *Nature PhysiCs |*, 18(5):476–477, 5 2022. ISSN 1745-2473. doi: 10.1038/s41567-022-01548-5.

[113] J. Gehlen, J. Li, C. Hourican, S. Tassi, P. P. Mishra, T. Lehtimäki, M. Kähönen, O. Raitakari, J. A. Bosch, and R. Quax. Bias in O-Information Estimation. *Entropy*, 26(10):837, 9 2024. ISSN 1099-4300. doi: 10.3390/e26100837.

[114] S. Sengupta, S. Basak, and R. A. Peters. Particle Swarm Optimization: A survey of historical and recent developments with hybridization perspectives. *Machine Learning and Knowledge Extraction*, 1(1):157–191, 2019. doi: 10.3390/make1010010.

[115] W.-N. Chen and D.-Z. Tan. Set-based discrete particle swarm optimization and its applications: a survey. *Frontiers of Computer Science*, 12(2):203–216, 4 2018. ISSN 2095-2228. doi: 10.1007/s11704-018-7155-4.

[116] X. Yu, W.-N. Chen, T. Gu, H. Zhang, H. Yuan, S. Kwong, and J. Zhang. Set-Based Discrete Particle Swarm Optimization Based on Decomposition for Permutation-Based Multiobjective Combinatorial Optimization Problems. *IEEE Transactions on Cybernetics*, 48(7):2139–2153, 7 2018. ISSN 2168-2267. doi: 10.1109/TCYB.2017.2728120.

[117] C. Qiu and F. Xiang. Feature selection using a set based discrete particle swarm optimization and a novel feature subset evaluation criterion. *Intelligent Data Analysis*, 23(1):5–21, 2 2019. ISSN 1088467X. doi: 10.3233/IDA-173735.

[118] T. F. Varley, M. Pope, J. Faskowitz, and O. Sporns. Multivariate information theory uncovers synergistic subsystems of the human cerebral cortex. *Communications Biology*, 6(1):451, 4 2023. ISSN 2399-3642. doi: 10.1038/s42003-023-04843-w.

[119] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(85):2825–2830, 10 2011.

[120] J. T. Melchior, S. E. Street, T. Vaisar, R. Hart, J. Jerome, Z. Kuklenyik, N. Clouet-Foraison, C. Thornock, S. Bedi, A. S. Shah, J. P. Segrest, J. W. Heinecke, and W. S. Davidson. Apolipoprotein A-I modulates HDL particle size in the absence of apolipoprotein A-II. *Journal of Lipid Research*, 62:100099, 2021. ISSN 00222275. doi: 10.1016/j.jlr.2021.100099.

[121] T. A. Musliner, H. J. Michenfelder, and R. M. Krauss. Interactions of high density lipoproteins with very low and low density lipoproteins during lipolysis. *Journal of lipid research*, 29(3):349–61, 3 1988. ISSN 0022-2275. doi: 10.1016/s0022-2275(20)38540-0.

[122] G. Peeters, C. Hourican, M. Lees, R. Quax, R. Melis, A. Kok, N. M. van Schoor, and M. Olde Rikkert. Symptoms provide predictive information on daily functioning beyond signs and diseases in middle-aged and older adults, particularly those with multimorbidity. *Age and Ageing*, 55(2):afag046, 3 2026. ISSN 1468-2834. doi: 10.1093/ageing/afag046.

[123] M. M. J. Nielen, M. J. J. C. Poos, and A. M. Gommer. Chronische aandoeningen en multimorbiditeit: Cijfers en context huidige situatie, 2018. URL https://www.volksgezondheidenzorg.info/onderwerp/chronische-ziekten-en-multimorbiditeit/cijfers-context/huidige-situatie#definities.

[124] T. T. Makovski, S. Schmitz, M. P. Zeegers, S. Stranges, and M. van den Akker. Multimorbidity and quality of life: Systematic literature review and meta-analysis. *Ageing Res Rev*, 53:100903, 2019. doi: 10.1016/j.arr.2019.04.005.

[125] A. Ryan, E. Wallace, P. O'Hara, and S. M. Smith. Multimorbidity and functional decline in community-dwelling adults: a systematic review. *Health Qual Life Outcomes*, 13:168, 2015. doi: 10.1186/s12955-015-0355-9.

[126] P. B. Tran, J. Kazibwe, G. F. Nikolaidis, I. Linnosmaa, M. Rijken, and J. van Olmen. Costs of multimorbidity: a systematic review and meta-analyses. *BMC Med*, 20(1):234, 2022. doi: 10.1186/s12916-022-02427-9.

[127] G. B. Hughes LD, McMurdo ME. Guidelines for people not for diseases: the challenges of applying uk clinical guidelines to people with multimorbidity. *Age Ageing*, 42(1):62–9, 2013. doi: 10.1093/ageing/afs100.

[128] R. Silva-Tinoco, L. Castillo-Martínez, A. Orea-Tejeda, J. J. Orozco-Gutiérrez, O. Vázquez-Díaz, P. Montaño-Hernández, A. Flores-Rebollar, and A. Reza-Albarrán. Developing thyroid disorders is associated with poor prognosis factors in patient with stable chronic heart failure. *Int J Cardiol*, 147(2):e24–5, 2011. doi: 10.1016/j.ijcard.2009.01.012.

[129] M. G. M. Olde Rikkert, R. W. Vingerhoets, N. van Geldorp, E. de Jong, and H. A. A. M. Maas. [atypical clinical picture of covid-19 in older patients]. *Ned Tijdschr Geneeskd*, 164, 2020.

[130] E. KR. Nonspecific and atypical presentation of disease in the older patient. *Geriatrics*, 53(2):50–2,, 1998.

[131] E. O. Hoogendijk, D. J. H. Deeg, S. de Breij, S. S. Klokgieters, A. A. L. Kok, N. Stringa, E. J. Timmermans, N. M. van Schoor, E. M. van Zutphen, M. van der Horst, J. Poppelaars, P. Malhoe, and M. Huisman. The longitudinal aging study amsterdam: cohort update 2019 and additional data collections. *European Journal of Epidemiology*, 35:61–74, 2020. doi: 10.1007/s10654-019-00541-2.

[132] G.-O. K. van Buuren S. mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 45(3):1, 2011. doi: 10.18637/jss.v045.i03.

[133] National Cancer Institute. NCI Dictionary of Cancer Terms. https://www.cancer.gov/publications/dictionaries/cancer-terms, 2024. Accessed 2024.

[134] D. DJ. Functional limitations, 2025. URL https://lasa-vu.nl/topics/functional-limitations/.

[135] A. B. Barrett. Exploration of synergistic and redundant information sharing in static and dynamical Gaussian systems. *Physical Review E*, 91(5):052802, 2015. doi: 10.1103/PhysRevE.91.052802.

[136] M. C. Rockwood K, Mitnitski AB. Some mathematical models of frailty and their clinical implications. *Reviews in Clinical Gerontology*, 12(2):109–117, 2002. doi: 10.1017/S0959259802012236.

[137] D. M. W. Kriegsman, D. J. H. Deeg, J. T. M. van Eijk, B. W. J. H. Penninx, and A. J. P. Boeke. Do disease specific characteristics add to the explanation of mobility limitations in patients with different chronic diseases? a study in the netherlands. *J Epidemiol Community Health*, 51(6):676–85, 1997. doi: 10.1136/jech.51.6.676.

[138] L. Sukumaran, A. Winston, J. Anderson, M. Boffito, F. A. Post, M. Sachikonye, P. W. G. Mallon, L. Waters, J. Vera, F. Burns, and C. A. Sabin. Multimorbidity frameworks impact the composition of patterns and their associations with patient-reported outcomes among people with hiv. *J Multimorb Comorb*, 15:26335565251331732, 2025. doi: 10.1177/26335565251331732.

[139] C. Hourican, G. Peeters, R. J. F. Melis, A. Kok, N. M. van Schoor, S. Wezeman, M. Lees, M. G. M. Olde Rikkert, and R. Quax. Interpreting higher-order dependence in multimorbidity using cohort data: A partial information decomposition approach. *PLOS Computational Biology*, 22(6):e1014386, 2026. doi: 10.1371/journal.pcbi.1014386.

[140] A. I. Luppi, P. A. M. Mediano, F. E. Rosas, D. J. Harrison, R. L. Carhart-Harris, D. Bor, and E. A. Stamatakis. What it is like to be a bit: an integrated information decomposition account of emergent mental phenomena. *Neuroscience of Consciousness*, 2021(2), 11 2021. ISSN 2057-2107. doi: 10.1093/nc/niab027.

[141] D. Kernick, C. A. Chew-Graham, and N. O'Flynn. Clinical assessment and management of multimorbidity: Nice guideline, 5 2017. ISSN 14785242.

[142] K. V. Patel, J. M. Guralnik, E. A. Phelan, N. M. Gell, R. B. Wallace, M. D. Sullivan, and D. C. Turk. Symptom Burden Among Community-Dwelling Older Adults in the United States. *Journal of the American Geriatrics Society*, 67:223–231, 2 2019. ISSN 15325415. doi: 10.1111/jgs.15673.

[143] H. E. Whitson, L. L. Sanders, C. F. Pieper, M. C. Morey, E. Z. Oddone, D. T. Gold, and H. J. Cohen. Correlation between symptoms and function in older adults with comorbidity. *Journal of the American Geriatrics Society*, 57:676–682, 4 2009. ISSN 00028614. doi: 10.1111/j.1532-5415.2009.02178.x.

[144] T. G. Willadsen, V. Siersma, D. R. Nicolaisdottir, D. Jarbol, A. D. Guassora, S. Reventlow, and R. Køster-Rasmussen. Symptom burden in multimorbidity: A population-based combined questionnaire and registry study from Denmark. *BMJ Open*, 11, 4 2021. ISSN 20446055. doi: 10.1136/bmjopen-2020-041877.

[145] E. O. Hoogendijk, N. M. van Schoor, Y. Qi, M. Visser, J. C. Swinkels, M. I. Broese van Groenou, A. A. L. Kok, T. J. Holwerda, H. R. W. Pasman, B. D. Onwuteaka-Philipsen, S. Remmelzwaal, E. van Ingen, T. G. van Tilburg, A.-C. van Haaster, M. van der Horst, J. Poppelaars, D. J. H. Deeg, and M. Huisman. The Longitudinal Aging Study Amsterdam: design and cohort update 2025. *European Journal of Epidemiology*, 40(6):705–720, 6 2025. ISSN 0393-2990. doi: 10.1007/s10654-025-01238-5.

[146] F. S. Luppino, L. M. de Wit, P. F. Bouvy, T. Stijnen, P. Cuijpers, B. W. J. H. Penninx, and F. G. Zitman. Overweight, Obesity, and Depression: A Systematic Review and Meta-analysis of Longitudinal Studies. *Archives of General Psychiatry*, 67(3):220, 3 2010. ISSN 0003-990X. doi: 10.1001/archgenpsychiatry.2010.2.

[147] A. Pan, Q. Sun, S. Czernichow, M. Kivimaki, O. I. Okereke, M. Lucas, J. E. Manson, A. Ascherio, and F. B. Hu. Bidirectional association between depression and obesity in middle-aged and older women. *International Journal of Obesity*, 36(4):595–602, 4 2012. ISSN 0307-0565. doi: 10.1038/ijo.2011.111.

[148] O. Pietiläinen, M. Laaksonen, O. Rahkonen, and E. Lahelma. Self-Rated Health as a Predictor of Disability Retirement – The Contribution of Ill-Health and Working Conditions. *PLoS ONE*, 6(9):e25004, 9 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0025004.

[149] J. Schnittker and V. Bacak. The Increasing Predictive Validity of Self-Rated Health. *PLoS ONE*, 9(1):e84933, 1 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0084933.

[150] G. Traversy and J.-P. Chaput. Alcohol Consumption and Obesity: An Update. *Current Obesity Reports*, 4(1):122–130, 3 2015. ISSN 2162-4968. doi: 10.1007/s13679-014-0129-4.

[151] J. L. Butler, P. Gordon-Larsen, L. M. Steffen, J. M. Shikany, D. R. Jacobs, B. M. Popkin, and J. M. Poti. Associations of 5-year changes in alcoholic beverage intake with 5-year changes in waist circumference and BMI in the Coronary Artery Risk Development in Young Adults (CARDIA) study. *PLOS ONE*, 18(3):e0281722, 3 2023. ISSN 1932-6203. doi: 10.1371/journal.pone.0281722.

[152] M. Ryu, H. Kimm, J. Jo, S. J. Lee, and S. H. Jee. Association between Alcohol Intake and Abdominal Obesity among the Korean Population. *Epidemiology and health*, 32:e2010007, 5 2010. ISSN 2092-7193. doi: 10.4178/epih/e2010007.

[153] L. de Koning, A. T. Merchant, J. Pogue, and S. S. Anand. Waist circumference and waist-to-hip ratio as predictors of cardiovascular events: meta-regression analysis of prospective studies. *European Heart Journal*, 28(7):850–856, 3 2007. ISSN 0195-668X. doi: 10.1093/eurheartj/ehm026.

[154] R. Xue, Q. Li, Y. Geng, H. Wang, F. Wang, and S. Zhang. Abdominal obesity and risk of CVD: a dose–response meta-analysis of thirty-one prospective studies. *British Journal of Nutrition*, 126(9):1420–1430, 11 2021. ISSN 0007-1145. doi: 10.1017/S0007114521000064.

[155] Y.-N. Ou, C.-C. Tan, X.-N. Shen, W. Xu, X.-H. Hou, Q. Dong, L. Tan, and J.-T. Yu. Blood Pressure and Risks of Cognitive Impairment and Dementia: A Systematic Review and Meta-Analysis of 209 Prospective Studies. *Hypertension*, 76(1):217–225, 7 2020. ISSN 0194-911X. doi: 10.1161/HYPERTENSIONAHA.120.14993.

[156] D. Hughes, C. Judge, R. Murphy, E. Loughlin, M. Costello, W. Whiteley, J. Bosch, M. J. O'Donnell, and M. Canavan. Association of Blood Pressure Lowering With Incident Dementia or Cognitive Impairment: A Systematic Review and Meta-analysis. *JAMA*, 323(19):1934, 5 2020. ISSN 0098-7484. doi: 10.1001/jama.2020.4249.

[157] M. Liu, T. Hou, M. Nkimbeng, Y. Li, J. L. Taylor, X. Sun, S. Tang, and S. L. Szanton. Associations between symptoms of pain, insomnia and depression, and frailty in older adults: A cross-sectional analysis of a cohort study. *International Journal*

of Nursing Studies, 117:103873, 5 2021. ISSN 00207489. doi: 10.1016/j.ijnurstu.2021.103873.

[158] K. K. Brewster, M.-C. Hu, S. Zilcha-Mano, A. Stein, P. J. Brown, M. M. Wall, S. P. Roose, J. S. Golub, and B. R. Rutherford. Age-Related Hearing Loss, Late-Life Depression, and Risk for Incident Dementia in Older Adults. *The Journals of Gerontology: Series A*, 76(5):827–834, 4 2021. ISSN 1079-5006. doi: 10.1093/gerona/glaa242.

[159] D. Poelarends, S. E. Kramer, C. Smits, and P. Merkus. The prevalence of patient-reported cognitive complaints and dementia risk factors in the audiology clinic. *International Journal of Audiology*, 64(8):832–839, 8 2025. ISSN 1499-2027. doi: 10.1080/14992027.2024.2406882.

[160] J. Huang, X. Wang, H. Zhu, D. Huang, W. Li, J. Wang, and Z. Liu. Association between grip strength and cognitive impairment in older American adults. *Frontiers in Molecular Neuroscience*, 15, 11 2022. ISSN 1662-5099. doi: 10.3389/fnmol.2022.973700.

[161] M. Gray, J. M. Glenn, and A. Binns. Predicting sarcopenia from functional measures among community-dwelling older adults. *AGE*, 38(1):22, 2 2016. ISSN 0161-9152. doi: 10.1007/s11357-016-9887-0.

[162] M. Curtis, L. Swan, R. Fox, A. Warters, and M. O'Sullivan. Associations between Body Mass Index and Probable Sarcopenia in Community-Dwelling Older Adults. *Nutrients*, 15(6):1505, 3 2023. ISSN 2072-6643. doi: 10.3390/nu15061505.

[163] J. W. Kay and R. A. A. Ince. Exact partial information decompositions for Gaussian systems based on dependency constraints. *Entropy*, 20(4):240, 2018. doi: 10.3390/e20040240.

[164] D. J. Kleitman. On dedekind's problem: The number of monotone boolean functions. *Proceedings of the American Mathematical Society*, 21(3):677–682, 3 1969. ISSN 0002-9939. doi: 10.1090/s0002-9939-1969-0241334-6.

[165] A. Makkeh, D. O. Theis, and R. Vicente. BROJA-2PID: A Robust Estimator for Bivariate Partial Information Decomposition. *Entropy*, 20(4):271, 2018. ISSN 1099-4300. doi: 10.3390/e20040271.

[166] R. G. James, C. J. Ellison, and J. P. Crutchfield. dit: a Python package for discrete information theory. *Journal of Open Source Software*, 3(25):738, 2018. doi: 10.21105/joss.00738.

[167] L. Koçillari, G. M. Lorenz, N. M. Engel, M. Celotto, S. Curreli, S. Blanco Malerba, A. K. Engel, T. Fellin, and S. Panzeri. Sampling bias corrections for accurate neural measures of redundant, unique, and synergistic information. *bioRxiv*, 6 2024. doi: 10.1101/2024.06.04.597303.

[168] C. B. Weir and A. Jan. BMI Classification Percentile And Cut Off Points. In *StatPearls*. StatPearls Publishing, 2019.

[169] A. J. Cruz-Jentoft, G. Bahat, J. Bauer, Y. Boirie, O. Bruyère, T. Cederholm, C. Cooper, F. Landi, Y. Rolland, A. A. Sayer, S. M. Schneider, C. C. Sieber, E. Topinkova, M. Vandewoude, M. Visser, M. Zamboni, I. Bautmans, J. P. Baeyens, M. Cesari, A. Cherubini, J. Kanis, M. Maggio, F. Martin, J. P. Michel, K. Pitkala, J. Y. Reginster, R. Rizzoli, D. Sánchez-Rodríguez, and J. Schols. Sarcopenia: Revised European consensus on definition and diagnosis, 1 2019. ISSN 14682834.

[170] S. M. Hunt, S. P. McKenna, J. McEwen, J. Williams, and E. Papp. The Nottingham health profile: Subjective health status and medical consultations. *Social Science and Medicine. Part A Medical Psychology and Medical*, 15:221–229, 1981. ISSN 02717123. doi: 10.1016/0271-7123(81)90005-5.

[171] L. S. Radloff. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement*, 1(3):385–401, 6 1977. ISSN 0146-6216. doi: 10.1177/014662167700100306.

[172] A. S. Zigmond and R. P. Snaith. The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica*, 67(6):361–370, 6 1983. ISSN 0001-690X. doi: 10.1111/j.1600-0447.1983.tb09716.x.

[173] S. Guloksuz, L. Pries, and J. Van Os. Application of network methods for understanding mental disorders: pitfalls and promise. *Psychological medicine*, 47(16): 2743–2752, 12 2017. ISSN 0033-2917. doi: 10.1017/s0033291717001350.

[174] M. Malgaroli, A. Calderon, and G. A. Bonanno. Networks of major depressive disorder: A systematic review. *Clinical Psychology Review*, 85:102000, 2 2021. doi: 10.31234/osf.io/wb53a.

[175] A. O. Cramer, C. D. Van Borkulo, E. J. Giltay, H. L. Van Der Maas, K. S. Kendler, M. Scheffer, and D. Borsboom. Major Depression as a Complex Dynamic System. *PLOS ONE*, 11(12):e0167490, 12 2016. ISSN 1932-6203. doi: 10.1371/JOURNAL.PONE.0167490.

[176] M. Horton and A. E. Perry. Screening for depression in primary care: A Rasch analysis of the PHQ-9. *BJPsych Bulletin*, 40(5):237–243, 2016. ISSN 2056-4694. doi: 10.1192/pb.bp.114.050195.

[177] E. I. Fried and A. O. J. Cramer. Moving Forward: Challenges and Directions for Psychopathological Network Theory and Methodology. *Perspectives on Psychological Science*, 12(6):999–1020, 11 2017. ISSN 1745-6916. doi: 10.1177/1745691617705892.

[178] J. C. Franklin, J. D. Ribeiro, K. R. Fox, K. H. Bentley, E. M. Kleiman, X. Huang, K. M. Musacchio, A. C. Jaroszewski, B. P. Chang, and M. K. Nock. Risk factors for suicidal

thoughts and behaviors: A meta-analysis of 50 years of research. *Psychological Bulletin*, 143(2):187–232, 2017. doi: 10.1037/bul0000084.

[179] J. M. B. Haslbeck, D. Borsboom, and L. J. Waldorp. Moderated network models. *Multivariate Behavioral Research*, 56(2):256–287, 2021. doi: 10.1080/00273171.2019.1677207.

[180] K. A. S. Davis, J. R. I. Coleman, M. Adams, et al. Mental health in uk biobank: development, implementation and results from an online questionnaire completed by 157,366 participants: a reanalysis. *BJPsych Open*, 2020. doi: 10.1192/bjo.2019.100.

[181] Z. Su, R. Liu, Y. Wei, R. Zhang, X. Xu, Y. Wang, Y. Zhu, L. Wang, L. Liang, F. Wang, and X. Zhang. Temporal dynamics in psychological assessments: A novel dataset with scales and response times. *Scientific Data*, 11(1):1046, 2024. doi: 10.1038/s41597-024-03888-8.

[182] Z. Su, R. Liu, Y. Wei, R. Zhang, X. Xu, Y. Wang, Y. Zhu, L. Wang, L. Liang, F. Wang, and X. Zhang. Data from: Temporal dynamics in psychological assessments: a novel dataset with scales and response times. Zenodo, 2024. URL https://zenodo.org/records/10423537. Dataset accompanying Su et al. (2024), *Scientific Data*, 11, 1046.

[183] M. H. Davis. Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44(1):113–126, 1983. doi: 10.1037/0022-3514.44.1.113.

[184] K. Kroenke, R. L. Spitzer, and J. B. W. Williams. The phq-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9):606–613, 2001. doi: 10.1046/j.1525-1497.2001.016009606.x.

[185] R. L. Spitzer, K. Kroenke, J. B. W. Williams, and the Patient Health Questionnaire Primary Care Study Group. Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. *JAMA*, 282(18):1737–1744, 1999. doi: 10.1001/jama.282.18.1737.

[186] T. Scagliarini et al. Gradients of o-information: Low-order descriptors of high-order dependencies. *Physical Review Research*, 5:013025, 2023. doi: 10.1103/PhysRevResearch.5.013025.

[187] A. Kolchinsky. Partial information decomposition: Redundancy as information bottleneck. *Entropy*, 26:546, 2024. doi: 10.3390/e26070546.

[188] T. F. Varley. A scalable synergy-first backbone decomposition of higher-order structures in complex systems. *npj Complexity*, 2024. doi: 10.1038/s44260-024-00011-1.

[189] A. E. Goodwell and P. Kumar. Temporal information partitioning: Characterizing synergy, uniqueness, and redundancy in interacting environmental variables. *Water Resources Research*, 53(7):5920–5942, 2017. doi: 10.1002/2016WR020216.

[190] P. Gács and J. Körner. Common information is far less than mutual information. *Problems of Control and Information Theory*, 2(2):149–162, 1973.

[191] A. Jansma, P. A. M. Mediano, and F. E. Rosas. Fast Möbius transform: An algebraic approach to information decomposition. *Physical Review Research*, 7(3):033049, 2025. doi: 10.1103/32sq-z2jf.

[192] D. Gondek and T. Hofmann. Conditional information bottleneck clustering. In *Proceedings of the 3rd IEEE International Conference on Data Mining, Workshop on Clustering Large Data Sets*, pages 36–42, 2003.

[193] N. Slonim and N. Tishby. Agglomerative information bottleneck. In *Advances in Neural Information Processing Systems*, volume 12, pages 617–623, 1999.

[194] M. Greenacre. *Correspondence Analysis in Practice*. Chapman and Hall/CRC, 3rd edition, 2017. doi: 10.1201/9781315369983.

[195] M. Halford. prince: Multivariate exploratory data analysis in Python, 2023. URL https://github.com/MaxHalford/prince. Python package.

[196] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999. doi: 10.1038/44565.

[197] K.-C. Li. Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, 86(414):316–327, 1991. doi: 10.1080/01621459.1991.10475035.

[198] J. Koepke. sliced: Sliced inverse regression in Python, 2018. URL https://github.com/joshloyal/sliced. Python package.

[199] S. Wold, A. Ruhe, H. Wold, and W. J. Dunn, III. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM Journal on Scientific and Statistical Computing*, 5(3):735–743, 1984. doi: 10.1137/0905052.

[200] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, 4th edition, 2013.

[201] F. Fleuret. Fast binary feature selection with conditional mutual information. *Journal of Machine Learning Research*, 5:1531–1555, 2004.

[202] H. H. Yang and J. Moody. Data visualization and feature selection: New algorithms for nongaussian data. *Advances in Neural Information Processing Systems*, 12, 1999.

[203] G. Brown, A. Pocock, M.-J. Zhao, and M. Luján. Conditional likelihood maximisation: A unifying framework for information theoretic feature selection. *Journal of Machine Learning Research*, 13:27–66, 2012.

[204] I. Kononenko. Estimating attributes: Analysis and extensions of RELIEF. *Machine Learning: ECML-94*, pages 171–182, 1994.

[205] R. J. Urbanowicz, M. Meeker, W. La Cava, R. S. Olson, and J. H. Moore. Benchmarking relief-based feature selection methods for bioinformatics data mining. *Journal of Biomedical Informatics*, 85:168–188, 2018. doi: 10.1016/j.jbi.2018.07.015.

[206] F. Murtagh and P. Legendre. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *Journal of Classification*, 31(3): 274–295, 2014.

[207] Z. Huang. Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge Discovery*, 2(3):283–304, 1998. doi: 10.1023/A:1009769707641.

[208] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, volume 14, 2001.

[209] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300, 1995. doi: 10.1111/j.2517-6161.1995.tb02031.x.

[210] E. J. Lawrence, P. Shaw, D. Baker, S. Baron-Cohen, and A. S. David. Measuring empathy: Reliability and validity of the Empathy Quotient. *Psychological Medicine*, 34(5):911–920, 2004. ISSN 0033-2917. doi: 10.1017/S0033291703001624.

[211] J. Decety and C. Holvoet. The emergence of empathy: A developmental neuroscience perspective. *Developmental Review*, 62:100999, 2021. ISSN 0273-2297. doi: 10.1016/j.dr.2021.100999.

[212] T. F. Varley, P. A. M. Mediano, A. Patania, and J. Bongard. The topology of synergy: Linking topological and information-theoretic approaches to higher-order interactions in complex systems. *PLOS Computational Biology*, 21(11):e1013649, 2025. doi: 10.1371/journal.pcbi.1013649.

[213] N. M. Timme and C. Lapish. A tutorial for information theory in neuroscience. *eNeuro*, 5(3):ENEURO.0052–18.2018, 2018. doi: 10.1523/ENEURO.0052-18.2018.

[214] E. I. Fried, S. Epskamp, R. M. Nesse, F. Tuerlinckx, and D. Borsboom. What are 'good' depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis. *Journal of Affective Disorders*, 189: 314–320, 2016. doi: 10.1016/j.jad.2015.09.005.

[215] M. K. Forbes, A. G. C. Wright, K. E. Markon, and R. F. Krueger. Evidence that psychopathology symptom networks have limited replicability. *Journal of Abnormal Psychology*, 126(7):969–988, 2017. doi: 10.1037/abn0000276.

[216] M.-A. Schulz and K. Ritter. Measurement noise limits the advantage of nonlinear models over linear models in biomedical prediction, 2026.

[217] L. Paninski. Estimation of entropy and mutual information. *Neural Computation*, 15(6):1191–1253, 6 2003. ISSN 0899-7667. doi: 10.1162/089976603321780272.

[218] P. Venkatesh and G. Schamberg. Partial Information Decomposition via Deficiency for Multivariate Gaussians. *2022 IEEE International Symposium on Information Theory (ISIT)*, pages 2892–2897, 6 2022. doi: 10.1109/isit50566.2022.9834649.

[219] A. I. Luppi, P. A. Mediano, F. E. Rosas, N. Holland, T. D. Fryer, J. T. O'Brien, J. B. Rowe, D. K. Menon, D. Bor, and E. A. Stamatakis. A synergistic core for human brain evolution and cognition. *Nature Neuroscience*, 25:771–782, 6 2022. ISSN 15461726. doi: 10.1038/s41593-022-01070-0.

[220] L. F. Bringmann, T. Elmer, S. Epskamp, R. W. Krause, D. Schoch, M. Wichers, J. Wigman, and E. Snippe. What do centrality measures measure in psychological networks. *Journal of Abnormal Psychology*, 2019. doi: 10.1037/abn0000446.

[221] E. I. Fried, C. D. van Borkulo, A. O. Cramer, L. Boschloo, R. A. Schoevers, and D. Borsboom. Mental disorders as networks of problems: a review of recent insights. *Social Psychiatry and Psychiatric Epidemiology*, 52(1):1–10, 1 2017. ISSN 09337954. doi: 10.1007/s00127-016-1319-z.

[222] M. N. Hallquist, A. G. C. Wright, and P. C. M. Molenaar. Problems with centrality measures in psychopathology symptom networks: Why network psychometrics cannot escape psychometric theory. *Multivariate Behavioral Research*, 56(2):199–223, 2021. doi: 10.1080/00273171.2019.1640103.

[223] Z. P. Neal, M. K. Forbes, J. W. Neal, M. J. Brusco, R. F. Krueger, K. Markon, D. Steinley, S. Wasserman, and A. G. C. Wright. Critiques of network analysis of multivariate data in psychological science. *Nature Reviews Methods Primers*, 2(1):90, 2022. doi: 10.1038/s43586-022-00177-9.

[224] D. J. Robinaugh, R. H. Hoekstra, E. R. Toner, and D. Borsboom. The network approach to psychopathology: a review of the literature 2008–2018 and an agenda for future research. *Psychological Medicine*, 50(3):353–366, 2 2020. ISSN 0033-2917. doi: 10.1017/S0033291719003404.

[225] M. Y. Wei, C.-H. Tseng, and A. J. Kang. Higher-Order Disease Interactions in Multimorbidity Measurement: Marginal Benefit Over Additive Disease Summation. *The Journals of Gerontology, Series A: Biological Sciences and Medical Sciences*, 80(1), 12 2024. ISSN 1079-5006. doi: 10.1093/gerona/glae282.

[226] A. Prados-Torres, A. Calderón-Larrañaga, J. Hancco-Saavedra, B. Poblador-Plou, and M. van den Akker. Multimorbidity patterns: a systematic review. *Journal of Clinical Epidemiology*, 67(3):254–266, 2014. doi: 10.1016/j.jclinepi.2013.09.021.

[227] C. Franceschi, P. Garagnani, P. Parini, C. Giuliani, and A. Santoro. Inflammaging: a new immune-metabolic viewpoint for age-related diseases. *Nature Reviews Endocrinology*, 14(10):576–590, 2018. doi: 10.1038/s41574-018-0059-4.

[228] L. Ferrucci and E. Fabbri. Inflammageing: chronic inflammation in ageing, cardiovascular disease, and frailty. *Nature Reviews Cardiology*, 15(9):505–522, 2018. doi: 10.1038/s41569-018-0064-2.

[229] R. D. Gibbons, D. J. Weiss, E. Frank, and D. Kupfer. Computerized Adaptive Diagnosis and Testing of Mental Health Disorders. *Annual Review of Clinical Psychology*, 12(1):83–104, 3 2016. ISSN 1548-5943. doi: 10.1146/annurev-clinpsy-021815-093634.

[230] A. A. L. Kok, M. Huisman, E. J. Giltay, and G. Lunansky. Adopting a complex systems approach to functional ageing: bridging the gap between gerontological theory and empirical research. *The Lancet Healthy Longevity*, 6(3):100673, 2025. doi: 10.1016/j.lanhl.2024.100673.

[231] A. Fomins, Y. Sych, and F. Helmchen. Conservative significance testing of tripartite statistical relations in multivariate neural data. *Network Neuroscience*, 6(4):1243–1274, 2022. doi: 10.1162/netn_a_00259.

[232] A. Marx, A. Gretton, and J. M. Mooij. A Weaker Faithfulness Assumption based on Triple Interactions. In *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence (UAI)*, volume 161 of *PMLR*, pages 451–460, 2021.

[233] P. A. M. Mediano, F. E. Rosas, A. I. Luppi, H. J. Jensen, A. K. Seth, A. B. Barrett, R. L. Carhart-Harris, and D. Bor. Greater than the parts: a review of the information decomposition approach to causal emergence. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 380(2227), 7 2022. ISSN 1364-503X. doi: 10.1098/rsta.2021.0246.

[234] L. Faes, L. Sparacino, G. Mijatovic, Y. Antonacci, L. Ricci, D. Marinazzo, and S. Stramaglia. Partial Information Rate Decomposition. *Physical Review Letters*, 135(18), 10 2025. ISSN 0031-9007. doi: 10.1103/nrwj-n8lj.

[235] D. L. Vetrano, A. Roso-Llorach, S. Fernández, A. Calderón-Larrañaga, G. Onder, L. Fratiglioni, A. Marengoni, and A. Prados-Torres. Twelve-year clinical trajectories of multimorbidity in a population of older adults. *Nature Communications*, 11(1):3223, 2020. doi: 10.1038/s41467-020-16780-x.

[236] D. C. Mohr, M. Zhang, and S. M. Schueller. Personal sensing: Understanding mental health using ubiquitous sensors and machine learning. *Annual Review of Clinical Psychology*, 13:23–47, 2017. doi: 10.1146/annurev-clinpsy-032816-044949.

[237] J. Torous, J.-P. Onnela, and M. Keshavan. New dimensions and new tools to realize the potential of RDoC: digital phenotyping via smartphones and connected devices. *Translational Psychiatry*, 7:e1053, 2017. doi: 10.1038/tp.2017.25.

[238] I. Nahum-Shani, S. N. Smith, B. J. Spring, L. M. Collins, K. Witkiewitz, A. Tewari, and S. A. Murphy. Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. *Annals of Behavioral Medicine*, 52(6):446–462, 2018. doi: 10.1007/s12160-016-9830-8.

[239] D. McNeish. Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23(3):412–433, 2018. doi: 10.1037/met0000144.

[240] E. I. Fried. The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of Affective Disorders*, 208:191–197, 2017. doi: 10.1016/j.jad.2016.10.019.

[241] T. Yan and A. D. Procaccia. If You Like Shapley Then You'll Love the Core. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI-21)*, pages 5751–5759, 2021. doi: 10.1609/aaai.v35i6.16721.

[242] E. M. Andresen, J. A. Malmgren, W. B. Carter, and D. L. Patrick. Screening for depression in well older adults: evaluation of a short form of the ces-d (center for epidemiologic studies depression scale). *American Journal of Preventive Medicine*, 10:77–84, 1994.

[243] S. M. Hunt, J. McEwen, and S. P. McKenna. Measuring health status: a new tool for clinicians and epidemiologists. *Journal of the Royal College of General Practitioners*, 35:185–188, 1985.

[244] J. W. McEvoy, C. P. McCarthy, R. M. Bruno, et al. 2024 esc guidelines for the management of elevated blood pressure and hypertension. *European Heart Journal*, 45: 3912–4018, 2024. doi: 10.1093/eurheartj/ehae178.

[245] World Health Organization. Obesity: preventing and managing the global epidemic. report of a who consultation. Technical Report 894, World Health Organization, 2000.

[246] World Health Organization. Waist circumference and waist-hip ratio: report of a who expert consultation, geneva, 8–11 december 2008. Technical report, World Health Organization, Geneva, 2011.

[247] M. F. Folstein, S. E. Folstein, and P. R. McHugh. “mini-mental state”. a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12:189–198, 1975.

[248] Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey (NHANES), 2017–2018. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2020. URL https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2017. Accessed: 2026-03-19.

[249] A. Börsch-Supan. Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 7. SHARE-ERIC, 2022. Release version: 8.0.0.

[250] J. Banks, G. D. Batty, J. Breedvelt, K. Coughlin, R. Crawford, M. Marmot, J. Nazroo, Z. Oldfield, N. Steel, A. Steptoe, M. Wood, and P. Zaninotto. English Longitudinal Study of Ageing: Waves 0–9, 1998–2019. UK Data Service, 2023. Wave 6 used in this study.

[251] Health and Retirement Study. Harmonized HRS version C. University of Michigan, Institute for Social Research, 2024. URL https://hrs.isr.umich.edu/. Wave 13 (2016) used in this study. Harmonized HRS Version C. DOI 10.7910/DVN/8KXWPK could not be verified – using institutional URL instead.

[252] D. Cella, W. Riley, A. Stone, N. Rothrock, B. Reeve, S. Yount, D. Amtmann, R. Bode, D. Buysse, S. Choi, K. Cook, R. DeVellis, D. DeWalt, J. F. Fries, R. Gershon, E. A. Hahn, J.-S. Lai, P. Pilkonis, D. Revicki, M. Rose, K. Weinfurt, and R. Hays. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *Journal of Clinical Epidemiology*, 63(11):1179–1194, 2010. doi: 10.1016/j.jclinepi.2010.04.011.

[253] Centers for Disease Control and Prevention (CDC). Behavioral Risk Factor Surveillance System Survey Data, 2022. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2023. URL https://www.cdc.gov/brfss/annual_data/annual_2022.html. Accessed: 2026-03-19.

[254] Agency for Healthcare Research and Quality (AHRQ). Medical Expenditure Panel Survey (MEPS), 2022 Full Year Consolidated Data File (HC-243). U.S. Department of Health and Human Services, Agency for Healthcare Research and Quality, 2024. URL https://meps.ahrq.gov/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-243. Accessed: 2026-03-19.

[255] U.S. Census Bureau. Household Pulse Survey: Measuring Social and Economic Impacts During the Coronavirus Pandemic, Phase 4.2, Cycle 4 (December 2024). U.S. Census Bureau, 2025. URL https://www.census.gov/programs-surveys/household-pulse-survey.html. Accessed: 2026-03-19.

[256] G. Briganti, E. I. Fried, and P. Linkowski. Network analysis of Contingencies of Self-Worth Scale in 680 university students. *Psychiatry Research*, 272:252–257, 2019. doi: 10.1016/j.psychres.2018.12.080.

[257] A. Börsch-Supan. easySHARE. SHARE-ERIC, 2022. Release version: 8.0.0. Waves 1, 2, 4, 5, 6, 7, 8 used in this study.

[258] S. Gruber, C. Hunkler, and S. Stuck. Generating easySHARE: Guidelines, structure, content and programming. SHARE Working Paper Series 17-2014, Munich Center for the Economics of Aging (MEA), Max Planck Institute for Social Law and Social Policy, 2014.

[259] P. F. Lovibond and S. H. Lovibond. The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, 33(3):335–343, 1995. doi: 10.1016/0005-7967(94)00075-U.

[260] J. A. Taylor. A personality scale of manifest anxiety. *The Journal of Abnormal and Social Psychology*, 48(2):285–290, 1953. doi: 10.1037/h0056264.

[261] L. R. Goldberg. A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. J. Deary, F. De Fruyt, and F. Ostendorf, editors, *Personality Psychology in Europe*, volume 7, pages 7–28. Tilburg University Press, Tilburg, The Netherlands, 1999. Data obtained from the Open-Source Psychometrics Project, https://openpsychometrics.org/_rawdata/.

[262] M. Rosenberg. *Society and the Adolescent Self-Image*. Princeton University Press, Princeton, NJ, 1965. Data obtained from the Open-Source Psychometrics Project, https://openpsychometrics.org/_rawdata/.

[263] K. A. Brennan, C. L. Clark, and P. R. Shaver. Self-report measurement of adult attachment: An integrative overview. In J. A. Simpson and W. S. Rholes, editors, *Attachment Theory and Close Relationships*, pages 46–76. Guilford Press, New York, 1998. Data obtained from the Open-Source Psychometrics Project, https://openpsychometrics.org/_rawdata/.

[264] H. M. Hendin and J. M. Cheek. Assessing hypersensitive narcissism: A reexamination of Murray's Narcism scale. *Journal of Research in Personality*, 31(4):588–599, 1997. doi: 10.1006/jrpe.1997.2204.

[265] P. K. Jonason and G. D. Webster. The dirty dozen: A concise measure of the dark triad. *Psychological Assessment*, 22(2):420–432, 2010. doi: 10.1037/a0019265.

[266] R. A. Baer, G. T. Smith, and K. B. Allen. Assessment of mindfulness by self-report: The Kentucky Inventory of Mindfulness Skills. *Assessment*, 11(3):191–206, 2004. doi: 10.1177/1073191104268029.

[267] S. C. Kalichman and D. Rompa. Sexual sensation seeking and sexual compulsivity scales: Validity, and predicting HIV risk behavior. *Journal of Personality Assessment*, 65(3):586–601, 1995. doi: 10.1207/s15327752jpa6503_16.

[268] H. A. Guvenir and G. Demiroz. Dermatology. UCI Machine Learning Repository, 1998. URL https://archive.ics.uci.edu/dataset/33/dermatology.

[269] H. A. Guvenir, G. Demiroz, and N. Ilter. Learning differential diagnosis of erythemato-squamous diseases using voting feature intervals. *Artificial Intelligence in Medicine*, 13(3):147–165, 1998. doi: 10.1016/S0933-3657(98)00028-1.

[270] UCI Machine Learning Repository. Primary tumor. UCI Machine Learning Repository, 1988. URL https://archive.ics.uci.edu/dataset/83/primary-tumor.

[271] M. M. F. Islam, R. Ferdousi, S. Rahman, and H. Y. Bushra. Early stage diabetes risk prediction. UCI Machine Learning Repository, 2020. URL https://archive.ics.uci.edu/dataset/529/early-stage-diabetes-risk-prediction-dataset.

[272] M. M. F. Islam, R. Ferdousi, S. Rahman, and H. Y. Bushra. Likelihood prediction of diabetes at early stage using data mining techniques. In *Computer Vision and Machine Intelligence in Medical Image Analysis*, volume 992 of *Advances in Intelligent Systems and Computing*, pages 113–125. Springer, 2019. doi: 10.1007/978-981-13-8798-2_12. CORRECTION: year 2019 per verified metadata.

[273] S. Borzooei and A. Tarokhian. Differentiated thyroid cancer recurrence. UCI Machine Learning Repository, 2023. URL https://archive.ics.uci.edu/dataset/915/differentiated-thyroid-cancer-recurrence.

[274] F. Thabtah. Autism screening adult. UCI Machine Learning Repository, 2017. URL https://archive.ics.uci.edu/dataset/426/autism-screening-adult.

[275] F. Thabtah. Autism Spectrum Disorder Screening: Machine Learning Adaptation and DSM-5 Fulfillment. *Informatics for Health and Social Care*, 44:278–297, 2018. doi: 10.1145/3107514.3107515. CORRECTION: year 2018, volume 44, pages 278–297 per verified metadata.

[276] M. Zieba, J. M. Tomczak, M. Lubicz, and J. Swiatek. Thoracic surgery data. UCI Machine Learning Repository, 2014. URL https://archive.ics.uci.edu/dataset/277/thoracic-surgery-data. CORRECTION: year 2014 per Zięba et al. (2014), *Applied Soft Computing*, 14, 99–108.

[277] M. Zięba, J. M. Tomczak, M. Lubicz, and J. Świątek. Boosted SVM for extracting rules from imbalanced data in application to prediction of the post-operative life expectancy in the lung cancer patients. *Applied Soft Computing*, 14:99–108, 2014. doi: 10.1016/j.asoc.2013.07.016.

[278] M. Elter. Mammographic mass. UCI Machine Learning Repository, 2007. URL https://archive.ics.uci.edu/dataset/161/mammographic-mass.

[279] M. Elter, R. Schulz-Wendtland, and T. Wittenberg. The prediction of breast cancer biopsy outcomes using two CAD approaches that both emphasize an intelligible decision process. *Medical Physics*, 34(11):4164–4172, 2007. doi: 10.1118/1.2786864.

[280] L. Rubini, P. Soundarapandian, and P. Eswaran. Chronic Kidney Disease. UCI Machine Learning Repository, 2015. URL https://archive.ics.uci.edu/dataset/336/chronic+kidney+disease. Dataset.

[281] S. E. Golovenkin, J. Bac, A. Chervov, E. M. Mirkes, Y. V. Orlova, E. Barillot, A. N. Gorban, and A. Zinovyev. Myocardial infarction complications. UCI Machine Learning Repository, 2020. URL https://archive.ics.uci.edu/dataset/579/myocardial+infarction+complications. Dataset.

[282] S. E. Golovenkin, J. Bac, A. Chervov, E. M. Mirkes, Y. V. Orlova, E. Barillot, A. N. Gorban, and A. Zinovyev. Trajectories, bifurcations, and pseudo-time in large clinical datasets: Applications to myocardial infarction and diabetes data. *GigaScience*, 9 (11):giaa128, 2020. doi: 10.1093/gigascience/giaa128.

[283] E. Tasci and K. Camphausen. Glioma Grading Clinical and Mutation Features. UCI Machine Learning Repository, 2022. URL https://archive.ics.uci.edu/dataset/759/glioma+grading+clinical+and+mutation+features. Dataset.

[284] K. Fernandes, J. S. Cardoso, and J. Fernandes. Cervical Cancer (Risk Factors). UCI Machine Learning Repository, 2017. URL https://archive.ics.uci.edu/dataset/383/cervical+cancer+risk+factors. Dataset.

[285] Centers for Disease Control and Prevention. CDC Diabetes Health Indicators. UCI Machine Learning Repository, 2017. URL https://archive.ics.uci.edu/dataset/891/cdc+diabetes+health+indicators. Dataset, derived from CDC BRFSS 2015 survey data.

[286] S. M. Hammer and D. A. Katzenstein. AIDS Clinical Trials Group Study 175. UCI Machine Learning Repository, 1996. URL https://archive.ics.uci.edu/dataset/890/aids+clinical+trials+group+study+175. Dataset.

[287] S. M. Hammer, D. A. Katzenstein, M. D. Hughes, H. Gundacker, R. T. Schooley, R. H. Haubrich, W. K. Henry, M. M. Lederman, J. P. Phair, M. Niu, M. S. Hirsch, and T. C. Merigan. A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *New England Journal of Medicine*, 335(15):1081–1090, 1996. doi: 10.1056/NEJM199610103351501.

[288] A. C. Tanrikulu and O. Er. Mesothelioma's disease data set. UCI Machine Learning Repository, 2012. URL https://archive.ics.uci.edu/dataset/440/mesothelioma+s+disease+data+set. Dataset.

[289] University of Michigan Institute for Healthcare Policy and Innovation. National Poll on Healthy Aging (NPHA). UCI Machine Learning Repository, 2023. URL https://archive.ics.uci.edu/dataset/936/national+poll+on+healthy+aging+(npha). No working DOI available. Accessed: 2026-03-19.

[290] S. Kamal and M. ElEleimy. Hepatitis C Virus (HCV) for Egyptian patients. UCI Machine Learning Repository, 2017. URL https://archive.ics.uci.edu/dataset/503/hepatitis+c+virus+hcv+for+egyptian+patients. Dataset.

[291] J. C. L. Tavera and D. García. Mental health data of 776 Mexican medical students (PHQ-9, GAD-7 and Epworth Sleepiness Scale Scores). Figshare, 2018. URL https://figshare.com/articles/dataset/1305817.

[292] M. M. Syeed. MHP (Anxiety, Stress, Depression) Dataset of University Students. Figshare, 2024. URL https://figshare.com/articles/dataset/25771164. Author list may be incomplete; verify via DataCite.

[293] M. Z. Hasan, A.-A. Galib, and M. S. Uddin. A Dataset on the Mental Health of Undergraduate Students in Bangladesh. Mendeley Data, 2024. URL https://data.mendeley.com/datasets/f4z2bfv7vk/1.

[294] M. A. I. A. Miraz, S. Mondal, and S. Ahmed. PHQ-9 Student Depression Dataset. Mendeley Data, 2025. URL https://data.mendeley.com/datasets/kkzjk253cy.

[295] L. Martínez, E. Robles, V. Trofimoff, N. Vidal, A. D. Espada, N. Mosquera, B. Franco, V. Sarmiento, and M. I. Zafra. Subjective well-being and mental health among college students: Two datasets for diagnosis and program evaluation. *Data*, 9(3): 44, 2024. doi: 10.3390/data9030044.

[296] L. Martínez, E. Robles, V. Trofimoff, N. Vidal, A. D. Espada, N. Mosquera, B. Franco, V. Sarmiento, and M. I. Zafra. Subjective well-being and mental health among college students: Dataset. Mendeley Data, 2024. URL https://data.mendeley.com/datasets/bytb22nf7m/1. Dataset accompanying Martínez et al. (2024), *Data*, 9(3), 44.

[297] R. J. McNally, D. J. Robinaugh, G. W. Y. Wu, L. Wang, M. K. Deserno, and D. Borsboom. Mental disorders as causal systems: A network approach to posttraumatic stress disorder. *Clinical Psychological Science*, 3(6):836–849, 2015. doi: 10.1177/2167702614553230.

[298] M. Vervaet, L. Puttevils, R. H. A. Hoekstra, E. Fried, and M.-A. Vanderhasselt. Transdiagnostic vulnerability factors in eating disorders: A network analysis. *European Eating Disorders Review*, 29(1):86–100, 2021. doi: 10.1002/erv.2805.

[299] Y. Wu, W. Yan, Y. Wu, and K. Peng. Adaptation and validation of the Claremont Purpose Scale to measure Chinese adolescents' purpose in life. *Journal of Research on Adolescence*, 34:776–790, 2024. doi: 10.1111/jora.12944.

[300] Y. Wu, W. Yan, Y. Wu, and K. Peng. A large-scale evaluation of Chinese youth purpose and its associations with mental health. Open Science Framework, 2024. URL https://osf.io/2xbpy/. Dataset accompanying Wu et al. (2024), *Journal of Research on Adolescence*, 34, 776–790.

[301] Y. Yamada, D.-B. Ćepulić, T. Coll-Martín, S. Debove, G. Gautreau, H. Han, J. Rasmussen, T. P. Tran, G. A. Travaglino, et al. COVIDiSTRESS global survey dataset on psychological and behavioural consequences of the COVID-19 outbreak. *Scientific Data*, 8(1):3, 2021. doi: 10.1038/s41597-020-00784-9.

[302] R. D. Hays, K. L. Spritzer, B. D. Schalet, and D. Cella. PROMIS®-29 v2.0 profile physical and mental health summary scores. *Quality of Life Research*, 27(7):1885–1891, 2018. doi: 10.1007/s11136-018-1842-3.

[303] R. C. Kessler and K. R. Merikangas. The National Comorbidity Survey Replication (NCS-R): Background and aims. *International Journal of Methods in Psychiatric Research*, 13(2):60–68, 2004. doi: 10.1002/mpr.166.

[304] M. Alegria, J. S. Jackson, R. C. Kessler, and D. Takeuchi. Collaborative Psychiatric Epidemiology Surveys (CPES), 2001–2003 [United States]. ICPSR – Interuniversity Consortium for Political and Social Research, 2007. URL https://www.icpsr.umich.edu/web/ICPSR/studies/20240. Dataset. NCS-R depression module used from DS0002 (20240-0002).

[305] C. D. Ryff, D. M. Almeida, J. Z. Ayanian, D. S. Carr, P. D. Cleary, C. Coe, R. J. Davidson, R. F. Krueger, M. E. Lachman, N. F. Marks, D. K. Mroczek, T. E. Seeman, M. M. Seltzer, B. H. Singer, R. P. Sloan, P. A. Tun, M. Weinstein, and D. R. Williams. Midlife in the United States (MIDUS 2), 2004–2006. ICPSR – Interuniversity Consortium for Political and Social Research, 2019. URL https://www.icpsr.umich.edu/web/ICPSR/studies/4652. Dataset, version 8. Datasets #89–91 use variables from DS0001 (04652-0001).

[306] G. Ver Steeg and A. Galstyan. NPEET: Non-parametric entropy estimation toolbox, 2014. URL https://github.com/gregversteeg/NPEET. Python package.

[307] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu. scikit-feature: Feature selection repository in Python, 2018. URL https://github.com/jundongl/scikit-feature. Python package.

[308] R. Vallat. Pingouin: statistics in Python. *Journal of Open Source Software*, 3(31): 1026, 2018. doi: 10.21105/joss.01026.

[309] N. Tishby, F. C. Pereira, and W. Bialek. The information bottleneck method. *Proceedings of the 37th Annual Allerton Conference on Communication, Control, and Computing*, pages 368–377, 1999.

[310] J. Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006.

Acknowledgements

This is, by some distance, the section I most dreaded writing and the one I kept putting off. Gratitude is harder to be precise about than information theory, and it is difficult to put into words the gratitude to everyone that has helped me over the years. A doctorate looks, from the outside, like a solitary thing. It is not. These pages exist because a remarkable number of people decided, at one point or another, that I was worth helping — academically, socially, or both. Here is my (very incomplete) attempt to acknowledge them.

Supervisors

Marcel and **Geeske**, thank you for your patience as I tried to convey my ideas and results, and for helping me distill what was truly important from what was written in a way that only I could understand. I hope all those hours spent reading my many drafts and long emails were worth it.

Rick, thank you for the brainstorming. I have never met anyone else who can turn a single problem into a list of twenty ideas to explore — along with two entirely new problems to solve. It took us a while, but we eventually worked out how to reconcile your Dutch directness with my Irish habit of beating around the bush and expecting you to read between the lines. Not many people can say they spent the evening before their first-ever conference talk in a cocktail bar in Vienna at 10pm, rebuilding slides and running new simulations. You knew a few drinks would take the edge off the nerves — or at least convince me the nerves and the simulations were equally worth worrying about.

Mike, you joined as promotor in my third year and promptly took on the role of translator-in-chief between Rick and me. Thank you for fighting in my corner. My PhD journey may have ended, but I shall continue on my campaign to win us a little extra borrel budget.

Collaborators

Eric, this work would look very different without you. Thank you for helping me navigate those tricky python packages to get PID to actually compute and making sure we were doing it correctly. Your insistence on making sure everything could be easily explained to others definitely elevated my work. Your enthusiasm for our collaboration gave me a great deal of the energy I needed to finish our projects, and resulted in by far the longest appendix I've ever written (and the spontaneous drinks did not hurt either).

René, thank you for your constant enthusiasm and speedy feedback. The many holes you poked in my earlier work, and the questions you asked really made me stop and take a step back, for the better. Your ideas unquestionably elevated this thesis.

Sandra, thank you for all your help deciding how to curate the datasets, and our discussions about how our computational findings would translate into clinical meaning.
Debraj, thank you for the positivity and excitement you brought, and for your knack of asking exactly the right question.
Angela, thank you for your can-do attitude and your almost supernatural powers of data curation, with you, things simply got done.
Jie, thank you for the brainstorming on synergy and networks, and your constant enthusiasm.

The Computational Science Lab

I had the honour of sharing an office with some remarkable people over these past few years — some who offered stimulating conversation, some who offered guidance, some who asked for it, and some who simply wanted to keep me hydrated on those sunny days. **Dhruv**, thank you for the drinks, the music, and for reliably making me smile. **David**, thank you for the chats about Irish music, and for watering the poor office plants whenever Dhruv and I forgot they existed. **Frederike**, self-appointed social chair and keeper of all the latest gossip: thank you for organising the activities that bonded this group together, and for the morning coffees — you left very big shoes to fill. **Victoria**, co-social chair and much missed after you left: thank you for the boat rides, the drinks, the games evenings, and the BBQ you handed down to me — and sorry for being so bad at staying in touch. **Giulia**, my brief office mate, thank you for the decorating and making the place seem homely, and of course the good coffee chats. **Jori**, thank you for tolerating my laptop as it whirred like a rocket preparing for take-off as I started my PhD journey.
To my fellow PhD students and Postdocs: thank you for making the CSL somewhere I wanted to be. **Kyuri**, thank you for your eagerness to collaborate and for the sheer courage of your Dutch; the way you switch between Dutch and English mid-sentence is a trick I keep meaning to learn. **Casper**, thank you for so reliably making me doubt whether synergy is real at all. **Bas**, thank you for trying to lure me into calisthenics — maybe someday I will manage that handstand. For now I'm glad we get to improve our meubelmaken skills together. **Jelle**, thank you for the music, the games, and for introducing me to bouldering. **Ben**, your boardgame recommendations never missed; **Isaak**, thank you for the games evenings and wonderful dinners. **Katinka**, for all the little check-ins and stories about boating life; **Nikša**, for the food you laid on at the CSL seminars, which was invariably the best and most memorable thing about them; and **Nik**, thank you for the coffee chats, the games evenings and of course for all the work you did as my paranymph.
Thank you, too, for the corridor conversations, the coffees, and the general good company, to **Martina**, **Alex**, **James**, **Antonio**, **Roland**, **Konstantinos**, **Valeria**, **Kamiel**, **Shivam**, **Lucia**, **Frank**, **Joeri**, **Matteo**, and **Vera**. Good luck to you all as you navigate your PhD journey.

Friends

Somewhere during these past few years I realised that life outside academia is at least as important as the work itself. Research has a habit of expanding to fill every available hour, and it takes good people to remind you that there is a world beyond papers, deadlines, and reviewer comments. Thank you all for giving me reasons to close the laptop every now and then.

Parva, thank you for the coffees at Science Park, the walks through Vondelpark, the Christmas dinners, and for patiently listening to more PhD-related venting than anyone should reasonably have to endure. At this point, I think we've spent almost as much time discussing Marvel as higher-order interactions.

Fiona, thank you for the tennis, the board game evenings, the gossip, and for being such an effective distraction from work. I also owe you an apology for conducting perhaps the least ethical experiment of my PhD: After borrowing your coffee for several weeks, I decided to replace it, blissfully unaware that the tin's contents had borne little resemblance to its label. The weeks of unexplained insomnia that followed turned out to have a remarkably straightforward explanation.

Sarah, quite possibly the most competitive person I know, thank you for the escape rooms, the games evenings and for planning a great trip to Ireland together. You somehow became such a fixture at CSL borrels that most people simply assumed you were doing the MSc with us.

Salomé, our research never overlapped as much as we first imagined, thank you for the wonderful Georgian food, for rescuing my very first week of teaching by coming back to Amsterdam as an invited speaker, and for organising such a fantastic trip to Georgia this summer.

Carmen, thank you for convincing me that moving to Amsterdam was a good idea, and for making those first months here feel like home. Without you, this thesis may well have been written somewhere else entirely.

Javier, after years of hearing from Carmen about one of the smartest people she knew, it was nice to finally meet you on one of my visits to Amsterdam, months before I eventually moved here. People are often surprised to learn that you're one of the people I know longest in the Netherlands—perhaps because they now mostly associate you with the blind date that Carmen and I arranged all those years ago. Thank you for all the conversations and good company over the years, from tennis and board games to an afternoon spent at an alpaca farm café. I have accepted that the promised paella is destined to remain just that: a promise.

Jurre, thank you for sharing a home with me these past few years. Somehow we managed to fit a PhD, a mini apartment renovation, countless home-cooked dinners, board game evenings, and a frankly irresponsible amount of coffee into that time. Replastering walls, painting ceilings, and endlessly debating furniture choices turned out to be a surprisingly effective antidote to staring at code all day. I also don't think either of us has quite recovered from drinking our way around the Amsterdam Coffee Festival.

Lau, thank you for our wonderfully unconventional cooking sessions—we really should

do more of them.
Jennie, thank you for the game nights and all your design advice, for helping me turn vague ideas into actual furniture choices, and for proving that an apartment can always be improved just a little bit more.
Sanne, thank you for all the fun during my MSc years. I'm still optimistic that, one day, we'll finally organise that scavenger hunt we've been talking about for years.
Annemijn, thank you for one of my very first car journeys in the Netherlands, as I drove an automatic car for the first time ever on the wrong side of the road, and for demonstrating that good company is considerably more important than good directions. We may have taken the scenic route to Muiden, but we got there in the end.
Xuemeng and **Tom**, thank you for the many board game nights, for always being enthusiastic about whatever slightly unusual activity someone suggested next, and for the steady stream of house listings that somehow convinced us to move to Amstelveen. I'm looking forward to being neighbours, and to a much shorter cycle home after game nights.
Alexander, thank you for always being up for a drink in the sun and for saying yes to whatever adventure seemed like a good idea at the time. Whether it was finally making good on some of those promised cocktails, watching an Ajax game, going to a masquerade ball, or somehow ending up at a Hans Klok magic show, there was never a shortage of things to do.
Davit, thank you for stepping in to save our Hyrox team after all that training, and for organising such an unforgettable trip to Georgia.
Hongyu, thank you for some of the most wonderfully random adventures of my PhD: a campervan trip through the Ardennes, countless expeditions through Dutch thrift shops in search of uranium glass, and somehow always returning home with far more than either of us intended.
Jorien, my honorary "fake-Dutch" friend, thank you for proving that being Dutch does not necessarily mean knowing anything about the Netherlands. Whether we were searching beaches for clams, surviving bootcamp sessions, or trying to answer questions that every Dutch person was apparently supposed to know, life was always more entertaining with you around.
Emma, thank you for always making time to catch up whenever I'm back in Ireland and for ensuring I leave fully briefed on every piece of local gossip I could possibly have missed.

Family and Home

If research expands to fill your time, it is family that anchors you to who you are when the laptop is closed. To my family back home, thank you for being the ultimate safe harbour from the pressures of academia.
Mam and **Dad**, thank you for your unwavering belief in me, for your endless encouragement from across the Irish Sea, and for always being at the other end of the phone whenever the PhD felt a bit too heavy. I wouldn't have made it to the start line of this journey, let alone the finish, without you both. To my brothers, **Oisin** and **Fionn**, and

my sister, **Naoise**: thank you for ensuring that whenever I came home, I was thoroughly roasted, properly grounded, and reminded that having a doctorate doesn't excuse you from the tea round.
Granny, thank you for keeping me completely up to date with the precise comings and goings of absolutely everyone back home; you ensured I never truly felt far away. I also want to remember my **Nana**, who passed away during my studies. Thank you for always wanting to know if I was enjoying whatever it was I was doing, and for remembering every single little detail of my life, even the ones I managed to forget myself. I miss her deeply, and I know she would have loved to see this day.
To my extended family, who made my visits home so bright and kept life entertaining: **Mary**, thank you for the brilliant, adventurous boat tours around Amsterdam, and for supporting my academic journey in more ways than I can count. **Kevin**, thank you for always being proudly on standby to solve my IT needs; even if information theory is slightly outside the remit of your tech company, the enthusiasm was highly appreciated. And **Suzie**, as a fellow academic, you hold the unique title of being the only family member who actually read (or at least convincingly claims to have read) my papers. Thank you for the wonderful visit to Amsterdam and for understanding the strange world of academia better than most.
To my aunts, uncles and their partners: **Christina**, **Rosario**, **Ciaran** and **Grainne**, **Margaret** and **Pauric**, **Sean** and **Bonnie**, **Imelda** and **Patrick**, **Colm** and **Sheila**, **Deirdre** and **Michael**, **Maria** and **Thierry**, **Elaine** and **Noel**, **Ciaran**, and everyone else in the Hourican and the O'Reilly Clans. Thank you all for the warmth, laughter, and sense of belonging that made every trip home feel like a reminder of where I came from.
Heather, thank you for always having a bottle of bubbles ready and for your unwavering commitment to creating memories. Whether you were organising evenings at Hippie Fish, convincing us that a historic F1 event or a polo match was exactly how we should spend a day, taking us to Curaçao, or enthusiastically guiding us around your favourite garden centre, there was never a shortage of adventures. Thank you for your warmth, generosity, and for making me feel so welcome. **Louise**, thank you for all the adventures we've shared, from Hans Klok to trips around England and making sure Joshua wouldn't drown me in Curaçao.
Finally, my deepest and most special thanks go to **Joshua**, mijn liefje, mijn eigen NEE-hoorn. Thank you for being there through every high and low of this PhD, for your endless patience, for always knowing when I needed Maltesers, and for reminding me that there was a world beyond academia. Whether we were escaping on city trips, sitting in the sunshine with a drink, or simply spending a quiet evening together, you somehow made everything feel lighter. This thesis may be dedicated to my research, but these years have been infinitely better because I got to share them with you.

List of Publications

Publications included in this thesis

1. **Hourican, C.**, Peeters, G., Melis, R. J. F., Wezeman, S. L., Gill, T. M., Olde Rikkert, M. G. M., & Quax, R. (2023). Understanding multimorbidity requires sign-disease networks and higher-order interactions, a perspective. *Frontiers in Systems Biology*, 3, 1155599. (Chapter 2)

2. **Hourican, C.**, Li, J., Mishra, P. P., Lehtimäki, T., Mishra, B. H., Kähönen, M., Raitakari, O. T., Laaksonen, R., Keltikangas-Järvinen, L., Juonala, M., & Quax, R. (2024). Efficient search algorithms for identifying synergistic associations in high-dimensional datasets. *Entropy*, 26(11), 968. (Chapter 3)

3. Peeters, G.[*], **Hourican, C.**[*], Lees, M., Quax, R., Melis, R. J. F., Kok, A. A. L., van Schoor, N. M., & Olde Rikkert, M. G. M. (2026). Symptoms provide predictive information on daily functioning beyond signs and diseases in middle-aged and older adults, particularly those with multimorbidity. *Age and Ageing*, 55(2), afag046. ([*]*these authors contributed equally*) (Chapter 4)

4. **Hourican, C.**, Peeters, G., Melis, R. J. F., Kok, A., van Schoor, N. M., Wezeman, S. L., Lees, M., Olde Rikkert, M. G. M., & Quax, R. (2026). Interpreting higher-order dependence in multimorbidity using cohort data: A partial information decomposition approach. *PLOS Computational Biology*, 22(6), e1014386. (Chapter 5)

5. **Hourican, C.**, Dignum, E., Quax, R., & Roy, D. (2026). Embedding-based partial information decomposition for symptom networks at system scale. *Manuscript under review.* (Chapter 6)

Publications not included in this thesis

1. Gehlen, J., Li, J., **Hourican, C.**, Tassi, S., Mishra, P. P., Lehtimäki, T., Kähönen, M., Raitakari, O. T., Bosch, J. A., & Quax, R. (2024). Bias in O-information estimation. *Entropy*, 26(10), 837.

2. Wezeman, S. L., Melis, R. J. F., **Hourican, C.**, Quax, R., Gill, T. M., Olde Rikkert, M. G. M., & Peeters, G. (2026). Nonspecific symptoms explain functional limitations in older adults better than disease-specific symptoms. Accepted for publication in *European Geriatric Medicine*.

3. Wezeman, S. L., Melis, R. J. F., **Hourican, C.**, Quax, R., van Schoor, N. M., Kok, A. A. L., Olde Rikkert, M. G. M., & Peeters, G. (2026). Symptom-symptom correlations are stronger in community-dwelling older adults with multimorbidity than in those with one or no disease. *In preparation.*

4. Li, J., Bosch, J. A., Rydin, A. O., **Hourican, C.**, Koloi, A., Tassi, S., Mishra, P. P., Mishra, B. H., Kähönen, M., Lehtimäki, T., Raitakari, O. T., Laaksonen, R., Keltikangas-Järvinen, L., Juonala, M., & Quax, R. (2025). A multilayer network analysis of cardiovascular–depression comorbidity reveals symptom-specific molecular biomarkers. *Psychological Medicine*, 55, e316.

5. Koloi, A., Loukas, V. S., **Hourican, C.**, Sakellarios, A. I., Quax, R., Mishra, P. P., Lehtimäki, T., Raitakari, O. T., Papaloukas, C., Bosch, J. A., März, W., & Fotiadis, D. I. (2024). Predicting early-stage coronary artery disease using machine learning and routine clinical biomarkers improved by augmented virtual data. *European Heart Journal – Digital Health*, 5(5), 542–550.

