Mapping the Dynamics of Depression

Uncovering Temporal Patterns in Digital Biomarkers through Network Modelling

Dagmar Frenken

Delft University of Technology

TU Delft

Mapping the Dynamics of Depression

Uncovering Temporal Patterns in Digital Biomarkers through Network Modelling

by

Dagmar Frenken

Dagmar
Frenken

Chair/First supervisor:	Dr.ir. P. Heijnen
Second supervisor:	Dr.ir. H. Torkamaan
External supervisor:	Dr. G. Aalbers
Project Duration:	Feb, 2025 - Aug, 2025
Faculty:	Faculty of TBM, Delft

Cover: @ NetworkWorld
Style: TU Delft Report Style

TU Delft

Executive Summary

Depression is a common and often recurring mental health condition. Many people who experience depression go through periods of varying depression symptoms. Traditional ways to monitor depression focus on questionnaires and interviews during clinic visits. These methods are useful but have clear limits. They rely on memory, give only a snapshot of someone's condition, and may miss small but important changes in how a person feels or behaves outside of clinical assessments.

Today, wearable devices such as smartwatches and smartphones make it possible to track behaviours continuously in daily life. These behaviours include how much someone moves, how they sleep, and how their heart rate changes. These measurements are called digital biomarkers. They can be collected with minimal effort from the person and provide an objective picture of what happens in their body and behaviour from day to day. This method of using digital devices to measure behaviour and physical signals in real life is called digital phenotyping. It offers a new way to observe how people function outside the clinic, in their everyday environment, and helps detect subtle changes that may not be visible through traditional interviews or questionnaires.

This thesis explores how digital biomarkers interact with each other over time in individuals with varying levels of depression symptoms. The main idea is based on a network approach. In this approach, depression is not seen as one fixed condition, but as a system where different behaviours and symptoms can influence each other. A network illustrates which signals are connected and how changes in one signal can lead to changes in others. This makes it possible to see not only what changes, but how change happens.

Two kinds of networks were studied. The first is called a temporal network. This network shows how one signal can influence another across days. For example, if someone sleeps poorly one night, this may affect their heart rate or activity level the next day. In this way, a temporal network helps identify patterns where one behaviour leads to changes in another over time.

The second kind is a contemporaneous network. This network shows how signals relate to each other at the same moment in time. It shows which behaviours tend to go together at the same moment. For instance, someone may sleep less and also have a higher heart rate simultaneously. This kind of network gives insight into how behaviours cluster together.

To understand the structure of these networks, the study looked at two important features: centrality and connectivity. Centrality tells us which signals are the most influential in the system. A signal with high centrality has many strong links to other signals. In this study, heart rate and time spent in sleep stages other than rapid eye movement (REM) sleep (NREM) are central in the temporal network. Heart rate was also the most stable signal over time and often predicted changes in other areas. For example, when heart rate was higher than usual, people tended to sleep less that night.

Connectivity reflects how strongly the signals in a network are linked overall. A network with high connectivity has many strong connections between its signals, while a network with low connectivity has fewer or weaker links. This study found that people with more depressive symptoms had more strongly connected temporal networks. This means their behaviours were more tightly linked from day to day. A system like this may be less flexible, because a change in one signal is more likely to cause changes in others. In contrast, people with fewer symptoms had looser connections between signals, which may allow for more adaptability and recovery.

The contemporaneous networks were more stable across groups. Most same-moment relationships between variables remained similar regardless of symptoms. This suggests that the most meaningful changes in depression may show up not in what happens within one day, but in how patterns unfold across days.

To inform personalised monitoring, this study suggests it is more useful to track how digital biomarkers

interact over time, rather than looking at them in isolation. Also, increases in connectivity between signals may indicate reduced flexibility and a higher risk of symptom increase, allowing systems to give early alerts. Focusing on central signals like heart rate and NREM sleep may also help guide more efficient interventions.

These findings have implications for patients, clinicians, and the broader health care system. For patients, personalised monitoring could support earlier recognition of warning signs, but could also feel intrusive. For clinicians, digital tools may offer a more efficient way to intervene before symptoms escalate. On a system level, digital monitoring could shift care toward prevention, helping reduce pressure on mental health services and shorten waiting times.

At the same time, ethical concerns must be taken seriously. Many mental health apps lack clinical oversight and rely on unverified models. There is a risk of overpromising what these technologies can do, and of shifting attention away from lived experience toward patterns in data. Current legal frameworks often lag behind technical developments. Finally, without inclusive design, these tools may reinforce existing inequalities in access to care. Addressing these risks requires regulation.

This study also has several limitations. It did not model person-specific networks, nor did it track changes in network structure over time. While the current analysis captures group-level patterns, future work should explore how these dynamics vary within individuals and across illness phases . Centrality estimates were reported, but their interpretation in psychological data remains contested. Robustness checks were limited due to current software constraints for multilevel temporal models. These factors suggest that results should be interpreted as exploratory.

Future research should focus on making these methods more precise, personalised and clinically useful. This includes developing person-specific thresholds for network change, improving statistical robustness, and combining passive data with patient narratives or self-reports. Doing so can help close the gap between what is measured and what is meaningful.

In conclusion, this thesis shows that digital biomarkers interact in structured ways that change over time depending on symptom severity. Understanding these patterns may help build tools that support mental health in a more personal and timely way. In the future, this could help make digital phenotyping models not only descriptive but also predictive and actionable in real-world care.

Preface

I am thankful to the multiple people who have helped me along the way; writing this thesis has been both an academic and personal journey.

First and foremost, I want to express my gratitude to Dr.Ir. P. Heijnen, my first supervisor and chair, for the consistent guidance, helpful criticism, and willingness to let me experiment with my own ideas. I am also thankful to my second supervisor, Dr.Ir H. Torkamaan, for insightful methodological discussions and clear advice at key moments. A special thanks to my external supervisor, Dr. G. Aalbers, whose practical perspective and experience greatly sharpened the final version of this work.

This thesis marks the end of my master's programme, but also the beginning of a deeper interest in the intersection between mental health and data science. I hope it contributes, in a small way, to how we understand and support those living with depression.

Dagmar Frenken
Delft, July 2025

During the writing process of this thesis, language models such as ChatGPT (OpenAI), QuillBot, and Grammarly were used to improve the clarity, grammar, and academic tone of selected passages. These tools were applied to scan and refine existing text written by the author. All methodological decisions, conceptual frameworks, and interpretations of results were developed independently by the author. The final content was critically reviewed and revised by the author to ensure accuracy and originality.

Contents

List of Figures

Nomenclature

Symbols

Symbol	Definition	Unit
ρ	Partial correlation coefficient	-
β	Standardised regression coefficient	-
t	Time point (in days)	days
$t-1$	Previous time point (lag-1)	days
y_t	Vector of observed variables at time t	Varies
y_{it}	Value of variable i at time t	Varies
$y_{j,t-1}$	Value of variable j at previous time point $t-1$	Varies
μ	Vector of person-specific means	Varies
B	Temporal effect matrix ($p \times p$)	-
b_{ij}	Effect of variable j on variable i at lag-1	-
ζ_t	Innovation (error) term at time t	Varies
Σ	Covariance matrix of residuals ζ_t	-
ϵ_{it}	Residual of variable i at time t after temporal regression	Varies

1 Introduction

Depression is one of the leading causes of disability worldwide. It accounts for 14.3% of all years lived with disability across age groups, making it one of the most burdensome health conditions globally [32]. Depression is a complex condition caused by a combination of genetic, biological, psychological, and environmental factors [6], with symptoms such as (hyper)insomnia, loss of energy, feelings of worthlessness and guilt, reduced ability to think or concentrate, and even premature, self-induced death as a consequence [40]. Depression places a heavy burden not only on individuals but also on the healthcare system and society. People living with depression often experience long-term emotional, social, and economic consequences. Treatment costs per patient range from 2,471 to more than 51,000, not including indirect costs such as lost productivity or long-term disability [13].

For many patients, depression is not a single episode, but a lifelong condition. Instead of full recovery, many experience alternating phases of varying levels of symptoms. Approximately 55% of individuals with depression have multiple episodes across their lifespan [62]. This pattern highlights the need for ongoing, long-term monitoring, not just during acute episodes, but also in periods of apparent recovery. Timely detection of symptom changes is crucial to reduce severity through early intervention. It is also demonstrated that while patients may not be in active depression, often a range of symptoms is experienced, from less to more severe [33]. These symptoms vary widely between individuals and change over time, making the disorder highly heterogeneous and unpredictable.

The complexity of depression is a challenge to researchers. Patients with depression, in particular, have a complex treatment trajectory due to the heterogeneity of the disease, the unpredictability of depression, and the inconsistency of therapy outcomes [50]. However, current monitoring methods are limited. In most clinical settings, symptom tracking relies on self-report through questionnaires or interviews. These assessments are often infrequent and subject to recall bias, which is a type of memory error in which individuals inaccurately recall past experiences, leading to unreliable data [63]. As a result, important warning signs may go unnoticed, and treatment may not be adjusted in time. This is particularly problematic in daily life outside clinical care, where individuals might not be under active treatment. Early detection of symptoms is crucial to enable timely intervention before a full depressive episode develops. Therefore, more frequent, objective, and continuous forms of monitoring are needed to improve the management of depression over time.

1.1. Future of Depression Tracking

Current monitoring methods expose challenges for researchers due to self-reporting and infrequency. To solve these challenges, research has been done on continuous, objective monitoring methods, leading to several key concepts. The following paragraphs describe the key concepts used in this study.

1.1.1. Digital Phenotyping

New technologies have made it possible to collect health data in real time, during people's daily lives. A key concept in this development is digital phenotyping. As noted by Onnela et al. (2021), digital

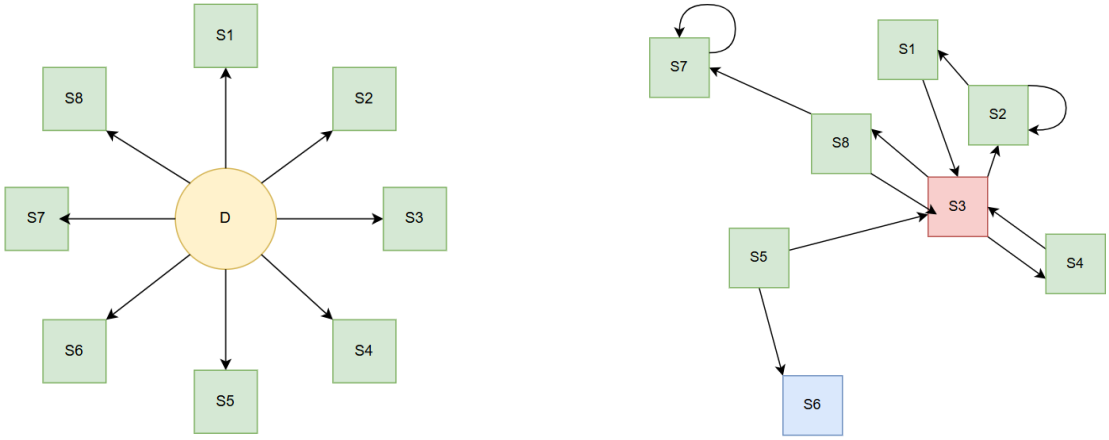
phenotyping refers to the “moment-by-moment quantification of the individual-level human phenotype in situ using data from personal digital devices” [46]. In simpler terms, it means using digital data to understand how someone functions in everyday life, not just in the clinic [31]. The data that is retrieved through digital phenotyping can be described as digital biomarkers. These are measurable characteristics, collected using wearables, that serve as indicators of biological processes, behaviour, or responses to treatment [61]. Unlike traditional biomarkers, which are often obtained through clinical tests or laboratory measurements, digital biomarkers are gathered passively or actively via devices like smartphones or wearables in real-world settings. These digital biomarkers can reflect behaviour, or physiology, such as sleep patterns, heart rate variability, or phone usage. By combining multiple digital signals, researchers could gain deeper insight into a person’s health status and detect meaningful changes over time. In mental health research, this can support earlier detection of depression, better monitoring of symptom fluctuations, and more personalised care [61].

1.2. Characteristics of Network Design for Depression

The high-frequency, multi-domain data collected through digital phenotyping offers the possibility to examine how digital biomarkers interact over time. The following section outlines the rationale of network-based approaches in this context.

1.2.1. Network Approach Model

In most traditional models, depression is seen as a single hidden disorder that causes a range of symptoms. This idea is called the common cause model [24], displayed in Figure 1.1a. It assumes that symptoms like insomnia, low mood, or fatigue all come from one underlying disease, depression, in this case. In this model, symptoms are treated as equal and interchangeable indicators of depression. However, research has shown that this way of thinking does not match the reality of how depression works [24, 14]. People with the same diagnosis often have very different symptoms [25]. Also, symptoms can influence each other directly. For example, sleeping poorly can lead to tiredness, which may reduce concentration. Difficulties concentrating can then result in problems at work or school, increasing stress and potentially worsening sleep. This creates a vicious cycle in which symptoms reinforce one another. This means that symptoms may not just be effects of a disease, but part of a system that feeds into itself. A network approach offers a different view than the common cause model. Instead of seeing depression as one cause with many effects, it sees depression as a group of connected elements that affect each other [24], see Figure 1.1b. This way of thinking helps us understand the complexity of depression. It allows researchers to look at how different aspects of someone’s behaviour and physiology interact, rather than treating them as separate or static. For this reason, network models are becoming more common in research on depression and digital health. Also, the network approach encourages researchers to pay attention to the unique role of individual symptoms, rather than treating them as equivalent. This leads to important clinical insights. In a symptom network, some symptoms are more central than others, meaning they are strongly connected to many other symptoms [24]. As shown in Figure 1.1b, a central symptom (in red) has many connections. While a peripheral one (in blue) is more isolated, meaning it has fewer or weaker connections and thus less influence on the rest of the system. Targeting a central node may lead to cascading effects, that is, changes in one variable may spread to others via their connections, throughout the system, while intervening on a peripheral node is less likely to produce widespread change [26]. Network models allow researchers to identify central signals using centrality measures, which quantify the strength and number of a node’s connections [18]. For example, if poor sleep is consistently central, improving it may also improve energy, mood, or activity levels.



(a) Common cause model

(b) Network model

Figure 1.1: The common cause model (a) and the network model (b)

While the original network theory focuses on symptoms, the same idea can be applied to digital biomarkers such as sleep, heart rate, or activity levels. These signals, often collected passively and continuously via wearable devices or smartphones, provide high-frequency insights into an individual's physiological and behavioural state. When modelled as a dynamic network, these biomarkers can reveal meaningful interactions and feedback loops across time, supporting a more personalised understanding of mental health in everyday life [54].

1.2.2. Network Characteristics

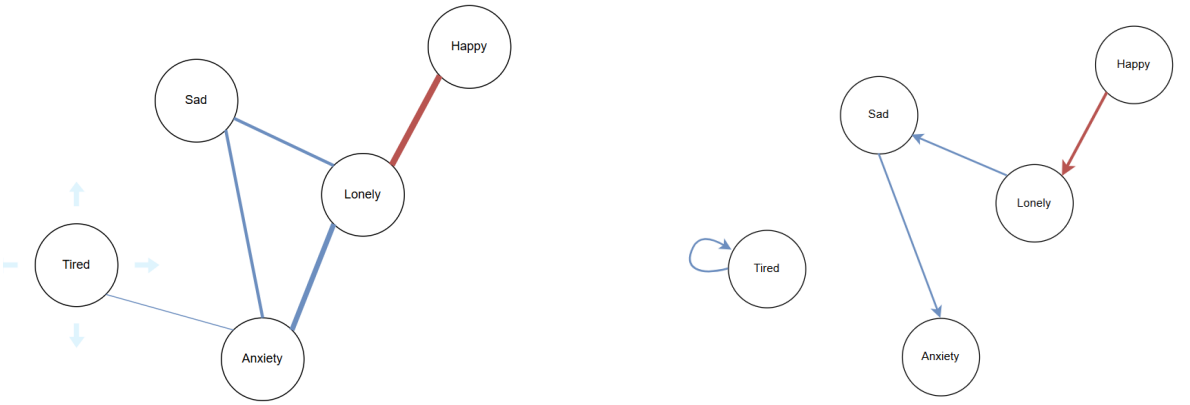
Network models represent complex systems as sets of nodes and edges. In the context of depression, each node represents a variable of interest. These can be symptoms (such as insomnia or low mood) or digital biomarkers (such as heart rate, step count, or sleep duration). Edges are the connections between nodes and reflect statistical relationships. An edge indicates that two variables are related in a meaningful way, either over time or within the same moment. There are two main types of networks that researchers use to study mental health over time: temporal networks and contemporaneous networks [18], these are the focus of this thesis.

Temporal networks (Figure 1.2b) show how one variable predicts another at a later time point. For example, if feeling lonely today increases the chance of feeling sad or anxious tomorrow, this would be represented as arrows from *Lonely* to *Sad* and *Anxiety* in the temporal network. These networks are based on estimation models that estimate how variables at time t influence variables at time $t + 1$. Arrows in the network represent directional influence, helping researchers understand how patterns unfold over time. For instance, in the figure, we also see that feeling *Tired* tends to persist across time (self-loop), and that feeling *Happy* today may reduce *loneliness* the next day.

Contemporaneous networks (Figure 1.2a) capture associations between variables that occur within the same time window. For example, if feeling *Lonely* often co-occurs with feeling *Sad* or *Anxious*, these variables are connected in the contemporaneous network. These connections are undirected and reflect partial correlations that remain after accounting for temporal effects. A partial correlation differs from a regular correlation because it measures the unique association between two variables while controlling for all other variables in the network. In this way, it shows direct relationships, rather than shared effects driven by other variables. Contemporaneous networks are especially useful for detecting fast-occurring processes, such as feeling *Lonely* and experiencing reduced *Happiness* at the same moment.

Both types of networks reflect *within-person* processes: they describe how an individual's state changes over time and how different variables influence each other within that same person. This is different from traditional research that looks at *between-person* differences, such as how people who sleep less tend to be less active on average. Instead, network models can show how, for one specific person,

sleeping poorly on a given night may lead to being less active the next day. This makes it possible to study how processes unfold in daily life [16].



(a) Contemporaneous symptom network: connections estimated within the same time window.

(b) Temporal symptom network: lagged relationships estimated from one time point to the next.

Figure 1.2: Illustrative examples of (a) contemporaneous and (b) temporal networks.

1.3. Knowledge Gap

In order to define the knowledge gaps in current research, a literature review has been conducted on studies applying network models to depression. The aim of this review was to examine how current methods capture the dynamic nature of depression, particularly in relation to digital signals collected through wearables. While previous research has shown that symptoms of depression are interrelated and can reinforce each other over time, most studies still rely on static models, short measurement periods, or isolated behavioural domains. Moreover, no studies have applied network models to continuously collected digital biomarker data to explore how these signals interact in daily life. As a result, it remains unclear how such digital biomarker signals behave over time within individuals with varying levels of depression symptoms, and how their dynamic patterns relate to symptom severity. This thesis addresses that gap by using network modelling to explore the temporal and contemporaneous relationships between digital biomarkers. In doing so, it contributes to the development of more personalised and responsive forms of depression monitoring.

1.4. Research Questions

To address the identified knowledge gap, the following research question has been developed:

How do digital biomarkers dynamically interact over and in time in individuals with varying levels of depression symptoms, and how can these interactions inform personalised monitoring?

To address this question, the following subquestions are examined:

1. What is the structure of the digital phenotyping biomarker data, and how can it be preprocessed for analysis?
2. What temporal and contemporaneous dynamics and centrality patterns of digital biomarkers can be identified in individuals with varying levels of depression symptoms?
3. How do digital biomarker dynamics and network connectivity differ across individuals with varying levels of depression symptoms?
4. How can insights from digital biomarker networks be applied to personalised monitoring and care in depression?

1.5. Research Objective and Societal Aim

The goal of this thesis is to better understand how digital biomarkers interact over time in people with varying levels of depression. Using network models, this study models both short-term (contemporaneous) and longer-term (temporal) connections between these signals. The aim is to explore whether these patterns differ between individuals with varying levels of depressive symptoms. By visualising these interactions as dynamic networks, the research aims to provide insight into the day-to-day structure of depression. These networks may help identify which signals are most central in people's lives and how they influence one another. In doing so, the study contributes to the scientific understanding of depression as a dynamic system and demonstrates how methods from network science can be applied to digital health data.

From a societal perspective, this work supports the development of more personalised and proactive mental healthcare. Many individuals experience fluctuations in depressive symptoms that are difficult to predict with current tools. Digital biomarkers collected via wearables offer a new opportunity to track changes continuously and objectively. However, turning these data streams into useful insights remains a challenge. This thesis addresses that gap by demonstrating how complex data can be structured and interpreted in a clinically meaningful way.

Finally, the research reflects the interdisciplinary nature of the CoSEM MSc programme. It connects data science, healthcare, and systems thinking, and deals with a real-world problem where human behaviour, technology, and care provision are closely linked. In this way, the project demonstrates how methods from complex systems engineering can contribute to more adaptive, data-informed mental health solutions.

1.6. Report Outline

This thesis is structured as follows:

- **Chapter 2 – Literature review**: This chapter provides an overview of the theoretical and methodological foundations of network models used to study mental health. It introduces the network approach in psychiatry and discusses prior research applying these methods to both symptom reports and digital biomarkers.
- **Chapter 3 – Methodology**: This chapter outlines the overall research design and analytic strategy. It describes the network modelling approach used in this thesis to estimate temporal and contemporaneous networks.
- **Chapter 4 – Data**: This chapter outlines the preprocessing procedures applied prior to modeling, including outlier handling, missing data imputation, and dimensionality reduction.
- **Chapter 5 – Results**: This chapter presents the results of the network analyses. It describes the estimated temporal and contemporaneous networks, compares network characteristics across symptom severity groups, and highlights key findings related to centrality and connectivity.
- **Chapter 6 – Discussion**: The final chapter interprets the main findings in light of existing literature. It reflects on the implications for personalised monitoring, discusses methodological strengths and limitations, and offers suggestions for future research.

2

Literature review

This chapter reviews key developments in the use of network models to study depression. It focuses on how temporal and contemporaneous networks are used to model dynamic interactions between symptoms or behavioural signals. Earlier studies using network analysis are discussed to provide context for the methodological and theoretical choices made in this thesis. The chapter concludes with the research gap this study addresses.

2.1. Search Strategy

Literature was identified using multiple databases, including Google Scholar, Scopus, and PsycINFO. These were chosen because of their accessibility and relevance to research in behavioural and psychological health. Search terms included combinations of *"depression"*, *"digital phenotyping"*, *"network analysis"*. Boolean operators were used to increase sensitivity, for example:

- `"depression" AND "digital phenotyping"`
- `"depression AND network AND Temporal OR Contemporaneous"`
- `"symptom network AND depression"`

Studies were included if they focused on depression and used either digital phenotyping or network analysis. Titles and abstracts were screened for relevance, after which the full texts of selected articles were reviewed. Reference lists of included papers were used to identify additional sources.

2.2. Why Personalised Monitoring for Depression Matters

Depression affects more than 280 million people worldwide [64] and represents a leading cause of disability across all age groups [41]. Although effective treatments for depression exist, many individuals still do not receive adequate care. Barriers such as stigma, financial cost, limited therapist availability, or geographic isolation can prevent people from accessing traditional face-to-face therapy [3]. These challenges have accelerated the development of digital interventions, which are treatment options provided through digital devices such as smartphones, web platforms, or other connected technologies. Their goal is to expand access to care and provide flexible support outside the clinical setting [35].

The COVID-19 pandemic further accelerated the adoption of digital interventions in psychiatry, making them a prominent component of (mental) health care. Research provides evidence for the efficacy and effectiveness of digital interventions for the treatment of depression [41]. However, most existing tools are still designed around fixed treatment protocols that apply the same content to all users. This "one-size-fits-all" approach overlooks the fact that depression is a highly heterogeneous condition; people differ in which symptoms they experience, how severe they are, and how they change over time [35]. This lack of individual tailoring contributes to high dropout rates, suboptimal outcomes, and limited long-term effectiveness [52]. It also reflects a broader limitation of current digital interventions: although they deliver therapy digitally, they often do not monitor the user dynamically. They treat people with

depression as static cases rather than fluctuating individuals [35].

In response, there is growing movement toward personalised monitoring, an approach that uses real-time data to track how symptoms evolve within individuals over time. Instead of applying the same treatment to everyone, personalised monitoring enables interventions to respond adaptively to a user's current state, symptoms, or behavioural patterns [35]. As Domhardt et al. (2021) point out, digital tools are uniquely suited to support this development. By collecting continuous data through digital phenotyping, they make it possible to capture behavioural fluctuations in everyday life, free from recall bias and detached from the constraints of the clinic [15].

Moreover, because digital platforms allow for longitudinal data collection, they open new doors for psychotherapy research. They enable the analysis of how change unfolds over time and offer the opportunity to develop models that forecast symptom trajectories for individual users [35]. In this way, personalised monitoring and digital interventions are not just scalable delivery methods; they are also a gateway to more personalised, responsive, and preventative care. This is especially relevant in the context of depression, a condition marked by fluctuating symptoms [62]. Without continuous monitoring, important early signs of depression symptoms may go unnoticed, delaying necessary adjustments in treatment. Personalised monitoring makes it possible to detect changes in real time, allowing for timely intervention.

2.2.1. Heterogeneity in Depression

Depression is not a uniform condition. People with depression can present with very different symptom profiles, illness trajectories, and treatment responses [24]. In fact, two individuals may both meet the diagnostic criteria for depression without sharing a single symptom. One might struggle with fatigue, insomnia, and self-blame, while another experiences sadness, weight loss, and suicidal thoughts. A study by Fried et al. (2015) identified more than a thousand unique symptom combinations in a sample of just a few thousand patients [27]. This symptom heterogeneity challenges the idea of depression as a single, coherent disorder and suggests instead that it may be a broad category of interacting problems. Some symptoms appear more closely linked to treatment response and functional impairment than others [26]. Moreover, symptoms may influence one another dynamically: insomnia may lead to fatigue, which may impair concentration and increase feelings of hopelessness. These interactions can reinforce depressive states and may serve as early warning signs of worsening mental health [26].

Most data-driven models still rely on general predictions derived from group-level data. These models typically capture population averages and assume that findings generalise across individuals. However, this assumption often fails to hold, especially in psychiatry. For instance, Muller et al. (2021) showed that a depression prediction model trained on a homogeneous student sample failed to generalise to a more diverse population [42]. This illustrates a broader issue in psychological and psychiatric modelling, particularly in the context of depression, which is characterised by heterogeneity. As Fisher et al. (2018) argue, most psychological processes are non-ergodic, meaning that group-level statistics do not necessarily reflect the dynamics observed within individuals [22]. Their study demonstrated that within-person variability can be up to four times greater than between-person variability, undermining the generalisability of group-based models.

To address the heterogeneity of depression, models are being explored that tailor predictions to individual users. These models often retain a shared structure but adjust specific parameters based on the person's unique data. This personalised approach helps to capture variability in symptom expression and leads to more accurate predictions [35]. Symptom or behaviour-based modelling may not only improve prediction but also help identify which patients benefit most from specific interventions. As Fried argues, adopting a symptom or behaviour-level perspective could unlock insights that may inform more effective and personalised treatments [26].

2.2.2. Limitations of Traditional Depression Assessment

The tools used to assess depression have changed little over the past decades. Self-report questionnaires remain the standard for screening and monitoring depressive symptoms. They rely on retrospective self-reports and assume that symptoms reflect a stable underlying condition. However, there is growing evidence that these assumptions are problematic [26].

One major limitation of traditional assessment tools is their lack of ecological validity. Mood, energy, sleep, and other behaviours can fluctuate considerably in daily life, often in response to situational factors such as stress, poor sleep, or social interactions. Because these instruments are typically administered only during clinical appointments or scheduled follow-ups, they fail to capture such short-term dynamics and may overlook early warning signs of worsening symptoms. As Insel (2018) notes, psychiatric diagnoses are often based on broad and static summaries that do not reflect how individuals function in daily life [31]. Such one-off diagnostic assessments cannot capture symptom change that could indicate depression symptoms.

In addition to conceptual limitations, traditional assessments are prone to practical issues such as recall bias and reduced objectivity. Asking individuals to summarise how they felt over extended periods can lead to memory distortions, particularly in those experiencing depression, who often show cognitive difficulties and fluctuating symptoms [63]. Furthermore, these tools depend solely on subjective evaluations and do not incorporate behavioural data, such as sleep, movement, or phone usage. This limits their ability to detect early changes in symptoms and further underscores the need for more ecologically valid, real-time assessment methods.

To overcome these limitations, research recommends more frequent, context-sensitive, and passive approaches to mental health monitoring. Digital phenotyping offers promising alternatives by enabling continuous, real-world data collection on mood, behaviour, and physiology [29, 46]. These approaches hold the potential to better capture the dynamic nature of depression and support more personalised, timely interventions.

2.3. The Promise and Pitfalls of Digital Phenotyping

Recent research demonstrates the potential of digital phenotyping to transform how we monitor depression over time. By leveraging data from smartphones and wearables, digital phenotyping allows researchers to passively and continuously track behaviours such as mobility, sleep, and phone usage, and other behaviours known to relate to mood and symptom severity [29]. Compared to conventional assessments, digital phenotyping offers key advantages: it reduces recall bias, enables large data collection, and supports within-person analyses of short-term symptom fluctuations [29]. These features make it especially suitable for disorders like depression, which often follow fluctuating trajectories. Moreover, longitudinal data open the door to advanced modelling techniques, such as dynamic network models, that can identify reinforcing or causal links between behaviours and symptoms over time [10, 48].

However, these opportunities come with considerable challenges. First, most existing studies suffer from short follow-up periods or small sample sizes, limiting generalisability and statistical power [48, 37]. In addition, issues remain around user engagement, data completeness, and privacy [29]. Missing data due to battery failures, disengagement, or device issues is common and can bias results [29].

Despite these limitations, the momentum for digital phenotyping is growing rapidly. As shown by Heckler et al. (2025), the number of studies using digital phenotyping in mental health has sharply increased since 2018. Most studies combine passive and active data modalities and use machine learning or regression-based models to detect symptom trajectories. Yet the field still lacks standardisation in data collection and analysis methods, which limits comparability across studies [29]. To unlock the full potential of digital phenotyping, future research must focus on integrating data across behavioural domains, ensuring user-centred design to improve engagement, and addressing privacy and ethical concerns more rigorously.

Overall, digital phenotyping holds great promise as a tool for monitoring depression, but its practical implementation requires methodological refinement, and fitting modelling tools.

2.4. Network Approaches in Psychiatry

Network approaches are a promising, novel avenue to model the dynamic interactions captured by digital phenotyping data. Network models allow researchers to examine how symptoms and behaviours influence one another over time. Instead of reducing depression to a single total score, this approach maps how symptoms or behaviours interact dynamically, enabling a more nuanced understanding of

mental states [8, 24]. In this framework, symptoms or behaviours are represented as nodes, and their statistical relationships are visualised as edges [17].

One consistent finding across studies is that network connectivity, defined as the overall strength of associations among symptoms, increases with the number of active symptoms. This phenomenon, observed in both simulations and empirical work, implies that more severe episodes of depression may be characterised by stronger and more self-reinforcing symptom interactions [14, 10, 48]. This insight has sparked growing interest in using network characteristics as early warning signals for depression symptoms or targets for intervention.

Network models not only describe associations but can also help identify symptoms or behaviours that play an important role in maintaining a disorder. These so-called “central symptoms” often have more or stronger connections to other nodes, and targeting them in treatment may lead to wider improvements across the network [26, 7, 14]. Castro et al. (2024) further showed that centrality is not a fixed property of a symptom, but depends on the context and can shift over time. In idiographic networks, where symptoms are modelled within individuals, they found that the most effective treatment targets may change as the symptom structure evolves. Their simulation studies demonstrated that removing highly central symptoms, particularly those with high degree centrality, had a larger impact on the network than removing symptoms at random or based on strength centrality [12]. Importantly, centrality-based interventions were more effective when centrality was reassessed after each change, using so-called “cascade attacks”. These findings highlight the need for repeated and personalised assessment if centrality is to be used as a clinical decision tool [12].

Yet, the interpretation of centrality remains contested. Bringmann et al. (2019) argue that centrality measures such as closeness and betweenness rely on assumptions from social network theory, which may not hold in psychological networks [11]. Epskamp et al. (2018) also show that even node strength, a more straightforward measure, can vary strongly in small or noisy datasets. They emphasise the need for accuracy checks, such as bootstrapped confidence intervals and case-dropping analyses, to assess the stability of centrality estimates [17]. However, these metrics are still often reported. In a simulation, Epskamp et al. (2018) demonstrate how betweenness and closeness centrality appear unequal in an estimated network, even when all nodes are equal in the true network. This example illustrates how centrality differences can arise from estimation error rather than real structure [17].

Taken together, network models represent a promising but methodologically demanding approach to understanding depression. Their capacity to model symptom dynamics through time, detect potential tipping points, and identify treatment targets makes them valuable tools for both research and clinical innovation.

2.5. Multilevel Vector Autoregressive Modelling

A currently popular network estimation method is multilevel vector autoregressive (mlVAR) modelling. Developed specifically for the analysis of intensive longitudinal data, multilevel VAR provides a model for estimating how symptoms or behaviours influence each other over time, both at the individual and population level [10]. The following sections discuss the development of multilevel VAR, explain its relevance for psychiatric research, and highlight its strengths and limitations based on empirical work.

From Autoregression to Multilevel VAR

To capture how psychological processes unfold over time, researchers have increasingly turned to time-series models. One of the most basic models is the autoregressive (AR) model, which estimates a variable based on its own previous values. This idea can be extended to multiple variables using vector autoregressive (VAR) models, which allow each variable to be predicted by the lagged values of all other variables in the system. Such models are well suited to studying feedback loops and symptom interactions, but they require long, uninterrupted time series for a single person, something that is rarely available in psychological or psychiatric research [10]. To address this, Bringmann et al. (2013) introduced the Multilevel Vector Autoregressive (mlVAR) model. This method combines the structure of VAR with multilevel modelling, allowing researchers to estimate both population-level patterns (fixed effects) and person-specific deviations (random effects). The model was specifically developed for use with experience sampling data, where subjective experiences are collected from multiple individuals.

It is also well suited to digital phenotyping studies, where repeated behavioural objective data are gathered over time using smartphones or wearables [10, 19]. By accounting for both within-person and between-person variation, multilevel VAR provides a powerful tool for studying how symptoms and behaviours influence one another over time, even when individual time series are relatively short or unevenly spaced.

Why Multilevel VAR for Psychiatry?

The mlVAR approach is well aligned with the network perspective on psychopathology, as it explicitly models interactions between symptoms over time [8]. Multilevel VAR enables researchers to estimate three types of networks from repeated measurements: (1) temporal networks, which capture lagged predictive effects among symptoms; and (2) contemporaneous networks, which reflect same-time associations after controlling for temporal influences [19]. This makes multilevel VAR align perfectly with network theory to uncover how symptoms dynamically influence each other both within and across individuals.

Bringmann et al. (2013) first applied this approach to data from patients with residual depressive symptoms and found that negative emotions such as sadness and worry reinforced one another across time. These findings support the idea that depressive symptoms may form self-sustaining feedback loops, a notion that cannot be captured by unidimensional models [10]. Epskamp et al. (2018) expanded on this framework by formalising how multilevel VAR can be used to estimate the three types of networks simultaneously. Their work also established practical tools and guidelines for estimating and interpreting these models, helping to make multilevel VAR more accessible and reproducible for researchers [19].

Opportunities and Limitations of Multilevel VAR

Several studies have demonstrated that multilevel VAR can uncover meaningful patterns in the temporal dynamics of symptoms and emotions. For instance, Pe et al. (2015) [48] found that individuals with depression exhibited denser temporal networks of negative emotions. In these individuals, emotions such as sadness and anxiety were more strongly linked across time. This phenomenon, referred to as emotional inertia, reflects a reduced capacity to shift out of negative emotional states and may contribute to the persistence of depressive episodes. Kivelä et al. (2024) [37] applied a similar modelling approach to examine the day-to-day fluctuations in suicidal ideation within a clinical population. Their use of multilevel VAR enabled the identification of temporal and contemporaneous associations, discussing how certain mental states may activate or sustain one another throughout the day. Although the study focused on suicide risk rather than depression, it illustrates the potential of multilevel VAR to capture short-term psychological dynamics that are not easily detected using static approaches.

Despite its strengths, multilevel VAR relies on several assumptions that may not hold in practice. One key assumption is that relationships between variables remain stable over time (stationarity). In some datasets, this assumption is difficult to meet, as symptom dynamics may shift due to treatment, life events, or seasonal effects [19]. Multilevel VAR also assumes that observations are evenly spaced and frequent, which is often not the case in naturalistic studies with irregular sampling [10]. Furthermore, most current applications rely solely on self-report data, potentially missing behavioural or physiological signals that individuals do not consciously report [48]. Lastly, passive digital phenotyping data such as sleep, movement, and heart rate are rarely included currently, despite their potential to capture important behavioural dynamics related to depression symptoms [55].

In conclusion, multilevel VAR is a promising tool for understanding how depression unfolds in daily life. It can separate what is happening within one person from general patterns seen across people. This may help us better understand the dynamic nature of depression.

2.6. Research Gap

The growing emphasis on personalised mental health care has highlighted the need for monitoring tools that reflect how depression unfolds within individuals over time. Traditional assessment methods, such as self-reported questionnaires, provide only limited, one-moment information. These tools often miss symptom fluctuations and rely on general assumptions that do not account for the heterogeneity depression. To support earlier intervention and improve personalised monitoring, there is a clear need

for more dynamic, individualised approaches to depression monitoring.

In response to these limitations, the field has seen a rise in the development of digital phenotyping: the continuous, real-world tracking of behaviour and physiology using smartphones and wearable devices. This approach offers a promising alternative to static assessment by capturing objective signals such as mobility, sleep, and heart rate, also called digital biomarkers. Several studies have shown that behavioural patterns like increased homestay, irregular sleep, and disrupted circadian rhythms are associated with worsening depressive symptoms. However, most of these studies remain descriptive or predictive and do not account for dynamic feedback between behaviours.

To better understand the time-dependent structure of mental health, researchers have increasingly turned to network models. These models reflect symptoms or behaviours not as passive reflections of an underlying disorder, but as active components in a dynamic system. Temporal and contemporaneous networks allow researchers to estimate how symptoms or behaviours influence one another across time, offering new insight into the mechanisms that sustain or intensify depressive episodes. Still, many network studies rely on group-level cross-sectional data and fail to distinguish between individual and population-level dynamics.

Multilevel VAR modelling addresses this limitation by allowing researchers to estimate temporal symptom networks at both the individual and group level using intensive longitudinal data. A growing number of studies have used multilevel VAR to investigate symptom dynamics and emotional rigidity in depression, showing that negative emotions tend to persist and reinforce one another over time. However, most existing multilevel VAR applications rely solely on self-reported data collected over short time frames. Passive digital signals, despite their availability and clinical relevance, are rarely integrated into these models. No studies have used multilevel VAR to evaluate if depression is related to how behavioural signals influence each other over time [55]. Applying multilevel VAR to passive data may reveal how behavioural signals influence each other over time, supporting earlier symptom detection and personalised care.

This thesis aims to address this gap by applying multilevel VAR to a digital phenotyping dataset that comprises continuous passive monitoring data collected with wearables. It investigates how sleep, physical activity and heart rate interact with each other over time. It evaluates whether these interactions differ across different levels of depression symptoms. By doing so, the study seeks to advance our understanding of how different domains of functioning co-evolve during the course of depression. This approach not only builds on recent developments in digital phenotyping and network modelling, but also takes a step toward more personalised forms of mental health monitoring.

3

Methodological approach

Understanding how digital biomarkers interact over time requires a modelling approach that captures the temporal dependencies and individual variability in individuals with varying levels of depression. This chapter outlines the methodological framework used to analyse passive digital biomarker data. The rationale for selecting the multilevel VAR model is described, the steps taken to select relevant variables and construct networks are discussed, and how these networks were interpreted and evaluated is explained. By combining temporal and contemporaneous analyses, this approach aims to uncover dynamic patterns across domains in relation to depression.

3.1. Multilevel VAR Justification

To investigate dynamic relationships among passively collected digital biomarkers in depression, network models were estimated with multilevel VAR. This model is specifically suited to analyse intensive longitudinal data, where repeated observations are nested within individuals over time [10]. The multilevel structure allows for simultaneous estimation of fixed effects, representing average temporal dynamics across the population, and random effects, capturing person-specific differences from this average [19]. Multilevel VAR can thus reveal how dynamics operate both at the group level (the average person) and at the individual level (specific persons).

Multilevel VAR explicitly models autoregressive and cross-variable lagged effects. This makes it possible to detect feedback loops that are theorised to underlie many mental disorders [8, 10]. By regressing each variable on its own and others' lagged values, multilevel VAR can reveal how behavioural states such as sleep quality or physical activity influence each other from day to day.

Compared to 'black-box' machine learning models, such as random forests or neural networks, which are commonly used in digital phenotyping research [55], multilevel VAR has the advantage of producing transparent and interpretable relations. This interpretability is crucial for hypothesis-driven research in psychiatry, where understanding the direction and magnitude of dynamic effects is often more important than the exact value of such relationships [49, 39]. Moreover, unlike traditional time-series models that require long uninterrupted sequences from a single individual, multilevel VAR is designed for the relatively short and partially incomplete time series typically found in mental health studies [10, 19].

Finally, multilevel VAR aligns conceptually with the network theory on psychopathology, which conceptualizes mental disorders as dynamic systems of interacting symptoms and behaviours [8]. By modelling the within-person processes, multilevel VAR offers a powerful tool for exploring these complex dynamics using a large amount of passive data retrieved from wearables.

3.2. Node Selection

Nodes in the network represent variables from the dataset. This thesis selected variables from three domains of passively collected digital biomarkers: activity, sleep, and heart rate. These domains were chosen based on their relevance to depressive symptomatology. Each domain contained multiple vari-

ables. This study used a data-driven approach to select the most relevant variables per domain. The domains and the variables selected for main analysis are discussed below.

3.2.1. Domain Selection

Three domains were included: activity, sleep, and heart rate. These domains were selected based on previous research in the literature review (see Chapter 2) and reflect core behavioural and physiological processes implicated in depression. Previous research has identified consistent associations within these domains and depressive symptom severity. All data were collected passively using wearable devices and aggregated per day to ensure equidistant time points, allowing for consistent temporal modelling.

- **Activity:** Variables related to movement and intensity were included because reduced physical activity is significantly linked to the presence of depression [48].
- **Sleep:** Sleep disturbances such as reduced total sleep time, altered sleep stages, and increased time in bed are both symptoms and predictors of depressive episodes. Sleep variability has also been linked to depression symptoms and poorer treatment outcomes in previous researches [26].
- **Heart rate:** Heart rate and heart rate variability are indicators of autonomic nervous system function, which is often impaired in individuals with depression. Meta-analytic evidence shows that people with depression exhibit significantly lower heart rate variability compared to healthy controls, with more severe symptoms linked to further reductions in heart rate variability [36].

This study focused exclusively on three passively collected domains: activity, sleep, and heart rate. This choice reflects the aim to model objective behavioural and physiological dynamics without relying on self-report data. As a result, the networks show how the biomarkers change over time, but do not include mood or thoughts directly. This means that the results offer a limited but objective view of daily functioning in depression, without capturing subjective experience.

3.2.2. Variable Selection

Exploratory factor analysis (EFA) was used as a preparatory step before estimating network models. The goal was to select variables within domains by identifying groups of highly correlated variables and retaining only the most representative ones. This helps simplify the network structure and improves estimation stability by reducing multicollinearity. EFA is a statistical method that typically uncovers clusters of observed variables that share variance, which are assumed to reflect underlying hidden constructs [20]. This makes EFA more suitable for understanding psychological or behavioural structure [44, 23]. Although Principal Component Analysis (PCA) is often applied by default in statistical software, it is not ideal when the goal is to uncover meaningful constructs, because it includes both shared and unique variance. Therefore it may obscure the underlying constructs by capturing variance that is not theoretically meaningful [51]. For example, PCA might combine deep sleep and time in bed into one component, even though only deep sleep reflects true sleep quality.

In this study, EFA was used somewhat controversially, not to interpret latent factors, but to support variable selection. It served to identify clusters of related variables within each domain and select a single, representative variable per cluster. Since only one variable per factor is retained, it must capture a meaningful aspect of the domain. The latent factors themselves were not included in the network; only the highest-loading observed variable from each factor was used to ensure better interpretability of the final networks.

To determine how many factors to retain, two standard criteria were combined. First, the Kaiser criterion, which keeps all factors with an eigenvalue greater than 1.0. Second, the scree plot (Figure 3.1), where the point of steep drop-off, the “elbow”, signals the optimal number of factors. When both pointed to the same solution, this was taken as support for the chosen number of factors [51]. To facilitate the interpretation of the factor solution, varimax rotation was employed. This method, which assumes that the resulting factors are uncorrelated, helps produce a simple structure where each variable loads clearly on one factor [23, 51].

To check whether factor analysis was appropriate for the data, two tests were used: Bartlett’s test and the Kaiser-Meyer-Olkin (KMO) measure. Bartlett’s test checks if the variables are related enough to each other to form factors. A significant result means there is enough shared variance between

variables. The KMO measure indicates how suitable the data are for factor analysis overall, with values above 0.6 generally considered acceptable.

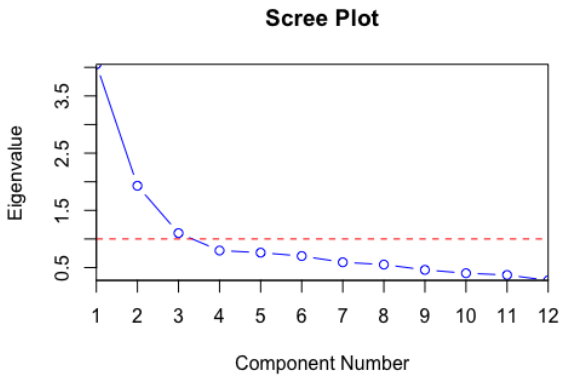


Figure 3.1: Illustrative scree plot with eigenvalue

Because the data consist of repeated measurements per individual, all variables were first averaged per participant to create a single-level dataset for factor analysis .

To improve the stability of the network estimation, redundancy within each domain was reduced by removing highly correlated variables. Keeping many similar variables can mislead results, such as inflated centrality estimates [17]. Although this reduction can lead to fewer and less strong connections in the final network, such trade-offs are common in network psychometrics to ensure models remain interpretable [10].

Although exploratory factor analysis (EFA) was used to reduce redundancy and select representative variables within each domain, some correlations between the selected variables can still appear in the network models. This is not necessarily a problem and can be explained in two ways. First, EFA was not used to include the latent factors themselves in the network, but only to identify which observed variables were most strongly associated with each factor. Only one observed variable per factor was retained. While the underlying factors are rotated to be statistically independent (orthogonal) through varimax rotation, the selected variables that represent them are not automatically completely uncorrelated. Some overlap between these variables may remain, even though the latent structure is orthogonal. Second, the factor analysis was performed on between-person average data, while the network models are based on time series data that capture within-person dynamics. It is possible that two variables do not show strong associations when comparing individuals on their averages, but still show meaningful connections within a person over time. For example, someone's heart rate and activity level may not correlate strongly across people on average, but may still influence each other day to day within individuals [10]. Therefore, it is not surprising if some connections between variables reappear in the network, even after using EFA for variable selection. This reflects a trade-off in network analysis: reducing redundancy improves interpretability and estimation, but may not remove all associations between selected variables. These remaining connections may still reflect meaningful dynamics and are kept in the analysis to capture relevant patterns.

3.3. Edge Selection

The selection of edges in the network models used in this study is a result of the chosen estimation model, the multilevel VAR model. This model estimates how variables predict each other over time (temporal effects) and within the same time window (contemporaneous effects). As such, the edges in both the temporal and contemporaneous networks are not selected manually, but follow from the coefficients estimated by the multilevel VAR procedure. In the next sections, the mathematical structure of the multilevel VAR model will be explained, and how its parameters are transformed into temporal and contemporaneous networks.

3.3.1. From Autoregressive to Multilevel VAR

To understand the multilevel VAR model, it is useful to build it up in three steps: starting with a basic autoregressive (AR) model for a single variable, then extending to a vector autoregressive (VAR) model for multiple variables, and finally incorporating a multilevel structure to allow for individual differences [10, 19].

Step 1: Autoregressive model (AR). The autoregressive model, is the simplest way to model change over time in a single variable. It assumes that the value of a variable at time t depends on its own value at the previous time point $t-1$. For example, for a variable y measured repeatedly for person p, we can write:

$$y_{t,p} = \mu_p + \phi_p \cdot (y_{t-1,p} - \mu_p) + \zeta_{t,p}$$

where:

- $y_{t,p} \in \mathbb{R}$ is the observed value of the variable at time t for person p,
- $\mu_p \in \mathbb{R}$ is the person-specific average (mean),
- $\phi_p \in \mathbb{R}$ is the autoregressive parameter, capturing how strongly the variable at $t-1$ predicts itself at t,
- $\zeta_{t,p} \sim \mathcal{N}(0, \sigma^2)$ is a normally distributed error term.

This model captures temporal dynamics for one variable only, and does not include interactions with other variables.

Step 2: Vector autoregressive model (VAR). The AR model looked at one variable over time. But in practice, we often want to study how multiple variables influence each other across time. For this, a vector autoregressive model (VAR) can be used. In the VAR model, no longer just one variable is predicted from its past, but instead predict the whole set of variables from their values at the previous time point. This is done using vectors and matrices. The model for person p becomes:

$$\mathbf{y}_{t,p} = \boldsymbol{\mu}_p + \mathbf{B}_p(\mathbf{y}_{t-1,p} - \boldsymbol{\mu}_p) + \boldsymbol{\zeta}_{t,p}$$

In this equation:

- $\mathbf{y}_{t,p} \in \mathbb{R}^K$ is a vector containing all K variables at time t,
- $\boldsymbol{\mu}_p$ is a vector of average values (means) for person p,
- $\mathbf{B}_p$ is a $K \times K$ matrix of regression coefficients: each element $b_{ij}^{(p)}$ tells us how variable j at time $t-1$ predicts variable i at time t,
- $\boldsymbol{\zeta}_{t,p}$ is a vector of error terms, assumed to be normally distributed.

Compared to the AR-model, the VAR-model adds complexity in two ways: (1) it includes multiple variables instead of one, and (2) it allows variables to influence each other across time. For example, not only can "sadness" predict future "sadness," but "sleep problems" at time $t-1$ can also predict "sadness" at time t. This makes the model especially useful when studying systems of interacting symptoms or behaviors over time.

Step 3: Multilevel VAR (mlVAR). The VAR model allows studying of how multiple variables influence each other over time, but it assumes that these dynamics are the same for every person. It was established in the Literature Review that depression is a heterogeneous disease, and thus, individual deviations must be incorporated in the model. To account for these individual differences, the VAR model is extended into a multilevel structure. This is known as the multilevel VAR model, introduced by Bringmann et al. [10].

The key idea is that each person has their own version of the temporal coefficient matrix $\mathbf{B}_p$, but that these personal matrices are variations around a shared population-average matrix $\mathbf{B}$. Formally, this is written as:

$$\mathbf{B}_p = \mathbf{B} + \mathbf{U}_p$$

where:

- $\mathbf{B}$ contains the average (fixed) effects across all people,
- $\mathbf{U}_p$ contains the random deviations for person p, assumed to follow a normal distribution.

This means that instead of estimating a single network for everyone, multilevel VAR estimates a population-level network and person-specific variations. This is especially useful when time series per person are short: by combining information across individuals, the model becomes more stable while still allowing for individual nuance.

In practice, each variable is modelled using a mixed-effects regression: the value of one variable at time t is regressed on the lagged values of all variables at time $t-1$, while allowing the strength of these effects to vary by person. For variable j, this results in the equation:

$$y_{t,p}^{(j)} = \beta_0^{(j)} + \sum_{k=1}^{K} \left(b_{kj}^{\text{fixed}} + b_{kpj}^{\text{random}} \right) \cdot y_{t-1,p}^{(k)} + \epsilon_{t,p}^{(j)}$$

In this equation:

- $\mathbf{y_{t,p}^{(j)}}$ is the observed value of variable j for person p at time t,
- $\boldsymbol{\beta_0^{(j)}}$ is the intercept term for variable j, representing its average value when all predictors are centered,
- The summation iterates over all K variables k, where for each predictor $\mathbf{y_{t-1,p}^{(k)}}$ (i.e., variable k at the previous time point), the model includes:
  - $\mathbf{b_{kj}^{\text{fixed}}}$, the fixed effect capturing the average influence of variable k on variable j across the entire sample,
  - $\mathbf{b_{kpj}^{\text{random}}}$, the random effect capturing how individual p's relationship between k and j differs from the group-level average,
- $\boldsymbol{\epsilon_{t,p}^{(j)}}$ is the residual (error term) for variable j, assumed to be normally distributed with mean 0 and constant variance.

In sum, compared to the standard VAR model, multilevel VAR adds a second layer: it captures not just group-level patterns, but also how individuals differ from these patterns. This makes it highly suited for psychological or clinical data where person-specific dynamics are of interest.

3.3.2. Estimating Temporal and Contemporaneous Networks

As Epskamp et al. [19] further explains, multilevel VAR enables the construction of two important types of networks:

- The *temporal network*, based on the fixed lagged effects $\mathbf{B}$, capturing directed connections over time.
- The *contemporaneous network*, based on the residual covariance structure $\boldsymbol{\Sigma}$, capturing undirected associations at the same time point.

Together, these networks offer a detailed picture of how symptoms or behaviors influence each other over time and within the same moment, both on average and per individual. The distinction between them lies in what part of the model they are derived from and what type of relationships they represent: delayed versus simultaneous, directional versus undirected.

Temporal network. The temporal network is constructed by estimating, for each variable j, how its value at time t is predicted by all variables at the previous time point $t-1$, while allowing these effects to vary across individuals. This is implemented in multilevel VAR through a series of mixed-effects regression models. For variable j, the model takes the following form:

$$y_{t,p}^{(j)} = \beta_0^{(j)} + \sum_{k=1}^{K} \left(b_{kj}^{\text{fixed}} + b_{kpj}^{\text{random}} \right) \cdot y_{t-1,p}^{(k)} + \epsilon_{t,p}^{(j)}$$

In this equation, each lagged variable $y_{t-1,p}^{(k)}$ is used to predict the current value of variable j, with a fixed effect b_{kj}^{fixed} representing the average group-level influence and a random effect b_{kpj}^{random} capturing person-specific deviations. The matrix of fixed effects $\mathbf{B} \in \mathbb{R}^{K \times K}$, composed of the b_{kj}^{fixed} terms, forms the basis of the group-level temporal network.

Each element b_{kj}^{fixed} reflects the directed influence of variable k at time $t-1$ on variable j at time t, controlling for all other variables in the system. This results in a *directed network*, where an edge from node k to node j indicates that changes in k are predictive of future changes in j [19].

Contemporaneous network. After fitting the temporal model, we examine the residuals $\boldsymbol{\zeta}_{t,p}$, which reflect what remains unexplained at time t after accounting for all lagged influences. These residuals are assumed to follow a multivariate normal distribution:

$$\boldsymbol{\zeta}_{t,p} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$$

The covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{K \times K}$ describes how variables covary within the same time point, after removing the influence of previous time points. To obtain unique associations between variable pairs, the matrix $\boldsymbol{\Sigma}$ is inverted and standardized to compute a matrix of partial correlations. This results in a contemporaneous network, an undirected graph where edges indicate instantaneous statistical associations that are not due to past values or shared effects of other variables [19].

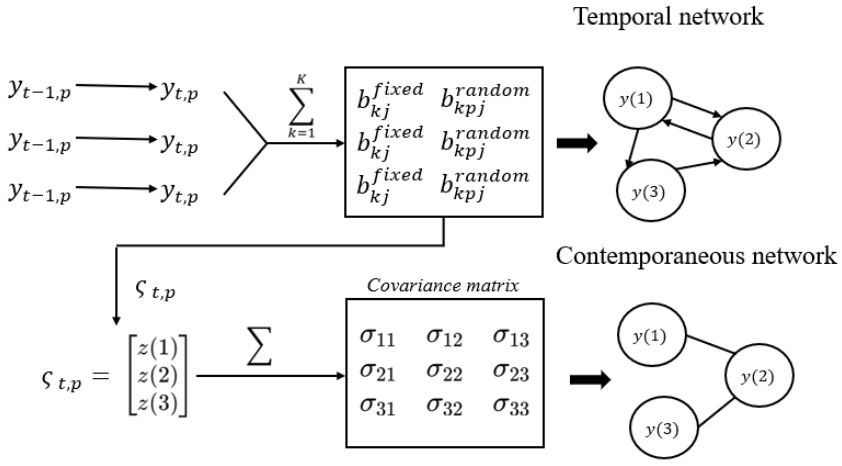


Figure 3.2: Schematic overview of multilevel VAR to temporal and contemporaneous networks

These two networks complement each other and together help us understand both how symptoms follow each other over time and how they appear at the same time. See Figure 3.2 for a schematic overview of how the multilevel VAR transforms into both networks.

3.4. Model Assumptions

Before estimating a multilevel VAR model, it is crucial to assess whether the data meet certain foundational assumptions. This section outlines the methodological handling of three key assumptions: equidistance, stationarity, and multivariate normality [10].

3.4.1. Equidistance

Equidistance refers to the assumption that the time intervals between successive measurements are equal. This ensures that time-based relationships in the model are measured over equal steps, which is important to interpret patterns correctly [10]. In this study, passive sensor data were aggregated per calendar day, with each day treated as a single time point. This daily aggregation guarantees equal distance between measurements. Although some days may be missing due to non-wear or technical issues, such missingness does not violate the equidistance assumption because the intervals themselves remain fixed. Therefore, temporal modelling can proceed under the assumption of equidistant data.

3.4.2. Stationarity

Stationarity is a key assumption in time series analysis and multilevel VAR modelling. It requires that a time series' statistical properties, such as its mean, variance, and autocorrelation, do not change over time. Violations of stationarity can lead to biased estimates, especially in the context of temporal effects [37]. If the data are not stationary, the model might wrongly estimate how one time point influences the next, leading to misleading results. To evaluate stationarity, the Augmented Dickey-Fuller (ADF) test is applied to each variable within each participant's time series to keep a multilevel structure. This test examines whether the series contains a unit root (indicator of non-stationarity) by testing the null hypothesis that a unit root is present. The ADF test is chosen above other tests for its reliability in autoregressive settings, which is the case in this study [56].

The Akaike Information Criterion (AIC) helps determine how many lagged observations (lags) to include in the ADF test. This is important because using too few lags may leave residual autocorrelation, which means that there are patterns in the data that are not captured by the model. By contrast, including too many lags adds unnecessary complexity and can reduce statistical power. The AIC addresses this trade-off by balancing model fit and parsimony. It does so by penalizing models with more parameters, using the following formula:

$$\text{AIC} = -2\ln(\hat{L}) + 2k$$

where $\hat{L}$ is the maximum likelihood of the model (a measure of how well the model fits the data) and k is the number of estimated parameters. Lower AIC values indicate a better-fitting model that avoids overfitting [2]. Participants whose time series fail the stationarity test across variables are excluded from the final sample. This approach ensures that the data used in the multilevel VAR model meet the required assumptions for valid estimation.

3.4.3. Multivariate Normality

Although multivariate normality is not a strict requirement for multilevel VAR estimation, evaluating the distribution of input variables helps to identify potential issues that may affect model performance or interpretation. To assess this assumption, univariate distributions are examined using the Kolmogorov–Smirnov (KS) test. This test evaluates the null hypothesis that a sample comes from a normal distribution [10]. Such deviations are common in intensive longitudinal and sensor-based data and are not inherently problematic for multilevel VAR modelling [37].

3.5. Network Interpretation

The focus of interpretation lies in understanding what the edges in each network represent, and how they can inform hypotheses about underlying mechanisms in depression. Rather than reflecting definitive causal pathways, these connections highlight associations that suggest which variables might potentially influence each other under real-world conditions [18].

Temporal interpretation In the temporal network, edges represent directed predictive effects from one variable to another across time. A directed edge from variable A to variable B indicates that deviations in A on day t are followed by changes in B on day $t+1$, controlling for all other variables. The edge sign reflects the direction of the association (positive or negative), and its magnitude indicates the strength of prediction. In addition to cross-variable predictions, temporal networks also contain autoregressive effects, which are edges from a node back to itself (self-loops). These loops represent the extent to which a variable is stable or persistent over time. These self-connections capture inertia: the tendency for states such as physical activity or heart rate to remain similar from one day to the next [18].

Temporal edges are particularly informative for identifying potential feedback loops or cascading effects between behavioural domains. However, these effects are not causal in a strict sense. They reflect statistical dependencies with a lagged structure but cannot account for unmeasured confounders or external influences [18].

Contemporaneous interpretation Edges in the contemporaneous network indicate partial correlations between variables measured on the same day, after removing all lagged influences. These associations are undirected and represent co-fluctuations that occur within the same time window, independent of previous states.

Such edges may reflect shared latent causes or residual coupling between variables. For example, two physiological signals might co-activate within a day due to common stress responses, even if no lagged effect exists. Although contemporaneous links do not imply directionality, they are still valuable for identifying clusters of synchronously active biomarkers that may represent subsystems [18].

3.6. Centrality Indices

To understand which behavioural and physiological signals may play a key role in the daily functioning of individuals with depression, this study examined centrality indices from network analysis. These metrics help identify which variables are most connected within the network. This aligns with the study's aim to explore how behavioural and physiological signals interact over time in individuals with varying levels of depression. By identifying which variables are most strongly linked to others within daily patterns, the analysis seeks to uncover key features that may drive or sustain depressive states. This can provide insight into how different processes in depression co-occur and reinforce each other, and which signals may serve as early indicators of change.

Three centrality metrics were calculated: strength, closeness, and betweenness. These are commonly used in psychological network research and each provides a different way of looking at node importance [18].

- **Strength:** the sum of absolute edge weights connected to a node. In temporal networks, this includes both incoming and outgoing predictive effects; in contemporaneous networks, it reflects total partial correlations.
- **Closeness:** the inverse of the average shortest path from a node to all others, indicating how efficiently a node can influence or be influenced by the rest of the network.
- **Betweenness:** the number of times a node lies on the shortest path between two others, highlighting its potential role as a bridge in the network.

Among these, strength is considered the most interpretable metric for psychological networks. Especially in the context of depression, highly central variables in terms of strength may represent key behavioural or physiological patterns that are tightly linked to other symptoms or markers. Identifying such nodes may help understand which features are most influential in the daily symptom dynamics of depression [11]. Closeness and betweenness were originally developed for flow-based systems such as social or transport networks, and their application to psychological data has been widely debated. Following recommendations [16, 11], strength centrality is the primary focus in this study, as it directly reflects local connectivity. Closeness and betweenness are reported for completeness, but are interpreted with restraint. The motivation for still including them is as follows; betweenness can highlight nodes that function as bridges between otherwise weakly connected symptom clusters. These bridging

nodes may be clinically important because they help explain how symptoms co-activate and spread across domains [16]. Closeness, while more abstract, can offer insight into how quickly a node could potentially affect the rest of the network. In the context of depression, a symptom or behavioral signal with high closeness may reflect a process that is indirectly linked to many other elements of the depressive state, and could therefore be a sensitive target for early intervention. Although both measures are more difficult to interpret than strength, their patterns may still offer exploratory insight. It would be unfortunate to miss potentially relevant structural patterns by excluding these metrics entirely; their inclusion supports a more complete and transparent representation of the network.

3.7. Connectivity Indices

Network connectivity is defined as the sum of the absolute edge weights in a network [16]. This index provides a single measure of how strongly all variables in the network are connected to one another. Higher network connectivity indicates that the digital biomarkers (or other variables) are more tightly coupled, which may reflect more rigid, self-reinforcing dynamics within the system. In contrast, lower connectivity suggests more loosely connected elements, potentially indicating greater flexibility or symptom independence. In the context of depression, higher connectivity has been associated with more severe and persistent symptom states, while lower connectivity may reflect resilience or recovery [14, 10]. As such, global connectivity is often used as a coarse indicator of systemic vulnerability or rigidity in psychopathological networks.

Connectivity was calculated separately for the temporal and contemporaneous networks. These values were then compared across symptom groups to explore if the overall structure and interactions strength of these networks depends on symptom severity. . It is important to note that connectivity is an aggregated metric and does not provide information about the direction or content of individual associations.

3.8. Symptom-Group Selection

To investigate whether network structure varies as a function of symptom severity, participants were classified into three groups based on their symptom status at repeated assessment points. Symptom status was determined using a combination of diagnostic and self-reported indicators, and categorized into the following levels:

- **Symptoms**: participants reported clinically significant depressive symptoms.
- **Mild symptoms**: participants reported a moderate level of symptoms.
- **No symptoms**: participants reported few or no depressive symptoms.

Separate multilevel VAR models were estimated for each symptom group using an identical variable set and model specifications. Importantly, symptom status could vary over time within individuals, meaning that a single participant could contribute data to more than one symptom group depending on their symptom trajectory. For example, a participant might begin the study with no symptoms, later develop mild symptoms, and eventually meet criteria for a symptoms state. Rather than assigning participants to only one group, this design allows network models to reflect the dynamic nature of symptom change and provides a more fine-grained understanding of symptom-specific network structures.

To ensure sufficient data per group, participants with fewer than 20 valid daily observations during a specific symptom state were excluded from that group's model. This threshold was used to support reliable network estimation.

Figure 3.3 illustrates this approach using two fictitious participants. Each line graph represents how a participant's symptom status evolved over time. Red arrows highlight which symptom segments were included in each network model. The lower panel shows simplified examples of the three resulting networks, estimated separately for data segments where participants were classified as symptoms, mild symptoms, or no symptoms. This visualization underscores how individual data can contribute flexibly to different network models, depending on the observed symptom patterns.

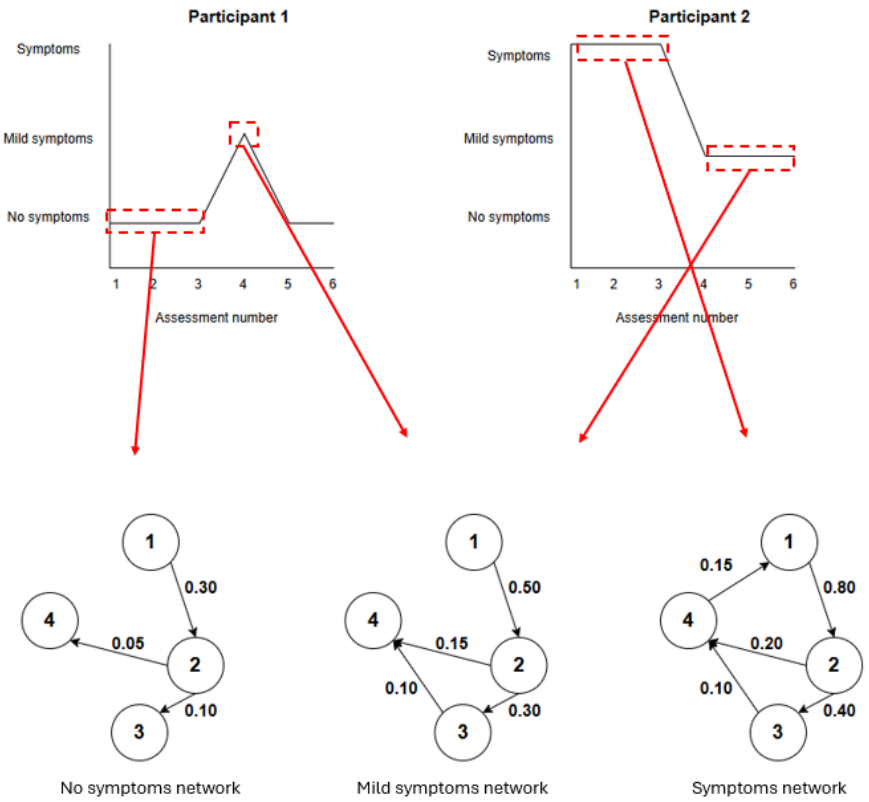


Figure 3.3: Visualisation of symptom-groups into networks

3.9. Software

All analyses were conducted using a combination of R (version 4.3.1) and Python (version 3.11.5). Data cleaning, and the assessment of model assumptions were primarily carried out in Python, using the `pandas` and `numpy` libraries. Exploratory factor analysis was performed in Python using the `factor_analyzer` package. For the estimation of multilevel VAR models, the `mlVAR` package in R was employed. Network visualizations and the calculation of centrality indices were done using the `qgraph` package. Additional data processing and plotting in R were conducted with the `tidyverse` collection of packages.

4

Data

This thesis analyzed a digital phenotyping dataset comprising wearable-based assessments of heart rate, sleep, and physical activity in individuals with varying levels of depressive symptoms. Participants were followed over an extended period, and daily summary data were collected through passive wearable devices. Each domain included multiple behavioural and physiological variables sampled on a daily timescale. In addition to wearable-based data, participants provided self-reports of depression symptoms at several time points during the study. These were used to classify individuals into three groups based on severity: 'no symptoms', 'mild symptoms', or 'symptoms'.

Rather than working with raw sensor streams, this thesis used pre-processed daily summary variables. No custom cleaning pipelines were applied at the raw data level. The variables retained reflect general and interpretable aspects of daily functioning, such as levels of physical activity, average heart rate, and sleep characteristics.

This chapter describes the process of preparing the dataset for network analysis. It includes an exploration of the data structure and quality, a description of how missing values and outliers were handled, and an explanation of the steps taken to select variables and check statistical assumptions. These preparations form the basis for the network modelling described in later chapters.

4.1. Data exploration

This section provides an overview of the initial assessment of data quality and structure. It focuses on identifying missing values, outliers, and other irregularities that could influence subsequent modelling steps.

4.1.1. Missing Values

All three datasets were systematically explored to assess data completeness. Since the number and nature of missing values varied across domains, a general rule was applied: variables with more than 20% missing data points were excluded from further analysis. This decision was made to ensure interpretability and reduce bias, following common practices in similar behavioral data research [59]. After applying this threshold, the number of retained variables per domain became more comparable in size. In this step, variables that primarily reflected derived or secondary statistical properties (such as percentiles or summaries) were also excluded.

4.1.2. Distribution of Variables

To assess the structure and plausibility of the data, the retained variables were explored in more detail. This phase focused on understanding and identifying irregularities such as outliers, extreme values, or implausible distributions that might reflect device errors. These explorations were done separately for each domain: sleep, heart rate, and physical activity. In this step, several irregularities were observed that required attention in the next phase of data preparation. For example, unusually high or low values sometimes appeared that did not align with typical human physiology, suggesting potential issues with

device functioning or wear time. Other signals, such as missing periods or unexpected peaks in activity or heart rate, also stood out. These issues were not immediately removed but flagged for further inspection and cleaning in the next section. Because the data were initially collected using passive sensors under real-world conditions, it was expected that such deviations would occur. The goal of this phase was not to filter or transform the data yet, but to build a clear picture of its structure, highlight potential problems, and inform the preprocessing pipeline.

4.2. Data preparation

This section describes the steps taken to prepare the dataset for modelling. It includes handling missing values, dealing with outliers, and selecting the variables that will be used. These steps are important to make sure the data is reliable and the results are not biased.

4.2.1. Outlier Detection and Treatment

To prepare the dataset for analysis, extreme values were carefully examined and adjusted where necessary. In digital phenotyping, particularly when using passive sensors, it is common to encounter unusual values. These can reflect genuine but rare behavioural patterns, but can also result from technical issues like device errors or improper use. Especially in the context of mental health research, it is important to distinguish between extreme but meaningful behaviour and measurement error. Rare behaviour should not be removed too quickly, as this could lead to the loss of valuable information. At the same time, leaving clear artefacts untreated can distort the results. The goal of this step was therefore to apply a method that protects the integrity of the data, while reducing the influence of values that fall outside the boundaries of what is realistically possible.

To separate true behavioural variation from measurement error, a two-step approach was used. First, implausible values were flagged based on predefined thresholds. These thresholds were derived from what is physiologically or technically feasible. Such threshold-based filtering is common in digital phenotyping research to flag implausible sensor data, as also applied in previous studies [21]. Examples include activity or heart rate values that exceed known human limits or device specifications.

Second, the values flagged in the first step were not removed from the dataset, but Winsorized. This means that only the values identified as implausible were adjusted by replacing them with the nearest defined boundary based on the interquartile range (IQR). The IQR bounds were calculated as:

$$\text{Lower bound} = Q_1 - 1.5 \times \text{IQR}, \quad \text{Upper bound} = Q_3 + 1.5 \times \text{IQR}$$

The IQR bounds were not used to detect outliers, but only to define a realistic upper and lower limit for replacement. This ensures that the adjusted values remain within the natural range of the data, without introducing artificial cut-offs. Values that fell within the predefined thresholds were left unchanged. This approach follows recommendations in the literature, which warns against automatic outlier removal due to the risk of inflated Type I errors. This means that removing outliers without careful consideration can lead to false-positive results, where the analysis suggests a significant effect that isn't truly there, especially in non-normal data [34, 4]. Winsorizing preserves the natural variance of the dataset and is therefore appropriate for time-series data. For Winsorization to make sense, the IQR-based replacement values should fall within the predefined thresholds of what is physiologically possible. This was the case here, which made it reasonable to use the IQR bounds as replacement values. The strength of this approach is that it combines domain knowledge with statistical methods, making outlier handling more grounded. At the same time, this approach comes with trade-offs. Some rare but real behaviours may be adjusted, and values close to the threshold can still shift the interpretation slightly. Since IQR values are based on the sample distribution, applying the same method to a different dataset might give slightly different results.

4.2.2. Handling Missing Values

To improve data quality, a number of steps were taken to deal with missing values. First, variables with high proportions of missingness had already been removed earlier in the process (see Section 5.1.2). In the next step, participants who had no data at all for one or more of the remaining variables were excluded. This choice was made because imputing entire variables for an individual would result in

fully artificial values. These values would carry no real information and could introduce serious bias. After removing these cases, the remaining dataset still contained some missingness across variables.

Therefore, several imputation strategies were explored. One option was to use model-based approaches such as K-nearest neighbor or random forest imputation [30]. These methods use patterns in the data to estimate missing values and can offer high accuracy in some cases. However, because this thesis is focused on the relationships between variables, using models to fill in data could create artificial associations that would not exist otherwise. This would make the results less trustworthy. As an alternative, simpler methods such as mean or median imputation were considered. These do not rely on modelling but have their own drawbacks. In this dataset, they led to a large number of repeated values, which flattened out individual differences and reduced the richness of the time-series data. Since variation between observations is a central part of behavioural and physiological signals, this loss of variability was undesirable. In the end, no imputation method was used. All rows that contained missing values were removed from the final dataset. This meant losing a part of the data, but it ensured that the remaining values were entirely based on real observations. This approach helps preserve the integrity of the analysis and supports more reliable interpretation of the results.

4.2.3. Variable Selection

To reduce redundancy and simplify the structure of the final models, exploratory factor analysis (EFA) was performed separately within each domain. This technique identifies clusters of highly correlated variables, from which one representative can be selected. Selecting a limited set of features in this way helps maintain interpretability and reduces the risk of inflated centrality estimates caused by multicollinearity [17]. Before applying factor analysis, its suitability was assessed using Bartlett's test of sphericity and the Kaiser-Meyer-Olkin (KMO) measure. Bartlett's test was significant across all three domains, indicating that the variables shared enough variance to justify factor extraction. The KMO values were acceptable, with strong adequacy for the activity domain and moderate adequacy for heart rate and sleep. Together, these results confirmed that the data were appropriate for exploratory factor analysis.

To determine the number of factors to retain in each domain, Kaiser's criterion (eigenvalues > 1) and visual inspection of the scree plot are used. These criteria suggested two clear factors in both the heart rate and step domains, and three factors in the sleep domain. Although an additional component appeared in the sleep domain, it was not retained due to limited theoretical interpretability.

Each retained factor was then interpreted based on the pattern of variable loadings. In the activity domain, one factor reflected the pace and intensity of movement, while the second captured the frequency of activity. In the heart rate domain, the first factor represented average heart rate, and the second captured fluctuations and extremes. In the sleep domain, one factor corresponded to sleep continuity, another to the composition of sleep stages, and a third to overall duration.

From these factors, one variable per factor was selected for the final model. Selection was based on the strength of loading and minimal overlap with other indicators. The result was a compact set of seven variables representing distinct behavioural dimensions: activity intensity (the extent to which an individual is physically active on a day), activity events (the extent to which physical activity is distributed across the day), heart rate (the average number of heart beats per minute), heart rate variability (the extent to which heart beat frequency changes), sleep effectivity (the extent to which an individual sleeps relative to being awake), time spent in NREM sleep phases (the extent to which an individual does not sleep in REM sleep phase), and total sleep time.

To check for residual multicollinearity at the within-person level, Pearson correlations were calculated separately for each participant. For each pair of selected variables, the proportion of participants with a correlation exceeding ($r > 0.6$) was computed. This was rare: no pair exceeded this threshold in more than 1% of cases. This suggests that the final variable set was sufficiently distinct to overestimation of relationships in multilevel VAR estimation.

This final variable set was used as input for all network models in this thesis.

4.3. Descriptive overview of the final dataset

After completing all preprocessing steps, the final dataset was reviewed to confirm that the retained data were consistent with expected behavioural and physical patterns. The selected variables showed realistic variation across participants, without extreme distortions, suggesting that data quality was preserved while device-related errors were minimised.

For physical activity, activity intensity showed plausible differences in movement levels between individuals. Activity events captured variation in how activity was distributed throughout the day, with some participants showing frequent short bursts and others displaying longer periods of movement.

Sleep variables also aligned with typical ranges. Total sleep time and sleep effectivity reflected nightly sleep duration and sleep quality, respectively, and showed expected variation between individuals. NREM sleep percentage captured time spent in non-REM sleep stages and varied within a plausible physiological range.

Heart rate indicators showed stable but individualized patterns. Heart rate displayed realistic daily averages, while heart rate variability reflected fluctuations that were both physically legitimate.

Together, these observations confirmed that the final variable set retained meaningful variation while excluding artefactual or implausible values, providing a solid foundation for network modelling.

4.3.1. Answer to Subquestion 1

What is the structure of the digital phenotyping data, and how can it be preprocessed and reduced for analysis?

This thesis integrates digital phenotyping data from three behavioural domains: steps, sleep, and heart rate. Each dataset was first explored separately to assess its structure, including descriptive characteristics, patterns of missingness, and potential outliers. Variables with more than 20% missing values were excluded, and clear outliers were addressed through a two-step procedure: domain-specific thresholds flagged implausible values, which were subsequently Winsorized using interquartile range (IQR) bounds. Participants with entirely missing data for one or more core variables were removed. No imputation was applied, in order to preserve the natural variation in the data and avoid introducing artificial correlations. Only complete cases were retained for analysis.

To reduce redundancy and improve interpretability, exploratory factor analysis (EFA) with varimax rotation was conducted separately for each domain to make a selection of variables. This allowed for the identification of latent constructs, with one representative variable selected per factor. The resulting set of seven variables, two for steps, two for heart rate, and three for sleep, offered a concise representation of the key behavioural features of the different domains.

4.4. Model Assumptions

To ensure that the dataset was suitable for multilevel VAR modelling, several assumptions were checked during preprocessing. This section briefly outlines how key requirements were addressed.

4.4.1. Equidistance

The data were aggregated at the daily level across all domains, which ensured a consistent time interval between observations. While occasional days were missing, the underlying structure of the dataset reflected a regular 24-hour measurement rhythm. This means that the assumption of equidistant time points, required for multilevel VAR, was adequately met.

4.4.2. Stationarity

To satisfy the requirements of multilevel VAR modelling, stationarity was evaluated at the individual level. The Augmented Dickey-Fuller (ADF) test was applied per variable and per participant, using automatic lag selection based on the Akaike Information Criterion (AIC). Participants who consistently showed non-stationary patterns across one or more variables were excluded from further analysis. This decision was based on the principle that multilevel VAR assumes weak stationarity within each time series, meaning that the statistical properties of the signal (such as mean and variance) remain stable over time.

While this step reduced the overall sample size, the resulting dataset still contained a satisfactory number of participants and observations. The stationarity filtering step was therefore considered both necessary and acceptable to ensure the reliability of the estimated networks.

4.4.3. Normality

To check whether the distributions of the selected variables were approximately normal, a statistical test for univariate normality was applied. This test compares the shape of each variable's distribution to that of a theoretical normal distribution. As is not uncommon, all variables showed some degree of deviation from normality. This was expected and does not pose a problem for the analyses used in this thesis.

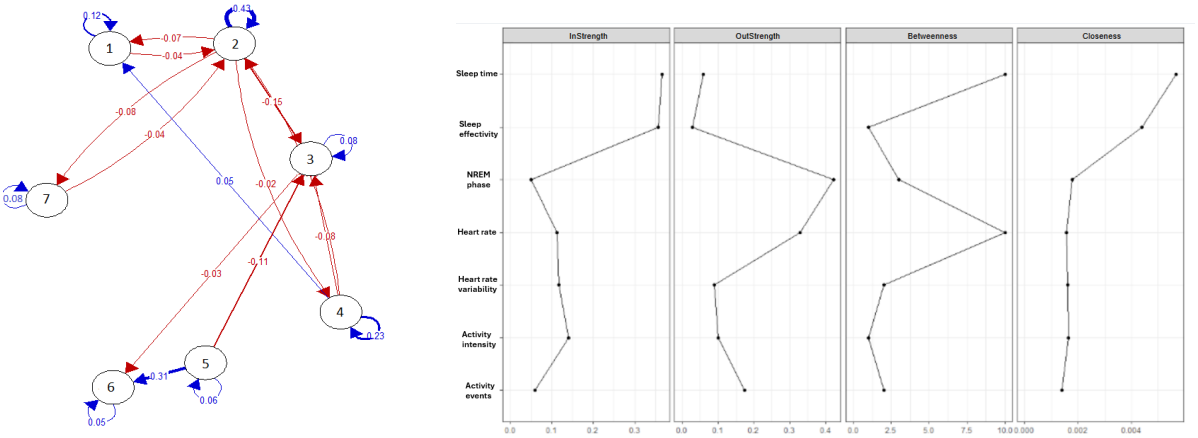
Multilevel VAR models are known to be relatively robust to non-normality, especially when the model structure is not overly complex [17, 37]. Because this thesis focuses on identifying broader patterns and associations, no data transformations were applied. This approach follows standard practice in the literature, where slight deviations from normality are typically tolerated. While non-normality might slightly reduce the sensitivity to detect very subtle associations, it does not affect the overall reliability or interpretability of the results. The networks produced in this study, therefore, highlight the most stable and meaningful connections between variables.

5

Results

This chapter presents the results of the network models. First, the general structure of the temporal and contemporaneous networks across the full sample is described. Also, centrality analysis is performed on these full samples. Next, these networks are examined separately across groups with different levels of symptom severity to explore how network dynamics vary. In each group, network structure and network connectivity are compared to assess shifts in dynamics and system flexibility.

5.1. Temporal Network: Whole Sample



(a) Temporal network whole group

(b) Centrality measures

Node	Variable	Node	Variable
1	Activity intensity	5	NREM phase
2	Heart rate	6	Sleep effectivity
3	Sleep time	7	Heart rate variability
4	Activity events		

Figure 5.1: Temporal network and strength metrics with corresponding variable legend

5.1.1. Interpretation: Temporal Network

Figure 5.1a shows the estimated temporal network for the full sample. Each directed edge represents a statistically significant lag-1 association between two variables, controlling for all others in the model. Blue edges indicate positive associations, while red edges indicate negative ones. All displayed effects met the $p < 0.05$ significance threshold.

Several strong autoregressive effects were found, reflecting temporal stability in some variables. Most notably, `heart rate` showed a strong autoregressive effect ($\beta = 0.43$, $p < .001$), indicating a high

degree of persistence: deviations in heart rate on a given day tend to continue into the next day. In practical terms, this means that if an individual's heart rate rises or falls relative to their baseline, it may take longer to return to that baseline. `Activity intensity`, a measure of physical activity, also showed moderate persistence over time ($\beta = 0.23$, $p < .001$), suggesting some day-to-day consistency in activity levels. In contrast, the sleep variables (`sleep time`, `NREM phase`, and `sleep effectivity`) had weak self-connections (all $\beta < 0.08$). These weaker autoregressive effects indicate that sleep patterns were more variable and less predictable based on the previous night, possibly due to greater influence from daily context or interactions with daytime behaviour.

The network also revealed meaningful connections between domains. A higher `heart rate` during the day was associated with reduced `sleep time` that night ($\beta = -0.15$, $p < .001$), suggesting that increased heart rate may negatively impact the sleep time that follows. Interestingly, longer `sleep time` predicted also a small reduction in `heart rate` on the following day ($\beta = -0.15$, $p < .001$), possibly indicating a restorative effect of sleep on cardiovascular functioning. Given the lagged data structure, this effect reflects how better sleep during the night after day $t-1$ is associated with a calmer heart rate on day $t+1$. This finding is consistent with previous research linking poor sleep with a more elevated heart rate [53].

`Activity intensity` showed several slight but notable effects. It was linked to a lower `heart rate` the next day ($\beta = -0.07$, $p = .004$), possibly reflecting improved cardiovascular recovery following active days. This finding aligns with earlier work suggesting that increased physical activity can trigger delayed nocturnal recovery of the autonomic nervous system, such as reduced heart rate or improved heart rate variability the following day [43]. `Activity events` also contributed to multiple downstream effects. It negatively predicted `sleep time` ($\beta = -0.078$, $p < .001$), suggesting that days with more fragmented activity may reduce overall sleep duration in the night that follows. Additionally, `activity events` predicted increased next-day `activity intensity` ($\beta = 0.046$, $p < .001$), showing temporal continuity in activity.

`Sleep time` was associated with reduced `sleep effectivity` the following night ($\beta = -0.032$, $p < .001$), suggesting that sleeping more is not always linked to better sleep quality. Longer sleep duration may reflect fragmented sleep. In contrast, more sleep spend in `NREM phase` predicted increased `sleep effectivity` the following night ($\beta = 0.31$, $p = .029$), indicating that deep, restorative sleep may improve subsequent sleep quality, which follows logical reasoning. Interestingly, more `NREM phase` sleep was also followed by less `total sleep` the next night ($\beta = -0.11$, $p = .000$). One possible explanation is that after a night of deep sleep, the body requires less recovery the following night, leading to shorter but sufficient sleep. This pattern suggests a self-regulating process in sleep dynamics, where sleep quality and quantity are continuously balanced by each other. This rebound effect is a known effect in literature [5].

Taken together, these findings indicate that `heart rate` functions as a stable and central signal in the daily physiological system, showing strong persistence and bidirectional links with sleep. Sleep variables were more variable and appear to respond sensitively to both heart rate and activity-related inputs. Physical activity indicators such as `activity intensity` and `activity events` showed cross-domain associations between sleep and heart rate systems. The overall structure of the temporal network points to interdependency across domains, in which changes in one system ripple through others.

5.1.2. Centrality Indices Temporal Network

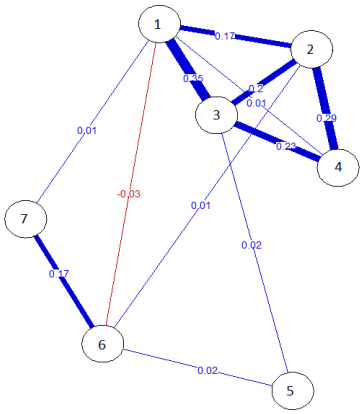
To understand the relative importance of each variable within the temporal network, three centrality metrics were examined: strength, closeness, and betweenness (Figure 5.1b). Among these, strength centrality is considered the most robust and directly interpretable in psychological networks [16].

`Sleep time` and `sleep effectivity` demonstrated the highest in-strength values, indicating that these sleep-related variables are strongly influenced by other components in the system. This suggests that fluctuations in other physiological or behavioural domains may shape sleep outcomes, rather than sleep acting as an initiating force. In contrast, variables such as sleep in `NREM phase` and `activity events` had relatively low in-strength, implying they are less affected by other processes. Regarding out-strength, sleep in `NREM phase` and `heart rate` showed the highest values. These variables ap-

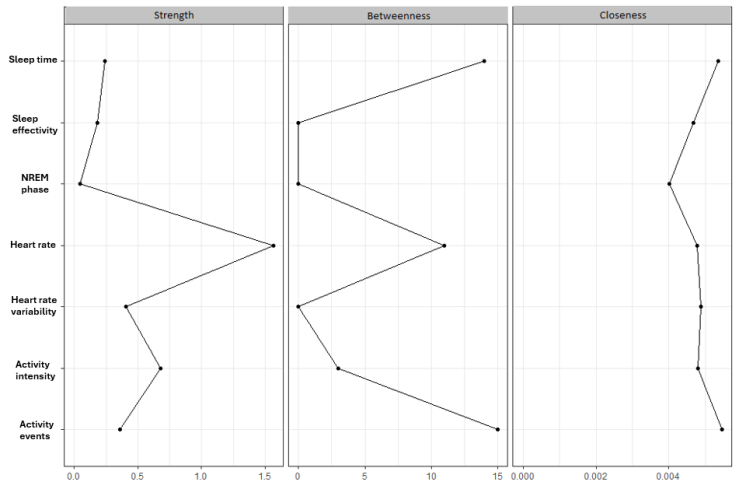
pear to influence other components, suggesting that they may function as important initiators within the system's daily dynamics. Interestingly, sleep in `NREM phase` stood out from the other sleep-related variables by showing a clear driving role, whereas `sleep time` and `sleep effectivity`, despite being central in the in-strength results, had lower out-strength values. This contrast implies that the composition of sleep, reflected in sleep architecture like NREM, may be more influential in shaping next-day functioning than the overall quantity or efficiency of sleep. In other words, it may not be how long or how efficiently one sleeps, but how restorative that sleep is, that drives change across domains.

Although closeness and betweenness centrality are considered less robust in psychological networks and interpreted with caution [16, 11], they were examined to provide additional context. Regarding betweenness centrality, both `sleep time` and `heart rate` frequently appeared on the shortest predictive paths between other variables. This implies a potential bridging role, enabling indirect interactions across otherwise weakly connected parts of the network. Closeness centrality followed a similar pattern: `sleep time` and `sleep effectivity` had the highest values, underscoring their integration within the network. These nodes are well-positioned to quickly transmit or receive influence.

5.2. Contemporaneous Network: Whole Sample



(a) Contemporaneous network whole group



(b) Contemporaneous node strength

Node	Variable
1	Activity events
2	Activity Intensity
3	Heart rate
4	Heart rate variability

Node	Variable
5	NREM phase
6	Sleep time
7	Sleep effectivity

Figure 5.2: Contemporaneous network and corresponding strength metrics with variable legend

5.2.1. Interpretation Contemporaneous Network

Figure 5.2a shows the contemporaneous network for the full sample, based on partial correlations between variables measured on the same day. All displayed edges are statistically significant ($p < 0.05$) and represent partial associations that control for all other variables in the system. Blue edges reflect positive relationships, and red edges indicate a negative one.

The strongest connection was found between `heart rate` (Node 3) and `activity events` (Node 1, $\rho = 0.345$), suggesting that increased physical activity is reliably accompanied by elevated heart rate. `Activity intensity` (Node 2) also showed moderate associations with both `activity events` ($\rho = 0.166$) and `heart rate` ($\rho = 0.196$), indicating a tight coupling between movement and cardiovascular signals within the same day.

Several associations in the contemporaneous network were very small and should be interpreted with caution, as they may represent false positives. For instance, `sleep effectivity` (Node 7) was only modestly associated with `sleep time` (Node 6, $\rho = 0.17$), which aligns with known links between sleep

quantity and quality [45]. However, most connections involving sleep were minimal. `Activity events`, a measure of activity fragmentation, showed only weak associations with sleep outcomes, including `sleep effectivity` ($\rho = 0.01$) and `sleep time` ($\rho = -0.03$). Similarly, the edge between `heart rate variability` (Node 4) and `activity events` (Node 1, $\rho = 0.013$) was close to the significance threshold and exhibited high variability. While these effects were statistically significant, their small magnitude and limited stability suggest they may not reflect robust or generalisable relationships.

Altogether, the contemporaneous network reveals clear clustering between physical activity and cardiovascular signals, with `heart rate` acting as a central component. In contrast, sleep-related variables play a more peripheral or bridging role, consistent with their delayed temporal alignment in the dataset.

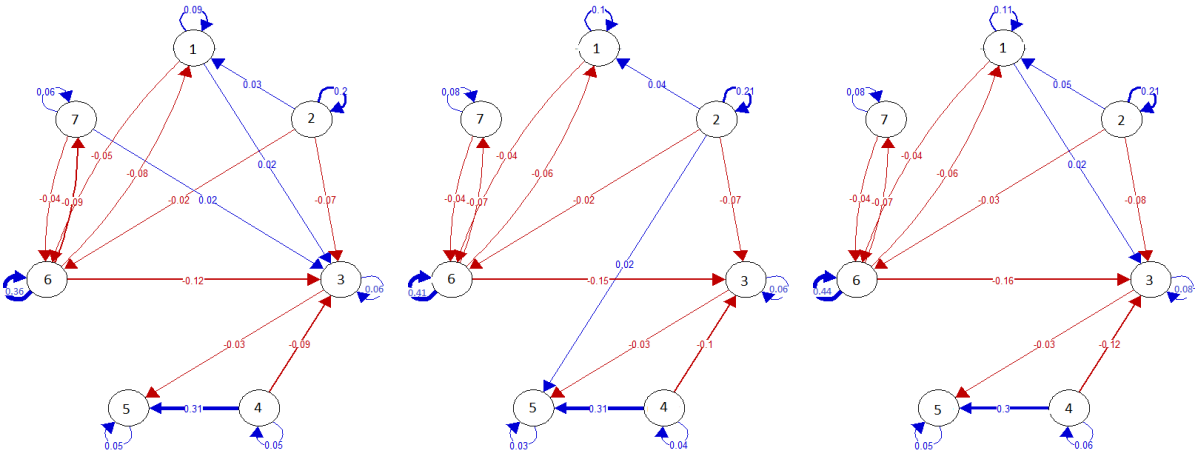
5.2.2. Centrality Indices Contemporaneous Network

Similarly to the temporal network, `heart rate` emerged as the most central node in the contemporaneous network. Its high strength indicates strong and consistent associations with other variables within the same day, suggesting that heart rate reflects ongoing physiological states that co-fluctuate with other domains in real-time. In contrast, sleep-related variables such as `sleep time`, sleep in `NREM phase`, and `sleep effectivity` showed lower centrality in this contemporaneous context.

Although interpreted with caution, betweenness centrality showed that `sleep time` and `activity events` frequently appeared on the shortest paths between other nodes. This pattern suggests a potential bridging role in facilitating indirect links between otherwise loosely connected variables. In contrast, variables such as `sleep effectivity`, sleep in `NREM phase`, and `heart rate variability` exhibited low betweenness values, indicating a more passive position in the network's structure.

Closeness centrality values were relatively uniform across variables but slightly elevated for `sleep time` and `activity events`. These results suggest that these nodes are more readily reachable from other parts of the network. Yet, as with betweenness, this metric is interpreted as descriptive rather than definitive.

5.3. Temporal Symptom Group Differences



(a) No symptoms **(b)** Mild symptoms **(c)** Symptoms

Figure 5.3: Temporal network plots by symptom group

Node	Variable	Node	Variable
1	Activity intensity	5	Sleep effectivity
2	Activity events	6	Heart rate
3	Sleep time	7	Heart rate variability
4	NREM phase		

5.3.1. Interpretation Temporal Symptom-Groups

Figure 5.3 displays the estimated temporal networks for participants grouped by symptom severity. Each plot shows significant lag-1 effects ($p < .05$) between the seven core variables. While the overall

structure is broadly consistent, several meaningful group differences emerge in the strength and pattern of temporal associations.

Across all groups, `heart rate` (Node 6) showed strong autoregressive effects, with the strongest self-loop in the symptoms group ($\beta = 0.44$), followed by the mild symptoms group ($\beta = 0.41$) and the no symptoms group ($\beta = 0.36$). This suggests greater temporal persistence in heart rate among individuals with more depressive symptoms, potentially reflecting lower physiological flexibility.

Differences were also observed in the pathway from `heart rate` to `sleep time` (Node 3). This negative association grew stronger with symptom severity, from $\beta = -0.12$ in the no symptoms group to $\beta = -0.16$ in the symptoms group. This could reflect elevated stress or arousal disrupting sleep in individuals with more symptoms [5].

Notably, in the no symptoms group, small but significant connections emerged from `activity intensity` (Node 1) and `sleep time` (Node 3) to `heart rate variability` (Node 7). These links were absent in the mild symptoms and symptoms groups. This pattern does not follow a clear trend and may be unstable, or sample-specific effects. Therefore, it should be interpreted with caution.

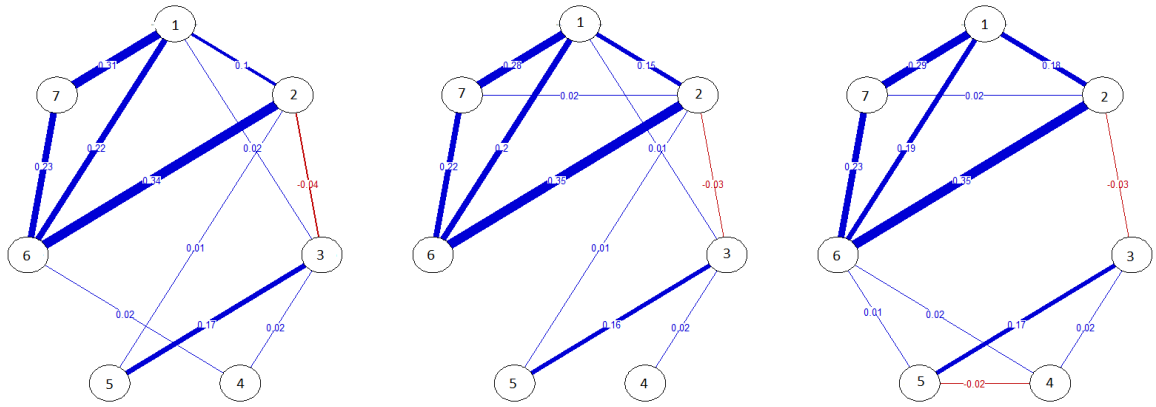
The effect of `heart rate` on next-day `heart rate variability` was strongest in the no symptoms group ($\beta = -0.09$), slightly weaker in the mild symptoms group ($\beta = -0.07$), and similarly reduced in the symptoms group. This suggests that in individuals with fewer symptoms, heart rate is a stronger predictor of how the autonomic system will regulate itself the following day. A similar pattern was reported by Hamilton et al. (2016), who found that higher mean heart rate predicted lower heart rate variability the next day. Crucially, this association was weaker in individuals with higher levels of depressive symptoms, supporting the idea that autonomic regulation becomes less responsive in the context of depression [28].

Lastly, the effect from sleep in `NREM phase` (Node 4) to `sleep time` (Node 3) became increasingly larger across groups ($\beta = -0.09$, -0.10, and -0.12 respectively). This indicates that when a larger share of sleep consists of deep NREM sleep, the person tends to sleep shorter the next night, especially in individuals with more symptoms. One possible interpretation is that higher NREM percentages reflect more efficient or restorative sleep, which may reduce the need for longer sleep duration the following night. This may reflect a rebound effect: after a night of increased deep sleep, individuals may experience shorter and more irregular sleep the following night. Similar rebound-like fluctuations in sleep duration and architecture have been observed in other studies on depression, pointing to instability in homeostatic sleep regulation [5].

5.3.2. Connectivity Indices Temporal Symptom-Groups

Network connectivity differed across symptom groups. The symptoms group showed the highest connectivity (2.02), suggesting that day-to-day changes in heart rate, sleep, and activity were more strongly linked. Participants with no symptoms showed moderately high connectivity (1.84), and the mild symptoms group had the lowest connectivity (1.79). One interpretation is that digital biomarkers become more interdependent when symptoms are more severe, reducing flexibility. In contrast, mild symptoms individuals may show less consistent coordination between systems, possibly reflecting a transition phase in which patterns are unstable. This interpretation is in line with previous studies using symptom networks, where higher connectivity has been found in individuals with more depressive symptoms [14, 10, 48]. In those studies, stronger connections between symptoms were thought to indicate a more rigid system, where changes in one symptom more easily affect others. The current results suggest that this principle may also apply to digital biomarkers, indicating that symptom severity is reflected not only in individual signals, but in the way physiological systems are connected.

5.4. Contemporaneous Symptom Group Differences



(a) No symptoms **(b)** Mild symptoms **(c)** Symptoms

Figure 5.4: Contemporaneous networks by symptom severity group.

Node	Variable	Node	Variable
1	Activity intensity	5	Sleep effectivity
2	Activity events	6	Heart rate
3	Sleep time	7	Heart rate variability
4	NREM phase		

5.4.1. Interpretation Contemporanous Symptom-Groups

Figure 5.4 displays the contemporaneous networks for the three symptom severity groups. All edges represent same-day partial correlations between variables, controlling for all others. While the core structure is largely preserved across groups, several shifts in edge strength and presence suggest subtle changes in physiological dynamics as symptom severity increases.

Several associations across groups were weak and should be interpreted with caution. For instance, the edge between `activity intensity` and `sleep effectivity` remained close to zero in all subgroups ($r < .05$), suggesting little to no immediate relation between physical activity fragmentation and perceived sleep quality. Similarly, associations between `heart rate variability` and other variables were small and inconsistent across groups, which may reflect either random variation or low effect sizes that lack practical relevance. These weak or unstable associations highlight the importance of focusing on robust network connections when drawing conclusions and raise the possibility of false positives due to the large number of estimated edges.

In summary, the contemporaneous network structure appears largely stable across symptom severity groups. While some minor shifts in edge strength were observed, most associations remained consistent in both direction and magnitude. This suggests that real-time relations between variables are relatively preserved, regardless of symptom severity. Unlike the temporal networks, which showed more clear differences in structure and coupling, the contemporaneous networks provide limited additional insight into group-level variation in depressive symptomatology.

5.4.2. Connectivity Indices Contemporaneous Symptom-Groups

Global connectivity in the contemporaneous networks showed relatively small differences across symptom groups. Connectivity was slightly higher in the symptoms group (1.53) compared to the no symptoms (1.48) and mild symptoms (1.45) groups. While this pattern is not as strong as the symptom-related differences found in the temporal networks, the small increase in contemporaneous connectivity among symptomatic individuals may still reflect subtle shifts toward tighter same-day coupling. This could indicate that with more symptoms, the system becomes slightly more reactive or interdependent, though the strength of change is modest.

Importantly, the relative consistency across groups implies that contemporaneous relationships are less

sensitive to symptom severity than day-to-day dynamics. Whereas temporal connectivity increased notably with symptom severity, contemporaneous connectivity remained comparatively similar. This difference suggests that the degree of symptoms primarily alters how digital biomarkers influence each other over time, rather than how they co-occur within a single day. Together, these findings underscore the added value of temporal modelling.

5.5. Answer to Subquestion 2

What temporal and contemporaneous dynamics, and centrality indices between digital biomarkers can be identified in individuals with varying levels of depression?

The temporal network showed that `heart rate` plays a central and stable role in daily physiological dynamics. It had the strongest autoregressive effect, meaning that heart rate tends to remain relatively consistent from day to day. It also influenced other domains. A higher heart rate predicted shorter sleep that night, while longer sleep was followed by a slightly lower heart rate the next day. These findings suggest that heart rate can interfere with recovery during sleep, while sleep may have a calming effect on the cardiovascular system. This aligns with existing evidence that elevated heart rate is associated with longer sleep latency and lower sleep quality, and that poor sleep can in turn worsen daytime cardiovascular functioning [57].

Physical activity also contributed to cross-domain effects. More activity events were followed by shorter sleep duration, and higher activity intensity predicted lower heart rate the next day. These effects were more minor but still indicate interaction between activity, sleep, and cardiovascular functioning. The sleep dynamics themselves exhibited a balancing pattern. More sleep in NREM phase predicted better sleep effectivity but less sleep time the next night, consistent with known rebound effects in sleep regulation established in models of sleep regulation, which states that sleep phases are homeostatically regulated and exhibit specific rebound effects after deprivation or satisfaction [47].

The centrality results give further insight into how these variables relate to the rest of the system. `Heart rate` and sleep in `NREM phase` had the highest out-strength, which means they influenced several other variables. `Sleep time` and `sleep effectivity` had the highest in-strength, meaning they were shaped by changes elsewhere in the system. `Sleep time` also had high betweenness and closeness values, indicating that it often connects otherwise separate parts of the network and plays a bridging role over time.

In the contemporaneous network, the strongest connections were found between `heart rate` and the activity variables `activity intensity` and `activity events`. This reflects the expected same-day coupling between movement and cardiovascular signals. Sleep variables were less central in this network.

Together, these findings show that heart rate and sleep in NREM phase are influential signal that connects different domains across days. Sleep duration and quality are more variable but play a role in linking the system over time. Activity shows both short-term synchrony and delayed effects on heart rate and sleep. Each domain contributes in different ways, and the structure of the networks helps to reveal which signals drive change and which respond to it.

5.6. Answer to Subquestion 3

How do digital biomarker dynamics and network connectivity differ across symptom subgroups?

The temporal networks showed differences across symptom subgroups. The strongest pattern was that overall connectivity increased with symptom severity. It was highest in the symptoms group, followed by the no symptoms group, and lowest in the mild symptoms group. This suggests that as symptoms become more severe, the coordination between digital biomarkers becomes tighter. A more strongly connected network may reflect a system that is less flexible, where changes in one domain are more likely to affect others. The lower connectivity in the mild symptoms group could reflect a transitional or unstable phase, where patterns are not yet settled. This matches earlier findings from symptom network studies, where higher symptom severity is linked to higher connectivity and reduced flexibility in the system [14, 10, 48].

Within the temporal networks, several group-specific patterns emerged. The autoregressive effect of mean heart rate was strongest in the symptoms group, indicating that heart rate becomes more persistent and possibly harder to regulate. The negative effect of heart rate on next-day sleep duration also became stronger with higher symptom severity, suggesting that elevated heart rate more strongly disrupts sleep in individuals with symptoms. Meanwhile, the association between heart rate and next-day heart rate variability weakened with symptom severity. In the no symptoms group, higher heart rate predicted lower variability the next day, but this link disappeared in the symptoms group. This pattern suggests a loss of cardiovascular adaptability, which aligns with research showing reduced autonomic flexibility in people with depression [28].

Other shifts were seen in sleep-related dynamics. The negative association between sleep in NREM phase and sleep duration that followed became stronger with symptom severity. This may reflect rebound-like processes in sleep regulation, where deeper sleep one night leads to shorter or more irregular sleep the next, especially in individuals with depressive symptoms.

In contrast, the contemporaneous networks showed much smaller group differences. The strongest same-day links, such as between heart rate and physical activity, were stable across groups. Minor shifts were visible, such as a weak association between activity and sleep duration in the asymptomatic group that was absent in the symptomatic groups. However, the overall similarity of contemporaneous networks indicates that symptom-related changes are more clearly expressed in how biomarkers influence each other across days than within a single day.

Overall, digital biomarker dynamics became more tightly coupled over time as symptoms increased, while same-day patterns remained relatively stable. These findings suggest that depressive symptoms are reflected most clearly in how digital biomarkers interact across days. As symptoms increase, these systems may become more synchronised but also more rigid, reducing the body's ability to respond flexibly to daily challenges.

6

Discussion

This chapter reflects on the key findings of the study and places them in the broader context of digital phenotyping, network theory, and personalised care in depression. It discusses how dynamic relationships between digital biomarkers vary with symptom severity and how these insights may inform monitoring and intervention strategies. Strengths, limitations, and directions for future research are also addressed.

6.1. Answer to Subquestion 4

How can insights from digital biomarker networks be applied to personalised monitoring in depression?

The results of this thesis suggest several concrete ways in which digital biomarker networks can support personalised monitoring in depression. First, heart rate emerged as a central node in both temporal and contemporaneous networks. Given its consistent self-predictive pattern and multiple links with other biomarkers, heart rate could serve as an indicator of system change. While high autoregressive effects suggest that heart rate is usually stable from day to day, this also means that deviations from its expected trajectory may be particularly informative. For instance, sudden increases in average heart rate or reductions in variability could signal a disruption in the system and an upcoming shift in mental state. In practice, this could be implemented through wearable sensors that alert users or clinicians to deviations from personalised baselines, prompting preventive action.

The sleep in NREM phase showed strong out-strength in the temporal network, indicating that it influences multiple other variables across domains. In contrast, sleep time and sleep effectivity had high in-strength, meaning they are shaped by changes elsewhere in the system. This suggests that sleep in NREM phase may act as a driver node within the dynamic system, while sleep time and sleep effectivity function more as downstream indicators. This interpretation aligns with previous network research showing that targeting highly central or influential nodes can lead to wider network changes [26, 12]. While most prior work focused on symptom networks, the current findings suggest that physiological driver nodes, such as sleep in NREM phase, could serve a similar role in digital biomarker networks. Sleep in NREM phase encompasses the stages of light and deep sleep in which essential restorative processes occur, including tissue repair, immune function regulation, and autonomic balance. These physiological roles provide a plausible mechanism for its influence on other variables across domains [38]. Improving sleep in NREM phase, through behavioural interventions such as promoting consistent sleep schedules, reducing late-night stimulation, or managing stress, may have cascading benefits across domains like heart rate regulation and physical activity. These results highlight the potential of using network centrality to inform behavioural targets for early intervention in depression.

A key finding was that temporal network connectivity increased with symptom severity. In individuals with more symptoms, variables were more tightly interconnected, suggesting reduced flexibility and increased vulnerability to cascading effects. This pattern mirrors earlier findings in symptom network research, where denser networks were associated with more severe depressive episodes and a higher risk of persistence [14, 48]. However, a key limitation of symptom networks is that they rely on self-

reports, which cannot be continuously monitored in daily life. The present study extends these findings by showing that similar connectivity patterns emerge in digital biomarkers, which can be passively and continuously tracked through wearables. This opens up new possibilities for depression symptom prediction. If network connectivity is monitored in real time, systems could detect when a person's digital biomarkers become unusually coupled. For instance, when temporal network connectivity exceeds a personalised threshold, it could trigger an automatic alert. This might prompt a clinician to check in, initiate a preventive intervention, or offer digital support. By responding early to signs of increased coupling between domains like sleep, heart rate, and physical activity, such systems could potentially prevent depression symptoms or reduce their severity.

Together, these findings suggest that personalised monitoring systems should not only track absolute values, but also changes in the structure and dynamics of behavioural signals over time. Future digital tools could visualise changes in network connectivity, and provide real-time feedback to patients and clinicians. This would support a more dynamic, approach to care that aligns with the fluctuating nature of depression.

6.2. Answer to the Main Research Question

How do digital biomarkers dynamically interact over and in time in individuals with varying levels of depression, and how do these interactions differ across levels of symptom severity to inform personalised monitoring?

This thesis shows that digital biomarkers in individuals with varying levels of depression form a dynamic system. Rather than fluctuating randomly, these signals demonstrate consistent lagged relationships that reflect day-to-day dependencies. For example, heart rate emerges as a stable signal that both influences and is shaped by activity and sleep. Physical activity contributes to next-day changes in sleep and cardiovascular patterns. Sleep variables were more temporally variable than heart rate and activity measures, but some played distinct roles in shaping or connecting other domains.

Temporal networks were especially effective in capturing these directional dynamics. They revealed how changes in one variable can lead to changes in another across time, which contemporaneous networks, based on same-day associations, could not detect as clearly. Importantly, these temporal dynamics varied by symptom severity. Individuals with more symptoms showed higher network connectivity in their temporal networks, suggesting a shift from a flexible and adaptive system to one that is more rigid and tightly coupled. This tighter coupling may increase vulnerability to cascading effects, where a change in one domain spreads more easily through the system. Interestingly, individuals with mild symptoms showed the lowest overall connectivity, possibly reflecting a transitional or unstable phase with weaker and more inconsistent coordination.

These findings also offer practical guidance for personalised monitoring. Heart rate and sleep in NREM phase emerged as central drivers in the network, suggesting that deviations in these signals may serve as early indicators of physiological disruption. Because these biomarkers can be continuously monitored through wearables, they provide a scalable foundation for real-time detection. In particular, the observed increase in temporal network connectivity with symptom severity suggests that monitoring changes in network connectivity, may be more informative for personalized monitoring. If a person's digital biomarkers become more strongly coupled than their historical baseline, this could indicate reduced adaptability and increased symptoms. Such a shift could automatically trigger a personalised alert, prompting clinician review, digital support, or behavioural intervention.

In this way, network-informed monitoring offers more than a descriptive map of behaviour: it enables dynamic, adaptive feedback loops between digital biomarker observation and timely intervention. By continuously evaluating both the strength and structure of behavioural signals, future tools may better reflect the fluctuating nature of depression and support truly personalised mental health care.

6.3. Implications for Stakeholders and Systems

The results of this study highlight how dynamic interactions between digital biomarkers can inform more personalised and timely monitoring in depression. These findings hold practical implications for several key stakeholders and the wider mental health care system.

For individuals with depression, continuous monitoring could offer earlier detection of deterioration, outside of clinical treatment periods. At the same time, constant monitoring risks feeling intrusive, reinforcing feelings of illness or triggering unnecessary worry. To be effective, systems must support insight without overwhelming the user, offering meaningful feedback that respects privacy and preserves daily life as a space for recovery, not surveillance. Earlier research has shown that user-centred design is essential to achieving this balance: it is a requirement for sustaining engagement and ensuring that monitoring technologies feel supportive rather than burdensome.

Clinicians could use network-based monitoring to complement traditional depression assessment methods. Rather than waiting for patients to report worsening symptoms, clinicians might detect early signals that indicate reduced resilience. This enables just-in-time interventions that are often less intensive, reducing the need for frequent in-person contact. As a result, clinicians can allocate their time more efficiently, offering timely support to more patients without increasing workload. Such systems could help expand care capacity and shift the focus from crisis management to early prevention.

Technology developers are responsible for designing systems that translate complex network data into actionable insights. These systems must be interpretable for clinicians, and minimally invasive for patients. Ideally, they operate in the background through wearable sensors and provide clinicians with suggestions when needed. Developers should collaborate with clinicians and users to ensure that digital tools match real-world needs.

On a systemic level, the findings support a shift in how mental health is understood and managed. Such interventions enable more preventive and individualised support, which could help systemic issues such as the current high pressure on health care systems and long waiting lists, especially after COVID-19 [1]. By enabling digital interventions, these systems can alleviate pressure on mental health services by optimising resource allocation, allowing limited clinical capacity to be distributed more efficiently across a growing patient population.

Ethical and Practical Reflections

Digital phenotyping promises personalised care, but it also raises ethical and social concerns. Many mental health apps lack clinical validation and do not follow medical guidelines, which can lead to misinformation, delayed treatment, or false reassurance [9]. The large number of unregulated tools makes it difficult for users and professionals to identify trustworthy options. Beyond technical risks, digital phenotyping introduces new forms of medical practice. Rather than explaining illness through causal mechanisms, digital tools focus on correlations between behavioural and physiological data [58]. This shift from causal understanding to predictive patterns can undermine psychological insight and the meaning patients attach to their experiences. If clinical attention turns to data patterns, deeper causes and personal narratives risk being ignored. The technology also redefines the relation between patients and clinicians. Real-time feedback and AI-driven alerts may raise false expectations about certainty or automation in care. These tools are not neutral: they shape how patients understand their bodies and how clinicians define health and illness [58]. Privacy and data governance also remain unresolved. Long-term monitoring of activity, sleep, and heart rate creates risks of data misuse, and privacy concerns. Yet current legal frameworks often lag behind technical developments [60]. Trust is a central concern, not only in data handling but in how systems are designed and explained to users. Finally, access to digital health tools is uneven. People without smartphones, digital literacy, or financial resources may be excluded. Without inclusive design and policy support, digital phenotyping could reinforce existing inequalities in mental health care [60].

In short, the value of digital phenotyping depends not only on technical progress, but on how its risks are addressed. Ethical reflection must go beyond regulation and include broader questions about how technology changes the meaning of health, care, and personhood.

6.4. Relation to Previous Research

This study builds on earlier research that suggests higher symptom severity in depression is linked to increased network connectivity and reduced resilience. Studies using static symptom networks [14, 10, 48] have consistently found that tightly coupled systems may be more vulnerable to cascading effects. These findings support the idea that depression is not only a matter of individual symptom levels, but

also of how symptoms and behaviours influence each other over time.

However, most of these studies have focused either on self-reported symptoms or isolated behavioural domains. This thesis extends previous work by integrating multiple digital biomarker signals and modelling their dynamics using network models. This approach makes it possible to estimate how different physiological domains interact over time and how these interactions differ across symptom severity levels.

The study thereby responds to recent calls for more dynamic and individualised models in mental health research, particularly within the field of digital phenotyping [15, 35]. Rather than relying on static, or self-reported data, the current analysis focuses on within-person dynamics and temporal coordination between digital biomarkers.

At the same time, the study remains exploratory in nature. Although the integration of passive data into multilevel VAR modelling is relatively novel, the findings should be interpreted with caution. As prior literature on centrality and network reliability has shown, estimated relationships may be sensitive to sample characteristics, model assumptions, or data quality [16, 11]. Moreover, the absence of self-report data means that interpretations remain limited to behavioural patterns, without direct reference to lived experience or emotional state.

In sum, the present study offers a modest but meaningful contribution to existing literature. It provides one of the first examples of applying multilevel VAR to wearable-derived biomarkers in the context of depression and explores how their temporal organisation shifts with symptom levels. This adds to a growing body of research seeking to move beyond static measurement and toward a more dynamic understanding of mental health.

6.5. Data-Driven Models: Value and Limits

This thesis used a data-driven approach to identify meaningful patterns in digital behaviour, without relying on subjective symptom ratings or predefined theoretical models. This is both a strength and a limitation. On the one hand, it allows for the discovery of new relationships that may not be visible in traditional frameworks. Objective sensor data are not biased by memory, social desirability, or interpretation. On the other hand, the results may miss important context. For example, a decrease in heart rate variability could reflect stress, but also relaxation after exercise or a technical artefact.

Moreover, purely data-driven models may be limited in their ability to explain or guide personalised interventions. While they detect patterns, they do not explain why those patterns occur or how individuals experience them. This creates a gap between statistical signals and lived meaning. Future research should aim to bridge this gap by combining digital data with clinical interviews, diaries, or ecological momentary assessments (EMA). Still, even without subjective input, this study provides a valuable framework for identifying candidates for follow-up or intervention.

From a theoretical standpoint, data-driven modelling also raises questions. Because the feature selection was not theory-guided but data-driven, the results may be less aligned with existing psychological constructs.

6.6. Strengths and Limitations

The unique strength of this study is that it analyzed a large, real-world dataset with digital phenotyping data to uncover the temporal dynamics between heart rate, sleep, and physical activity. In doing so, it moves beyond the limitations of self-report data. The inclusion of multiple physiological and behavioural domains allowed for insight into the dynamics of daily functioning.

Several limitations should be noted. First, centrality measures are reported in this research, but their meaning in psychological data is unclear. As Bringmann et al. (2019) [11] point out, these measures were made for social networks and rely on assumptions that often do not fit psychological symptoms, such as flow between nodes or shortest paths. Especially, closeness and betweenness centrality may be misleading. Results based on centrality should therefore be treated with care.

Also, no person-specific networks were estimated. Although the multilevel structure captures both group-level effects and individual deviations, the analysis did not explicitly model networks at the indi-

vidual level due to time constraints. Given the emphasis on personalised monitoring, this is a missed opportunity, as relevant dynamics may differ across individuals. Second, centrality and connectivity were not tracked over time, which limits inferences about system shifts or deterioration at the individual level.

Also although statistical significance was determined using a conventional threshold ($p < .05$), this approach comes with limitations in the context of high-dimensional network models. With many edges tested simultaneously, the likelihood of false positives increases. As such, network results should be interpreted as exploratory rather than definitive, and effect size patterns should be prioritised alongside significance. Future studies could improve precision by applying false discovery rate correction or Bayesian methods that better account for uncertainty.

Lastly, an essential limitation of this study is that no formal robustness analyses were conducted for the estimated networks. While methods such as bootstrapping are commonly used to assess the stability of edge weights and centrality indices, these techniques are primarily developed for cross-sectional or between-subject networks [17]. As Epskamp et al. (2018) note, “bootstrapping accuracy is primarily developed for between-subjects cross-sectional networks. For time-series models such as VAR or multilevel VAR, the interpretation and performance of these methods is still under development” [17, p. 621]. Although it is technically possible to implement manual case-dropping or resampling procedures for temporal and contemporaneous models, such approaches are highly time-consuming and would require custom programming and computational resources beyond the scope of this thesis. As a result, this study does not include robustness checks for the temporal networks. Future research may explore more advanced tools for evaluating the stability of multilevel VAR models once such methods become more standardized and accessible.

Taken together, these strengths and limitations highlight the potential of digital network models for understanding depression, while also pointing to necessary improvements for more robust and personalised applications.

6.7. Implications and Future Directions

This thesis contributes to the idea that digital biomarkers can be used not just for tracking symptoms, but for modelling the structure of behavioural systems. A key insight is that temporal connectivity increases with symptom severity, possibly signalling reduced flexibility. This could be developed into early warning systems that track rising coupling before depression symptoms arise. Such tools would help shift care from a reactive to a preventive approach.

Another direction is to examine how network patterns vary within individuals over time. This would require estimating person-specific networks and studying how connectivity, centrality, or edge strength change across phases of risk or recovery. Doing so would provide insight into the relation between connectivity measures, for example, and symptom levels, which is essential in the development of network thresholds that could indicate deterioration.

A further avenue is to improve the modelling tools. Current multilevel VAR methods are limited in terms of bootstrapping and sensitivity checks. More robust methods would increase confidence in dynamic monitoring tools.

Altogether, future work could focus on linking dynamic network features to meaningful clinical outcomes. This would help make digital phenotyping models not only descriptive but also predictive and actionable in real-world care.

References

[1] Nafiso Ahmed et al. “Mental health in Europe during the COVID-19 pandemic: a systematic review”. In: *The Lancet Psychiatry* 10.7 (2023), pp. 537–556.

[2] Hirotugu Akaike. “A new look at the statistical model identification”. In: *IEEE transactions on automatic control* 19.6 (2003), pp. 716–723.

[3] Laura Helena Andrade et al. “Barriers to mental health treatment: results from the WHO World Mental Health surveys”. In: *Psychological medicine* 44.6 (2014), pp. 1303–1317.

[4] Quentin André. “Outlier exclusion procedures must be blind to the researcher’s hypothesis.” In: *Journal of Experimental Psychology: General* 151.1 (2022), p. 213.

[5] Roseanne Armitage. “Sleep and circadian rhythms in mood disorders”. In: *Acta Psychiatrica Scandinavica* 115.s433 (2007), pp. 104–115. DOI: 10.1111/j.1600-0447.2007.00967.x.

[6] Robert H Belmaker and Galila Agam. “Major depressive disorder”. In: *New England Journal of Medicine* 358.1 (2008), pp. 55–68.

[7] Claudia van Borkulo et al. “Association of symptom network structure with the course of depression”. In: *JAMA psychiatry* 72.12 (2015), pp. 1219–1226.

[8] Denny Borsboom et al. “Network analysis of multivariate data in psychological science”. In: *Nature Reviews Methods Primers* 1.1 (2021), p. 58.

[9] Elisa Brietzke et al. “Integrating digital phenotyping in clinical characterization of individuals with mood disorders”. In: *Neuroscience & Biobehavioral Reviews* 104 (2019), pp. 223–230.

[10] Laura F Bringmann et al. “A network approach to psychopathology: New insights into clinical longitudinal data”. In: *PloS one* 8.4 (2013), e60188.

[11] Laura F Bringmann et al. “What do centrality measures measure in psychological networks?” In: *Journal of abnormal psychology* 128.8 (2019), p. 892.

[12] Daniel Castro et al. “Centrality measures in psychological networks: A simulation study on identifying effective treatment targets”. In: *Plos one* 19.2 (2024), e0297058.

[13] Silvia Coretti, Filippo Rumi, and Americo Cicchetti. “The social cost of major depression. A systematic review”. In: *Rev. Eur. Stud.* 11 (2019), p. 73.

[14] Angélique OJ Cramer et al. “Major depression as a complex dynamic system”. In: *PloS one* 11.12 (2016), e0167490.

[15] M Domhardt et al. *More light? Opportunities and pitfalls in digitalized psychotherapy process research. Front Psychol 12: 544129*. 2021.

[16] Sacha Epskamp, Denny Borsboom, and Eiko I Fried. “Estimating psychological networks and their accuracy: A tutorial paper”. In: *Behavior research methods* 50 (2018), pp. 195–212.

[17] Sacha Epskamp and Eiko I Fried. “A tutorial on regularized partial correlation networks.” In: *Psychological methods* 23.4 (2018), p. 617.

[18] Sacha Epskamp et al. “Personalized network modeling in psychopathology: The importance of contemporaneous and temporal connections”. In: *Clinical Psychological Science* 6.3 (2018), pp. 416–427.

[19] Sacha Epskamp et al. “The Gaussian graphical model in cross-sectional and time-series data”. In: *Multivariate behavioral research* 53.4 (2018), pp. 453–480.

[20] Leandre R Fabrigar et al. “Evaluating the use of exploratory factor analysis in psychological research.” In: *Psychological methods* 4.3 (1999), p. 272.

[21] Maria Faurholt-Jepsen et al. “Differences in mobility patterns according to machine learning models in patients with bipolar disorder and patients with unipolar disorder”. In: *Journal of Affective Disorders* 306 (2022), pp. 246–253.

[22] Aaron J Fisher, John D Medaglia, and Bertus F Jeronimus. “Lack of group-to-individual generalizability is a threat to human subjects research”. In: *Proceedings of the National Academy of Sciences* 115.27 (2018), E6106–E6115.

[23] J Kevin Ford, Robert C MacCallum, and Marianne Tait. “The application of exploratory factor analysis in applied psychology: A critical review and analysis”. In: *Personnel psychology* 39.2 (1986), pp. 291–314.

[24] Eiko I Fried. “Problematic assumptions have slowed down depression research: why symptoms, not syndromes are the way forward”. In: *Frontiers in psychology* 6 (2015), p. 309.

[25] Eiko I Fried. “The 52 symptoms of major depression: Lack of content overlap among seven common depression scales”. In: *Journal of affective disorders* 208 (2017), pp. 191–197.

[26] Eiko I Fried and Angélique OJ Cramer. “Moving forward: Challenges and directions for psychopathological network theory and methodology”. In: *Perspectives on Psychological Science* 12.6 (2017), pp. 999–1020.

[27] Eiko I Fried and Randolph M Nesse. “Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR* D study”. In: *Journal of affective disorders* 172 (2015), pp. 96–102.

[28] Jessica L Hamilton and Lauren B Alloy. “Atypical reactivity of heart rate variability to stress and depression across development: Systematic review of the literature and directions for future research”. In: *Clinical psychology review* 50 (2016), pp. 67–79.

[29] Wesllei Felipe Heckler et al. “Digital phenotyping for mental health based on data analytics: A systematic literature review”. In: *Artificial Intelligence in Medicine* (2025), p. 103094.

[30] Jia-Hao Hsu et al. “Digital phenotyping-based bipolar disorder assessment using multiple correlation data imputation and lasso-MLP”. In: *IEEE Transactions on Affective Computing* (2023).

[31] Thomas R Insel. “Digital phenotyping: a global tool for psychiatry”. In: *World Psychiatry* 17.3 (2018), p. 276.

[32] Spencer L James et al. “Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017”. In: *The lancet* 392.10159 (2018), pp. 1789–1858.

[33] Lewis L Judd. “The clinical course of unipolar major depressive disorders”. In: *Archives of general psychiatry* 54.11 (1997), pp. 989–991.

[34] Julian D Karch. “Outliers may not be automatically removed.” In: *Journal of Experimental Psychology: General* 152.6 (2023), p. 1735.

[35] Alexander Kathan et al. “Personalised depression forecasting using mobile sensor data and ecological momentary assessment”. In: *Frontiers in digital health* 4 (2022), p. 964582.

[36] Andrew H Kemp et al. “Impact of depression and antidepressant treatment on heart rate variability: A review and meta-analysis”. In: *Biological Psychiatry* 67.11 (2010), pp. 1067–1074.

[37] Liia MM Kivelä et al. “Examining contemporaneous and temporal associations of real-time suicidal ideation using network analysis”. In: *Psychological Medicine* 54.12 (2024), pp. 3357–3365.

[38] Meir H Kryger, Thomas Roth, and William C Dement. *Principles and practice of sleep medicine E-book: Expert consult-online and print*. Elsevier Health Sciences, 2010.

[39] Imogen E Leaning et al. “From smartphone data to clinically relevant predictions: A systematic review of digital phenotyping methods in depression”. In: *Neuroscience & Biobehavioral Reviews* 158 (2024), p. 105541.

[40] Anna McKeever, Mark Agius, and Pavel Mohr. “A review of the epidemiology of major depressive disorder and of its consequences for society and the individual”. In: *Psychiatria Danubina* 29.suppl. 3 (2017), pp. 222–231.

[41] Isaac Moshe et al. “Digital interventions for the treatment of depression: A meta-analytic review.” In: *Psychological bulletin* 147.8 (2021), p. 749.

[42] Sandrine R Müller et al. “Depression predictions from GPS-based mobility do not generalize well to large demographically heterogeneous samples”. In: *Scientific Reports* 11.1 (2021), p. 14007.

[43] Tero Myllymäki et al. “Effects of exercise intensity and duration on nocturnal heart rate variability and sleep quality”. In: *European journal of applied physiology* 112.3 (2012), pp. 801–809.

[44] Megan Norris and Luc Lecavalier. “Evaluating the use of exploratory factor analysis in developmental disability psychological research”. In: *Journal of autism and developmental disorders* 40.1 (2010), pp. 8–20.

[45] Maurice M Ohayon et al. “Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan”. In: *Sleep* 27.7 (2004), pp. 1255–1273.

[46] Jukka-Pekka Onnela. “Opportunities and challenges in the collection and analysis of digital phenotyping data”. In: *Neuropsychopharmacology* 46.1 (2021), pp. 45–54.

[47] Sung-Ho Park and Franz Weber. “Neural and homeostatic regulation of REM sleep”. In: *Frontiers in psychology* 11 (2020), p. 1662.

[48] Madeline Lee Pe et al. “Emotion-network density in major depressive disorder”. In: *Clinical Psychological Science* 3.2 (2015), pp. 292–300.

[49] George D Price et al. “Using digital phenotyping to capture depression symptom variability: detecting naturalistic variability in depression symptoms across one year using passively collected wearable movement and sleep data”. In: *Translational Psychiatry* 13.1 (2023), p. 381.

[50] David Proudman, Paul Greenberg, and Dave Nellesen. “The growing burden of major depressive disorders (MDD): implications for researchers and policy makers”. In: *Pharmacoeconomics* 39.6 (2021), pp. 619–625.

[51] Steven P Reise, Niels G Waller, and Andrew L Comrey. “Factor analysis and scale revision.” In: *Psychological assessment* 12.3 (2000), p. 287.

[52] A John Rush et al. “Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR* D report”. In: *American Journal of Psychiatry* 163.11 (2006), pp. 1905–1917.

[53] Amirreza Sajjadieh et al. “The association of sleep duration and quality with heart rate variability and blood pressure”. In: *Tanaffos* 19.2 (2020), p. 135.

[54] Alireza Sameh et al. “Digital phenotypes and digital biomarkers for health and diseases: a systematic review of machine learning approaches utilizing passive non-invasive signals collected via wearable devices and smartphones”. In: *Artificial Intelligence Review* 58.2 (2024), p. 66.

[55] Marília Pit dos Santos et al. “Machine learning applied to digital phenotyping: a systematic literature review and taxonomy”. In: *Computers in Human Behavior* 161 (2024), p. 108422.

[56] Lina Sjösten. *A Comparative Study of the KPSS and ADF Tests in terms of Size and Power*. 2022.

[57] Sleep Foundation. *Sleeping Heart Rate: What’s Normal and Why It Matters*. Accessed: 2025-07-20. 2024. URL: `https://www.sleepfoundation.org/physical-health/sleeping-heart-rate`.

[58] Giovanni Stanghellini and Federico Leoni. “Digital phenotyping: Ethical issues, opportunities, and threats”. In: *Frontiers in Psychiatry* 11 (2020), p. 473.

[59] Statistics Netherlands (CBS). *Which proportion of missing values in the data is allowed?* Accessed: 2025-04-07. 2023. URL: `https://www.cbs.nl/en-gb/society/nature-and-environment/indexes-and-trends--trim--/trim-frequently-asked-questions/which-proportion-of-missing-values-in-the-data-is-allowed-`.

[60] Ana Tomičić, Anamaria Malešević, and Anto Čartolovni. “Ethical, legal and social issues of digital phenotyping as a future solution for present-day challenges: a scoping review”. In: *Science and engineering ethics* 28.1 (2022), p. 1.

[61] Srikanth Vasudevan et al. “Digital biomarkers: convergence of digital health technologies and biomarkers”. In: *NPJ digital medicine* 5.1 (2022), p. 36.

[62] Judith Verduijn et al. “Reconsidering the prognosis of major depressive disorder across diagnostic boundaries: full recovery is the exception rather than the rule”. In: *BMC medicine* 15 (2017), pp. 1–9.

[63] J Elisabeth Wells and L John Horwood. “How accurate is recall of key symptoms of depression? A comparison of recall and longitudinal reports”. In: *Psychological medicine* 34.6 (2004), pp. 1001–1011.

[64] World Health Organization. *Depression*. `https://www.who.int/news-room/fact-sheets/detail/depression`. Accessed: 2025-07-16. 2023.