Multilayer Network Analysis with Psychological Data: Promises and Pitfalls

Jolien Venhuizen, s105070

Supervisors and advisors: F. Hasselman, R. Kievit, N. S. De Boer & L. De Bruin

Radboud Interfaculty Complexity Hub

Complex Systems Lab, Behavioural Science Institute, Radboud University

Major Research Project (SOW-BS028)

M. Luijten

July 2022

Abstract

It has been suggested that multilayer network models could be promising mathematical tools for the study of complex human behavioural phenomena. One of the reasons is that these models can be used to correct for differences between different data sources, which traditional network models cannot. However, multilayer networks models are only a few times empirically applied to psychological data and an important question is if we indeed can and should correct for differences between sources by conceptualizing the network as multi-layered. The aim of the current proof-of-concept study is to empirically apply multilayer network analysis on psychological, multisource data and document choices and difficulties in the process. Multisource data (mental health, physical health and cognitive function) from the Survey of Health, Ageing and Retirement in Europe (SHARE) dataset of elderly people was used. The multilayer network was estimated using Sparse Network And Component Supplemental (SNACS) model estimation, which uses a pre-processing step to estimate the underlying component structure of the data and takes this into account when estimating a regularized partial correlation network. Results showed that a) there is an underlying structure in the multisource data and; b) taking this into account results in a different network model. However, the results heavily depend on the modelling decisions made, hampering the robustness of the findings. Moreover, the interpretation of the results is not straightforward. Nonetheless, this study indicates that taking into account differences between sources might be a necessary first step when using multisource data. Future research directions are suggested.

Keywords: multilayer, multisource, network analysis, psychological, complexity

Multilayer Network Analysis: Promises and Pitfalls

Researchers in the psychological sciences emphasize the need for the collection and analysis of large, multi-source, multidisciplinary datasets and collaboration between disciplines to understand psychological phenomena (Bartel et al., 2013; Silverman & Loscalzo, 2012). This is necessary, because factors that underlie these phenomena exceed the boundaries of individual disciplines. This highlights the importance of understanding how data from different disciplines or domains relate to one another. For instance, how do variables typically collected in psychological sciences relate to those collected in neuroscience and biology, and how do these relations constitute the emergence of specific behaviour? Taking into account the relations between these different domains might broaden our understanding, open new treatment venues and increase the predictive accuracy of our models (Tio et al., 2020). Hence, there is a need in psychological science to develop tools that allow us to study psychological phenomena in a multidimensional manner.

Multisource Data and Complex Systems

The trend to use multisource data when analysing a psychological phenomenon is in line with a view that has gained popularity in the past decades: the notion that human behaviour is a complex dynamical system and should be analysed as such (e.g. Benthem de Grave et al., 2020; Blanken et al., 2021; Borsboom, 2017; Braun et al., 2018; Brooks et al., 2020; De Boer et al., 2021; Foote, 2007; Hasselman, 2022; Heino et al., 2020; Ladyman et al., 2013; Norman et al., 2011; Olthof et al., 2021; Robinaugh et al., 2020; Tanskanen et al., 2007). A complex dynamic system is a composition of parts and wholes in which interdependencies between parts lead to the emergence of wholes (Van Geert, 2019). In order to understand this complex behaviour, interactions between the parts of the system are of interest, rather than parts in isolation (Hinch et al., 2006). This is because complex dynamic behaviour is considered to be the result of interactions between processes across multiple spatial and temporal scales

(interaction-dominant dynamics). In other words, it is characterised by emergence, self-organisation, scaling, interference and synergies (see e.g. Heino et al., 2020; Ihlen & Vereijken, 2010; Szary et al., 2015; Turvey, 2007; Van Orden et al., 2003). When translated to psychological science, the dynamic system (this can for example be human behaviour or mental health or psychopathology) emerges from self-organized interactions between parts of this system, such as interactions within and between biological, environmental, psychological and neurological factors (Olthof et al., 2021). This approach is thus applicable to psychological phenomena. What this means for the study of psychological phenomena will now be further explained.

Psychological Complex Phenomena

Most studies that have been done in psychological science are in line with the assumption of independence, which is the idea that interacting factors interact in an additive manner, which allows them to be studied in isolation (Van Orden et al., 2003). This assumption has led psychological scientists to focus on unique contributions of specific factors to psychological phenomena. In contrast, in the complex adaptive systems framework, interacting factors are seen as inseperable because processes in one subsystem (e.g. formed by psychological factors) are dependent on the processes of all other related systems (Van Orden et al., 2012). Empirical studies that compared both assumptions have found evidence in favor of the interdependence assumption, where taking into account interactions between subsystems (which can for example be originated from different sources) increased the amount of explained variance (Ihlen & Vereijken, 2010; Wijnants, 2014). This implies that complex human phenomena, such as psychopathology, need to be treated and studied as a whole, and thus with multiple levels of explanation (Olthof et al., 2021; Van Geert, 2019; Wallot & Kelty-Stephen, 2018).

An example of a psychological phenomenon that can be seen as a whole, is the concept depression (Van Geert, 2019). Depression is studied in many different disciplines and theoretical frameworks and is therefore reflected in different kind of factors, studied by distinct domains. For instance, the factors persistent sad emotion (psychological factor), rumination (cognitive factor), inactivity (behavioural factor) and loss of appetite (physiological factor) all reflect depression. Moreover, these different variables of different domains often interact with each other, for instance when rumination → sad emotion → inactivity → rumination. This is called a feedback loop, because rumination is both the reason for the sad emotion but indirectly also the consequence of it. These feedback loops within the system make the phenomenon complex (Borsboom & Cramer, 2013). An important consequence of conceptualizing a phenomenon as an interdependent whole of subsystems (e.g. behavioural, physiological etc.), is thus that the phenomenon needs to be studied as a whole, instead of studying only one subsystem of the phenomenom. This indeed calls for studies with data from multiple sources. Moreover, this conceptualization of psychological phenomena as complex systems creates an opportunity where techniques used in complexity science can be used to examine psychological phenomena.

Network Tools

One way to increase understanding of relationships within a complex system, is using statistical tools from network science (Barabási, 2016) to estimate relations among elements of the system of interest (Fried, 2020; Fried & Robinaugh, 2020). The network approach suits the complex systems approach, since networks can be used to analyse complex relationships such as feedback loops (described earlier). The popularity of such network psychometrics approaches have increased over the past decade (Borsboom, 2017; Robinaugh et al., 2020b). We will first explain what a network is. The simplest network can be seen as a visualisation of a correlation matrix of all the data. A network can consists of nodes and edges. Nodes in a

symptom network represent the variables, for example of a psychological construct like feeling depressed. Edges represent the relationships between the nodes and can be estimated in several, more, or less complex psychometrics (e.g. correlation or unique relationships). Using this approach has led to valuable insights (Koller & Friedman, 2009). However, most network studies have been performed with data from only one level of analysis (e.g. self-ratings of psychological states). This aspect is in line with the assumption of independence described above, and not with the assumption of interdependence which suits the complexity.

Levels of Explanation

The reason why the use of one level of analysis for complex multisource data is specifically problematic, is because multisource data is expected to consist of clusters. Clustering arises when some nodes are more alike than others. This can be the case when variables have different spatial or temporal scales, when there are different subsystems in the data or when nodes are originated from different sources (such as different datasets, measures or operationalisations). Multisource data is therefore expected to cluster based on the source it belongs to and therefore, edges between and within those clusters need to be treated differently according to the assumption of interdependence, where clusters are seen as subsystems that interact with each other (e.g. Ihlen & Vereijken, 2010; Wijnants, 2014). If these differences in edges within and between clusters are not taken into account, the representations of the estimated edges might be biased, which is the case for networks with one level of explanation (Tio et al., 2020). Moreover, Miller (2010) and Thomas and Sharp (2019) reviewed attempts to link different levels of explanation in a monolayer network and argued against popular monolayer network strategies for the use of multiple levels of explanation because of their inadequateness on logical as well as conceptual grounds. However, it is suggested to use multilayer network models for the study of psychological phenomena with multiple levels of explanation (Boccaletti et al., 2014; Braun et al., 2018; De

Boer et al., 2021; Hasselman, 2022; Nelson et al., 2017; Norman et al., 2011; Riese & Wichers, 2021).

Multilayer Network Models

In the past couple of years, multilayer network analysis has become increasingly popular after the physics community began to use them with names like multilayer networks (Jo et al., 2006; Kurant et al., 2007; Kurant & Thiran, 2006), network of networks (Zhou et al., 2006, 2007) and node-colored networks (Newman, 2003; Vazquez, 2006). These models could theoretically integrate different types of data (e.g. neurobiological, physiological, psychological, environmental) into one model, allowing for insights and analyses that could not be performed when focusing one layer of analysis only (Kivelä et al., 2014). Following the multilayer network framework of Kivelä and colleagues (2014), we conceptualize *multilayer network* as an overarching term for various network types. We consider a network type multilayer when the network consists of multiple layers with within- and between-layer edges. Every single layer represents a subsystem of the whole system. The most described goal of multilayer network analysis is to increase understanding about the dynamics (e.g. Nelson et al., 2017) and/or topological features (e.g. Kaiser et al., 2007) of a certain system. We will now explain different categories of multilayer network types in further detail.

Multilayer Network Categories

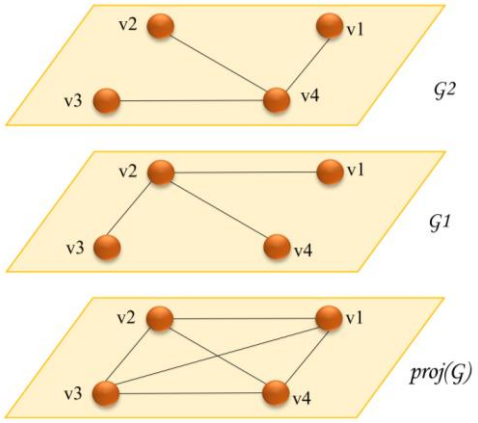
In this section, we will discuss the most important types of multilayer networks for psychological phenomena (for an extensive overview, see e.g. Bianconi, 2018; Boccaletti et al., 2014; Hammoud & Kramer, 2020; Kivelä et al., 2014). Multilayer networks can be categorized into two broad categories: edge-based and node-based multilayer networks (Hammoud & Kramer, 2020). We now briefly describe these categories and most common types that exist within those categories.

Edge-based. Edge-based networks are networks used to examine different types of relationships between nodes. Every layer represents another type of relationship and the network is node-aligned, which means that nodes can belong to multiple layers (for exceptions, see e.g. Buccafurri et al., 2013). The most used multilayer network type in empirical studies is edge-based, which is the multiplex network (Gluckman, 1995; Szell et al., 2010; Verbrugge, 1979), which can be seen in Figure 1. This network type originated from sociology and can for example be used to examine different types of one-to-one interactions between players of a game (e.g. positive connotation such as friendship, communication and trade, or negative connotation such as enmity, armed aggression, punishment), where every interaction type is placed on another layer (Szell et al., 2010).

Another edge-based network is the multilevel network (Figure 2). This network type uses subsets of nodes as wel as subsets of edges on every layer, where nodes can exist on multiple layers. This type is for example used for a transportation network where layers represent different transportation methods and not all stops (nodes) are used by all transportation methods (Aleta et al., 2017). Within edge-based networks, one can also make a distinction between different types of nodes (for example based on gender) and hierarchical structures can be included in multilevel networks, such as individuals as part of organizations (Lazega et al., 2008). Moreover, when the same nodes are being examined at different points in time in one multilayer network, this can also be conceptualised as an edge-based network. Each layer can for example be used to construct the network at each timepoint. The focus in edge-based networks is on the interactions and dynamics within and between the layers of the system.

Figure 1

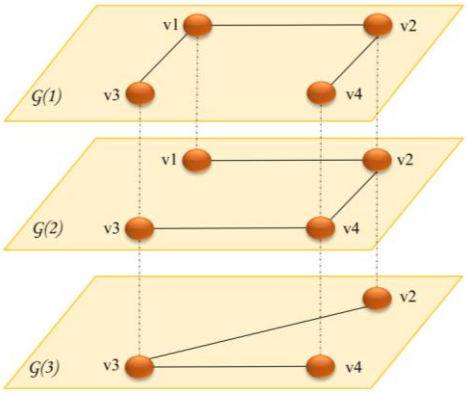
Multiplex Network



Note. All layers represent another type of relationship between the same nodes. Reproduced without permission from Hammoud and Kramer (2020).

Figure 2

Multilevel Network



Note. Every layer consists of a subset of nodes and edges. Between-layer edges only exist between the same nodes on different layers. Reproduced without permission from Hammoud and Kramer (2020).

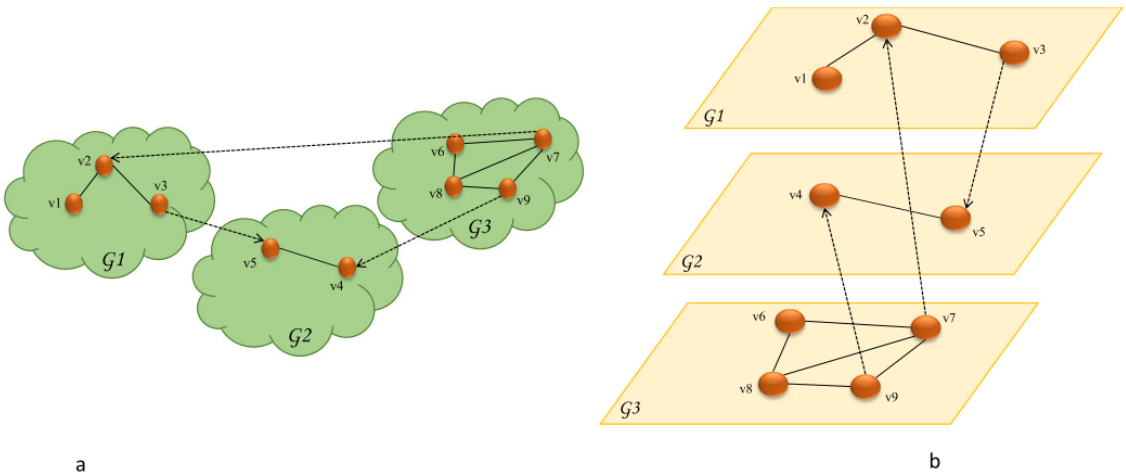
Node-based. In contrast to edge-based networks, node-based networks are not focused on different types of relationships but instead on different groups. The dataset is first divided into groups and every group is placed on a single layer. Most node-based networks are therefore layer-disjoint, which means that a node only exists on one layer. An example is the interconnected network (Figure 3), also called interacting network, or, network of networks

(e.g. Brummitt et al., 2012; Buldyrev et al., 2010; Criado et al., 2012; Donges et al., 2011; Rocklin & Pinar, 2013; Rosato et al., 2008). It can be seen as a collection of networks (e.g. subsystems or groups are placed on a single layer and every layer forms a network that interacts with other networks (Donges et al., 2011).

A more complex node-based network is the hypergraph (Chemero & Turvey, 2008). In this network type, different intersecting subnetworks can be analysed, with nodes that can belong to multiple subgroups and thus to multiple layers (Figure 4). These networks are therefore node-aligned. The edges in the network are a special kind and called hyperedges. These hyperedges can connect multiple nodes at the same time. Each hyperedge is mapped to a single layer (for mor information on hypergraphs and hyperedges, we refer to Chemero & Turvey, 2008). The reason this type is considered node-based is that the focus is on the grouping in the network. The focus in node-based networks in general is on the exploitation of the topology within the system. For instance, these network types might help achieve global synchronization between different layers (for more information on global synchronization, see Boccaletti et al., 2014).

Figure 3

Interconnected Network

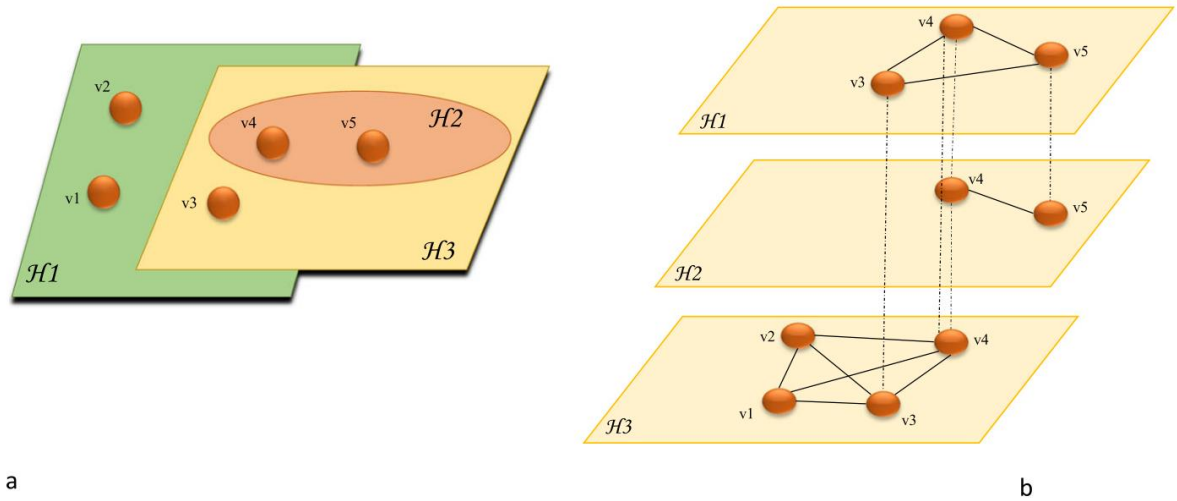


Note. **a.** The network in a monolayer view (left); **b.** The network in a multilayer view (right).

Nodes that belong to the same group are placed on the same layer. Reproduced without permission from Hammoud and Kramer (2020).

Figure 4

Hypergraphs



Note. **a.** The network in a monolayer view (left); **b.** The network in a multilayer view (right). Nodes that belong to the same sub-group are placed on the same layer. Note that the difference with an interconnected network is that a node can belong to multiple sub-groups and thus layers. Reproduced without permission from Hammoud and Kramer (2020).

Multilayer Networks in Psychology

We will now give some examples of the use of multilayer networks in psychology. Multilayer networks in psychology have been primarily proposed to be of use for the concept of psychopathology (Braun et al., 2018; Maatman, 2020; Masten et al., 2021; Olthof et al., 2021) but also on personality (Brooks et al., 2020), stress (Epel et al., 2018; Norman et al., 2011), brain and cognition (Simpson-Kent et al., 2021) and human experience (Hasselman & Bosman, 2020). To give a more in-depth idea of the possibilities when using multilayer networks in psychology, we will now explain two studies in further detail.

Simpson-Kent et al. (2021) examined the neurocognitive structure of general intelligence in struggling learners. In this study, three distinct monolayer networks were compared with one multilayer network that combined those networks. One layer was a cognitive network with verbal abilities with a focus on reading and two other layers were both a structural brain network: one layer consisting of grey matter cortical volume and one layer consisting of white matter fractional anisotropy. The authors found multiple positive and negative partial correlations between the layers of networks, indicating that brain and behaviour interact with each other.

Another study that uses empirical data is the study by Hasselman & Bosman (2020). A multilayer design was used to study human experience using Ecological Momentary Assessment (EMA) data, which were self-ratings of, for instance, self-esteem, depression and mood. The EMA data contains timeseries, since the same items are collected multiple times a day for multiple days. To take these timeseries into account, recurrence networks were created for every variable (for an explanation of recurrence networks, we refer to Hasselman & Bosman, 2020). These recurrence networks were then connected using a multilayer network design. Moreover, multiple between-layer edge psychometrics were compared in this study. This unique approach to multilayer networks is promising for the use of timeseries data, which provide more accurate information about the patterns in complex dynamic systems than studying variables on only one point in time (Wijnants, 2014).

The present study

The studies described in the last paragraph are unfortunately two of the very few empirical studies using multilayer networks in psychology. Most studies namely focus on the theoretical framework or on simulation studies. As a consequence, many conceptual and methodological challenges concerning the nature of psychological data are not yet put to the test. Specificallly, to date it is not clear yet how we should construct, analyse and interpret the

multilayer network model in psychology and what methodological and conceptual issues might arise when doing so, and how network structures of multilayer networks are different from network structures in monolayer networks.

To fill this gap, the present paper is structured as a guideline for researchers that are planning to use multilayer network analysis for multisource data to study psychological phenomena. The aim and structure of this study is threefold: 1) to provide an overview of steps and choices that need to be made when constructing a multilayer network for psychological phenomena; 2) to perform a proof-of-concept study and; 3) to discuss conceptual and methodological advantages and limitations of this approach. We follow through with a summary of the results and provide suggestions for future research.

Modelling Decisions when Construing Multilayer Networks: A Step-by-step Guide

Step 1: Goalsetting

The goal of your study can already constrain the modeling decisions that have to be made. First, the goal of your study will inform what kind of variables are of interest. Second, it will inform what type of multilayer network structure you should use. For instance, if you are interested in relationships between groups of variables, the model should be node-based; if you are interested in examining multiple relationship types between the same variables, your network should be edge-based.

Step 2: Choosing Variables

What variables to select for your model, will depend on your specific goals, on theoretical considerations, and on the dataset available. Here, we will focus specifically on the latter. One method to collect suitable data is combining data from different sources, for example by extracting information from different social-media websites (Magnani & Rossi, 2013). Another method is to directly collect different types of data from one source, such as

mobile-phone billing, where text messages as well as phone calls can be extracted (Zignani et al., 2014). A third method is to use observation and differentiate between different types of interactions (Szell et al., 2010). A fourth method is to directly collect different types of data from the same participants, such as neurological and psychological data.

As always when using psychological data, be aware of interpretation difficulties, for example as a consequence of overlap in concepts, unclear definitions, instrumental limitations, interpretation difficulties and missingness (Jensen et al., 2022). If possible, avoid them (see e.g. Bringmann et al., 2020) but if not, these difficulties have to be taken into account when interpreting the results.

Step 3: Multilayer Network Design

The next step is to design the multilayer network. There are many different possibilities and one can get quite creative with it. However, the goal of the study and the chosen dataset will exclude many options, which will make it easier to choose. In designing your multilayer network analysis, the following decisions should be made: how to represent the layers (3.1); how to order them (3.2), and how to design the between-layer edges (3.3).

3.1 Representation of the layers. When designing the network, first think about what the layers should represent. If your goal requires an edge-based network, different types of relationships should be placed on different layers and layers should be node-aligned (e.g. nodes can exist on multiple layers). In contrast, if your goal requires a node-based network, different subsets of the data should be placed on different layers and the network should be layer-disjoint.

3.2 Ordering of the layers. If the goal is to include hierarchy in the layers, based on theoretical considerations, then the layers have to be ordered accordingly. If not, you can skip this paragraph. This hierarchy can be based on timescales or for example part-whole grouping

in the data (Lazega et al., 2008). Part-whole grouping is for example when an individual is part of a whole, such as an organization. Processes on the individual level are then placed on a lower layer than processes on the organizational level. When hierarchy is based on timescales, variables are grouped together that fluctuate on the same timescale. Subsystems with faster processes are typically placed on lower layers and higher layers are typically used for subsystems of a 'higher order', meaning slower processes. Most studies that use hierarchy have a clearly visible hierarchy in their data already (Campi et al., 1997; Jeong et al., 2001; Noh, 2003; Ravasz & Barabási, 2003; Schuster et al., 2000; Wasserman & Faust, 1994). This is for example the case in biology when movement is being studied, where some layer represents processes within a single cell, the next layer represents processes within a muscle and the next layer represents processes on a behavioural level (Alves et al., 2018; Berenstein et al., 2016; Gomez et al., 2020). In this example, the hierarchy is based on both part-whole relationship and timescale, since processes on cell level within a muscle (part-whole) fluctuate on a faster timescale than processes on the muscle level. If hierarchy indeed suits the goal and the theory, the layers should be ordered according to the theory.

3.3. Design of the between-layer edges. When the representation (and, if required, the ordering) of the layers is chosen, the nodes can be assigned to the corresponding layers. The next step is then to choose whether between-layer edges should be connecting the same nodes on different layers (e.g. in edge-based networks) or different nodes (e.g. in node-based networks). After finishing this step, it should be possible to draw an empty version of the multilayer network.

Step 4: Deciding on the Appropriate Method for Analysis

Listing all different techniques that have been used in multilayer networks is out of the scope of this article. Whilst the field will undoubtedly develop further in the upcoming years, there are various decisions that have to be taken into account when performing your analysis.

Decisions concerning network sparsity. In networks, the number of possible edges grows when using larger datasets, which is often the case in multilayer networks. Therefore, a common used technique is the lasso, which provides sparser network structures (Friedman et al., 2008; Kivelä et al., 2014). The sparseness is regularized by a hyperparameter, sometimes called gamma. Unfortunately, the best value for this hyperparameter is a complex function of the true network structure, which is usually unknown (Epskamp & Fried, 2018). This makes the choice of the hyperparameter somewhat arbitrary and it can be set based on the choice whether the estimation should be more conservative (e.g. *gamma* = 0.5), possibly resulting in omitting true edges from the network (Foygel & Drton, 2010) or if the goal is to discover (e.g. *gamma* = 0), possibly admitting edges that do not exist (Dziak et al., 2020). Network sparsity using regularization can be used for estimation of within-layer edges as well as between-layer edges.

Connecting layers. Given that the new contribution of multilayer network analysis to network science is predominantly the differentiation between within- and between-layer edges, the main new methodological challenges concern the between-layer edges. However, literature has primarily focused on the connections within layers and neglected between-layer connection strength and the quantification of relative, weighted rates of information relationships between-layers (Min & Goh, 2013). Therefore there are still many questions remaining about what the best way is to connect layers with each other. This is especially the case for node-based networks, since estimating between-layer edges should be less complicated in edge-based, node-aligned networks, where the between-layer edges connect the same node on different layers (Kivelä et al., 2014).

Current approaches to connect layers primarily use similarity-based between-layer edges (Vaiana & Muldoon, 2020). Examples of those methods are for instance simply using correlations, the edge overlap measure (which estimates similarity and coherence between

layers and its average) and the strength-based interlayer Mutual Information (Eroglu et al., 2018; for comparison of these between-layer measures, see Hasselman, 2022).

Step 5: Interpretation

The last step that needs to be taken before performing the analysis, is deciding how to evaluate the research questions. Depending on whether the study is exploratory or confirmatory, the evaluation operationalization can be more, or less concrete. If the goal is to compare network models, than this should be done based on theory compatibility (Schmank et al., 2021) and the proposed data-generating mechanism (Van Bork et al., 2021) and should not be done using only goodness-of-fit indices (Simpson-Kent et al., 2021). One common used strategy is to compute centrality indices and clustering coefficients to interpret the network model. However, one question that should be asked is if it makes sense to consider correlations between variables (we highly recommend to read Bringmann et al. (2020) for a critical review on interpretation of centrality measures). For instance, psychological networks are strikingly different from networks studied in most other disciplines, because the edges are usually known (Wasserman & Faust, 1994), whereas those in psychological networks represent unknown statistical relationships that need to be estimated. This poses novel problems for statistical inference, which is why interpretations of the results need to be done with caution (Epskamp & Fried, 2018). However, when you decide to use the centrality indices and clustering coefficients anyway, both the accuracy and stability of the results should be examined before interpreting the results, which can be done using bootnet (Costantini et al., 2015; Epskamp et al., 2017).

Proof-of-Concept Study

After presenting this general step-by-step guide, we will now apply one specific method to an empirical dataset: a proof-of-concept study. This proof-of-concept study is based on a simulation study by Tio, Waldorp and Van Deun (2020). In this study, a multilayer

network design was used to combine different models from multiple sources to see how these interact. The authors explain that it is a necessity when working with multisource data, to take into account the underlying structure of associations. They argue that different sources may inherently differ from one another, which could bias the estimation of what the edges represent, which is essentially a measure of an association. Therefore, Tio et al. conceptualized networks based on multisource data as multi-layered and performed a simulation study where an underlying structure was written in the data. First the underlying structure was estimated and then this was taken into account when estimating the relationships using a regularized partial correlations. This resulted in a multilayer network model which was then compared to several monolayer network models, estimated from the same data. Tio et al. found that the multilayer network model better captured the true relationships that were written in the data than the monolayer network models. This indicates that if there are differences between the different sources, a multilayer network model should be used instead of a monolayer network model. However, this method has not been used on empirical data yet. Therefore it is unknown whether using this method on empirical data would lead to the same conclusions. Moreover, it is expected that problems might occur when applying this multilayer method to empirical data and it would be valuable to document what these challenges are and, if possible, how these can be dealt with. These questions will be addressed in the present proof-of-concept study.

Step 1: Goalsetting

In this proof-of-concept study, we want to examine if we have to take differences into account between within- and between-layer edges when using multidomain data. More specifically, we are interested in whether there is a potential component structure underlying the different empirical data sources we include (RQ1), and whether taking this into account, which results in a multilayer network, leads to different outcomes compared to monolayer

networks, estimated on the same dataset, that do not take this underlying structure into account (RQ2). Based on the simulation study of Tio et al. (2019), we hypothesize that the answer to both questions is yes (H1 and H2). Any conceptual and methodological challenges that arise during this application study will be documented and evaluated in the discussion.

Step 2: Choosing Variables

Ideally, as explained earlier, one would base the variables of the dataset on theoretical assumptions. However, for the goal of this study, namely to show how a node-based multilayer network can methodologically be created with psychological data, this is not as relevant because we are not interpreting the outcomes for theory-building or understanding of the phenomena. The reason for this is that we argue that we must withheld from theoretical interpretation as long as we have strong conceptual and methodological challenges and unknowns. Moreover, finding or collecting an ideal theory-based dataset was out of the scope and time of this study. We focused on finding a dataset with multiple sources or domains with at least psychological and physical or biological data with a connection to mental health, to demonstrate how these domains can be connected in multilayer networs. In the present study, we used a dataset with the domains mental health, psychical health and cognitive function from the Survey of Health, Ageing and Retirement in Europe project (SHARE-project; Börsch-Supan, 2022), Wave 7 (Bergmann et al., 2019). This data was collected from 77202 elderly people in 28 different countries, covering the whole European Union.

Exclusion of cases was based on practical reasons: for most subjects, not all scales of all these three domains were collected in Wave 7. We excluded these cases, which reduced the sample size to 13577 cases. The exclusion criteria of variables were informed by theoretical considerations: variables in the domains cognitive function and physical health were excluded if there was no reason to think it would be related to mental health. Furthermore, we concatenated some variables to change them from binary data to ordinal (e.g.

we concatenated yes/no answers on items of physical conditions). This was done because the matrix estimation needed for network analysis does not handle mixed data well (e.g. binary, ordinal and continuous data). Moreover, we used imputation to account for missing values (mice package; Van Buuren & Groothuis-Oudshoorn, 2011). After exclusions and concatenations, we had reduced the dataset to 22 variables. All variables, possible values and information about type and whether they were concatenated are listed in Appendix A. The script with all exclusions and operations on the data can be requested from the author.

Step 3: Multilayer Network Design

Our goal is already able to constrain some of our modelling decisions: to test our hypotheses, we conceptualized our data as divided based on the different domains (e.g. mental health, physical health or cognitive function) and placed every domain on another layer, which makes our network multilayer, node-based and layer-disjoint (step 3.1). We did not put hierarchy in our network because this was not the goal of this study and because there is no clear theory-based hierarchy visible in the data (step 3.2). Because our multilayer network is node-based, the connections between layers are automatically between different nodes and not between the same node on different layers, because this does not exist in our node-based model (step 3.3). Because one of our goals is to compare a multilayer network model with monolayer network models, we also estimate three monolayer network types that can be used for comparison.

Data standardization for multisource data can be done with Grand-mean standardization as well as Group-mean standardization. The Grand-mean standardization is the standardization method that is built in in the pre-process function in Regularized SCA described below. The Group-mean standardization is in literature for example being used when the interactions between groups are of interest. This method is also used in SNAC (Tio et al., 2020). Since between groups interactions (namely the different domains) is indeed our

goal, we think that the Group-mean standardization will suit our goal better than the Grand-mean standardization. For exploratory purposes, we decided to use both methods to see how this choice might influence the network structure. Therefore, all analyses described below are done twice: once with Grand-mean standardized data and once with Group-mean standardized data.

Step 4: Deciding on the Appropriate Method for Analysis

Monolayer Analysis Method

The first monolayer network is a correlation network that can be plotted from a correlation matrix. Although correlation networks can be very helpful and easy to interpret, spurious correlations might exist, for instance when two nodes share an edge with a third node. To control for this, partial correlation networks can be estimated, which is done in our second monolayer network model. This network controls for correlations with all other variables and therefore shows the conditional dependencies between the variables. The partial correlation coefficients can directly be estimated from the correlation matrix by taking the inverse of the matrix (Lauritzen, 1996).

Our third monolayer network model is the regularized (also called weighted) partial correlation network. To estimate this network, the graphical lasso (glasso; Friedman et al., 2019) is used. This lasso variant is specifically aimed at estimating partial correlation networks by inverting the sample variance-covariance matrix (Friedman et al., 2008). When estimating an edge between two nodes, this method controls for the relationships of these nodes with every other node in the network. This creates sparser networks, thereby decreasing the risk of potentially spurious connections by penalizing for more complex models and this enables interpretation of conditional dependencies between nodes (Epskamp & Fried, 2018). These regularization techniques have been shown to perform well in retrieving the true

network structure (Foygel & Drton, 2010; Friedman et al., 2008). This regularized partial correlation is also used in the SNACS function.

Multilayer Network Analysis

The technique used to estimate the multilayer network is based on the Sparse Network and Component function (SNAC) created by Tio et al., 2020. This two-step component-graphical model first performs a component analysis to measure between-layer relationships in data from 2 sources, before estimating the network. The aim in the publication by Tio et al. was to only connect models from different sources, therefore it does not estimate the within-layer relationships. We obtained the function from the original authors and made some adaptations to the function so it can handle more than two data sources and can also estimate within-layer relationships. We named this the Sparse Network and Component Supplemental function (SNACS). The script of this function can be found in Appendix B.

The function SNACS first estimates the component structure within the multilayer network and regenerates the data based on this structure, using Regularized SCA (version 0.5.4, Zhengguo & Van Deun, 2018). Next, the multilayer network is estimated using regularized partial correlation.

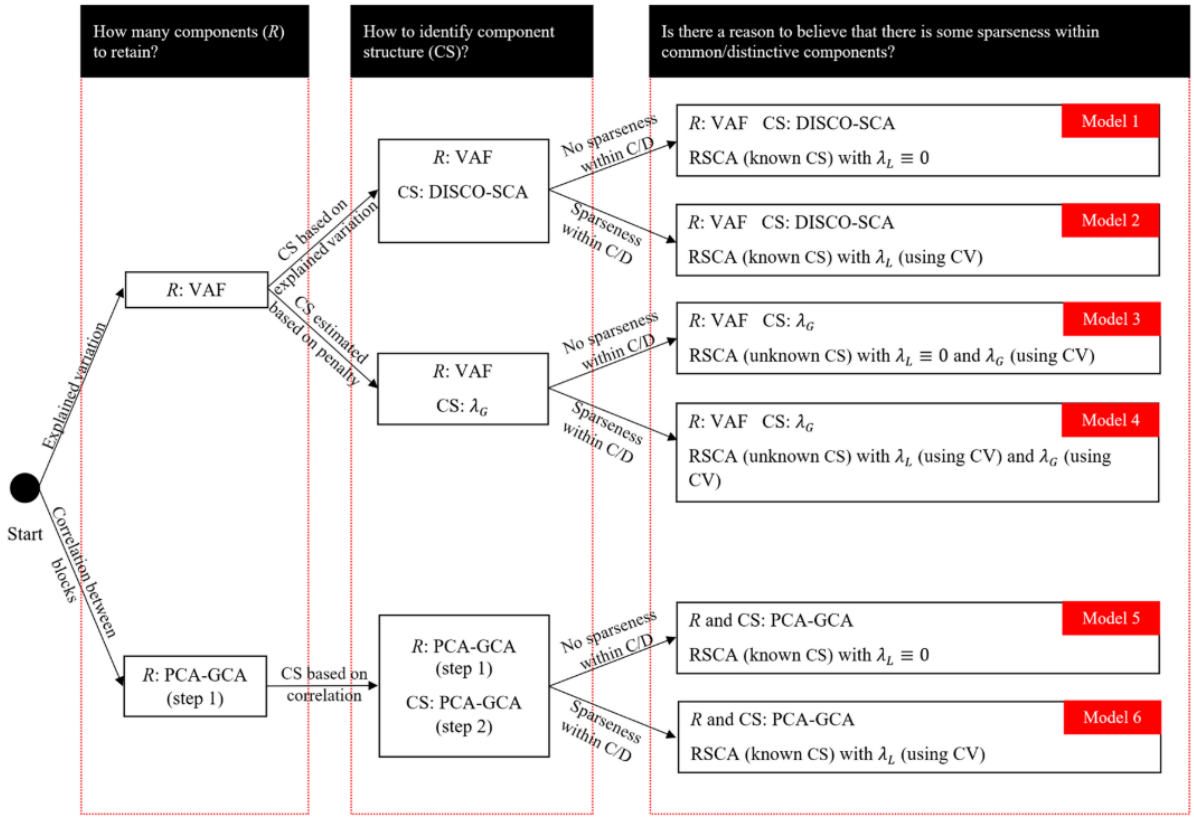
Regularized SCA. The core algorithm of Regularized SCA is based on simultaneous component analysis (SCA) which is a well-known method in psychometrics to integrate data (Smilde et al., 2017) and has already been used to combine data from multiple sources in biology as well as psychology (De Tayrac et al., 2009; Lock et al., 2013; Van Deun et al., 2013; Wilderjans et al., 2011). This method allows for identification of components for all sources. However, this method does not identify joint sources of variation that offer shared information across data sources and it does not identify unique variation that provides critical information on a subset of the data sources (Gu & Van Deun, 2019). This method can

therefore not be used when the goal is to only link a subset of specific variables of different sources. Moreover, the interpretation of components in SCA is based on all variables and therefore the interpretation of the results are more difficult (Van Deun et al., 2011). Because of these shortcomings, the package Regularized SCA has been developed that combines SCA with structured selection of variables. It allows to identify both the joint variation that is shared across the different data sources and the specific variation associated with one or a few of the data sources. Moreover, it is able to estimate component matrices with predefined structures, such as the structures of multiple data blocks (e.g. sources). This last option is the reason why this package is of interest for the estimation of multilayer networks.

In the Regularized SCA package used in SNACS, decisions have to be made within three steps before a model can be estimated. The different choice options lead to six different possible models (see Figure 5 for a flowchart for model selection in Regularized SCA; Gu & Van Deun, 2019). In step 1 it has to be decided whether to use either VAF or PCA-GCA to identify the number of components that need to be estimated. In step 2 the method for the component structure estimation needs to be chosen. In f the choice is to use the Group Lasso penalty, it also has to be decided whether the Group Lasso penalty needs to be imposed on the entire component matrix (block-wise method) or on every component of all data sources (component-wise method). In the experience of Gu and Van Deun (2019), the component-wise method is more useful in practice because the common and distinctive components are directly identified. However, they advise to use the block-wise method when users are not sure whether certain data blocks provide any information at all. In this method, entire blocks can be dropped from the analysis if they do not provide any information. In step 3 it has to be decided whether to achieve sparseness in the component structure.

Figure 5

Flowchart for Model Selection in Regularized SCA



Note. Model selection is based on 1) the measurement method of the amount of component; 2) if the component structure is known and 3) if you want sparseness within common and distinctive components. Note that CS stands for component structure, and C/D stands for common and distinctive components. Reproduced without permission from Gu & Van Deun (2019).

After the component analysis using Regularized SCA, SNACS can be used to regenerate the data based on the component structure. With this new data, a regularized partial correlation network is estimated in the same way as the monolayer regularized partial correlation network. After these two steps, SNACS is finished. The output of SNACS is a matrix that can be plotted using qgraph (Epskamp et al., 2012). This matrix can also be used to compute centrality indices and clustering coefficients for interpretation purposes, to see if the same nodes are central in the different network models.

Step 5: Interpretation

Centrality Indices

To determine what nodes are most important in the network structures, centrality indices will be computed. These indices give information about the patterns of the connections of the nodes as well as the tolerance of the network to the removal of specific nodes (Crucitti et al., 2004; Jeong et al., 2001). This can also be used as a guide for network interventions (Valente, 2012). We first explain the various indices. Degree centrality represents the number of connections a node has with other nodes (Freeman, 1979). The strength centrality is the generalization of the degree centrality to weighted networks by summing the weights of the connections of the node (Barrat et al., 2004; Newman, 2004). Another class of indices is based on the distance between nodes, which is defined as the length of the shortest path between two nodes (Brandes, 2001). Two common indices of this class are the closeness and betweenness centralities. The closeness centrality can be seen as the expected speed of arrival of something flowing through the network and the likelihood of a node to quickly affect other nodes or to be quickly affected by changes in other nodes (Borgatti, 2005). The closeness centrality represents the inverse of the sum of the distances of the focal node from all other nodes in the network (Sabidussi, 1966). The betweenness centrality is defined as the number of shortest path between other nodes, that travel through the focal node (Brandes, 2001; Freeman, 1979). This centrality assumes that if a node that is high in betweenness centrality is removed, the distance among other nodes will increase.

In the present study, we will focus on all these centrality indices to understand the patterns and important nodes in the different networks in the individual networks. Moreover, we can also use them to compare the networks, for instance if monolayer networks and multilayer networks have similar or different ordering of the centrality of the nodes. Clustering coefficients can serve the same purpose.

Clustering Coefficients

Clustering in networks is also very important for network structures and dynamics. The local clustering coefficient represents the number of connections among the neighbors (the directly connected nodes) of the focal node, divided by the maximum of possible connections among these neighbors (Watts & Strogatz, 1998). This coefficient can be interpreted as how redundant a node is, because removing a node when its neighbors have many interactions, will not make it much harder for the neighbors to interact (Latora et al., 2013). Unfortunately, the original clustering coefficients cannot take into account the edge weights of weighted networks but there are several new types that have been generalized to weighted networks (Saramäki et al., 2007), where Barrat and colleagues (2004) were the first to introduce this. Other types are the coefficient by Onnela and colleagues (2005) and one that is particularly suited for correlation networks (Kalna & Higham, 2007): the clustering coefficient by Zhang and Horvath (2005). All existing clustering coefficients can give different results. We refer to Saramäki and colleagues (2007) for a critical comparison.

Since the clustering coefficient of Watts and Strogatz (1998) cannot take into account the edge weights of our weighted networks, and the clustering coefficient of Zhang and Horvath is particularly suited for correlation networks, we will only focus on the coefficients proposed by Barrat and colleagues (2005) and Onnela and colleagues (2005).

Accuracy and Stability

Before interpreting the results of the analyses and indices, we will examine both the accuracy and stability of the results. Accuracy represents how prone the network estimation is to sampling variation and stability represents how similar interpretations remain with less observations (Epskamp et al., 2017). The stability of the outcomes increases with larger samples. To examine the stability of the results, the network estimation procedure and the centrality indices and clustering coefficients can be re-estimated using bootstrapping methods (Costantini et al., 2015; Epskamp et al., 2017).

Results

All statistical analyses and visualizations were performed using R (R Core Team, 2020), version 4.0.3 ("Bunny-Wunnies Freak Out"). Estimation of the networks was performed with the packages qgraph (Epskamp et al., 2012), igraph (Csardi & Nepusz, 2006) and graphical lasso (glasso; Friedman et al., 2019). Descriptive statistic of the data are listed in Appendix C.

Network Estimation

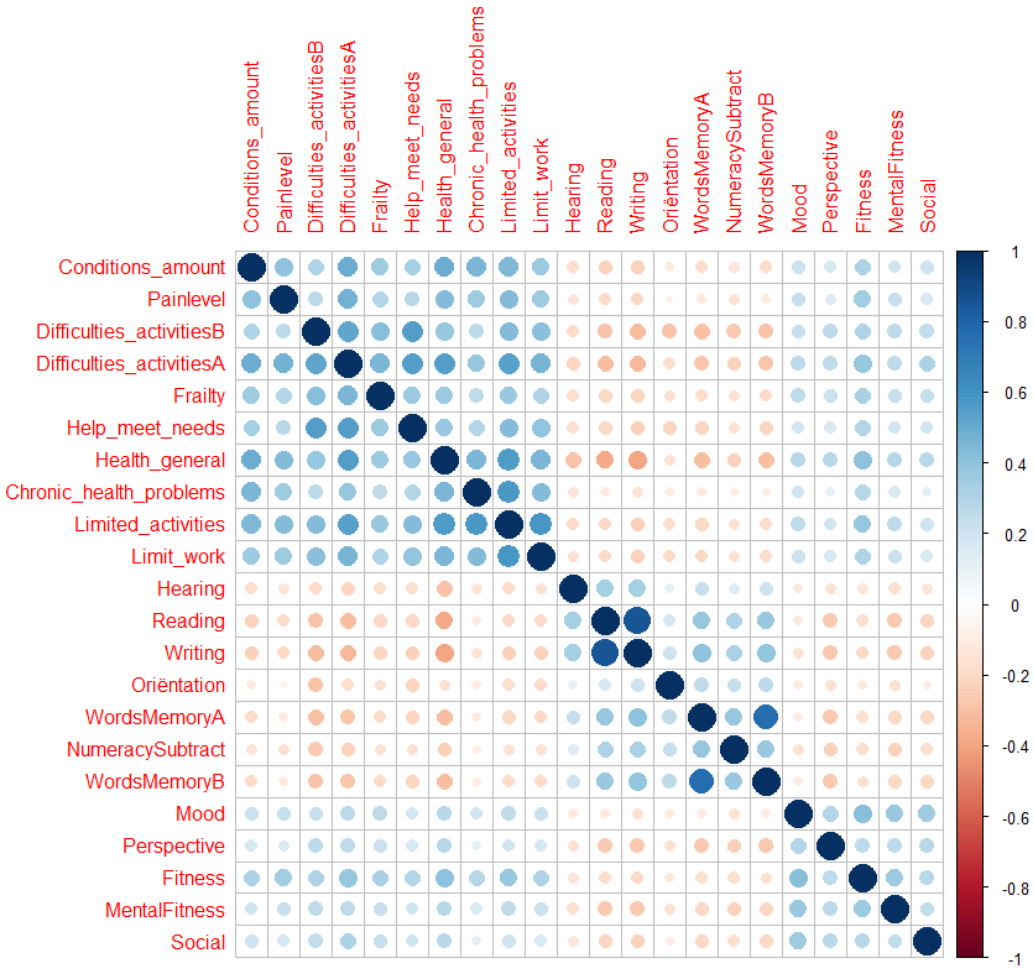
We started our network estimations by Group-mean standardizing the data and by Grand-mean standardization. The Grand-mean standardized data was produced using the pre-process function in Regularized SCA (Gu & Van Deun, 2019) that also weighs each block by taking into account the number of variables, since blocks can dominate each other when one block contains much more variables than another (Van Deun et al., 2009).

Monolayer Networks

Then we computed the correlation matrix of the data, see Figure 6. The figure indicates that the sources cluster together (e.g. three blocks of strong correlations are visible on the diagonal). This suits our assumption that data from different domains or sources might inherently differ from each other. This is something we hope to take into account in the multilayer network analysis. The matrices were the same for both the Grand-mean standardized and the Group-mean standardized data. This correlation matrix can directly be used to visualize the correlation network. From this matrix we also estimated the partial correlations (using package corpcor; Whittaker, 1990) and the weighted partial correlations (using glasso; Friedman et al., 2019) for both Group-mean and Grand-mean standardized data, which resulted in six network estimations.

Figure 6

Correlation Matrix of the Data



Note. Variables are ordered based on the source they belong to: the variables Conditions_amount to Limit_work belong to the source Physical health; the variables from Hearing to WordsMemoryB belong to the source Cognitive function and; variables Mood to Social belong to the source Mental health.

Multilayer Networks

We estimated two multilayer network models: model 1 with the Grand-mean standardized data and model 2 with the Group-mean standardized data. Because we do not have a priori knowledge or literature about the general component structure in our data, we need to use model 3 or model 4 of the Regularized SCA models, which are both meant for

estimation with unknown component structure. Because we want sparseness in our model, we want a lasso tuning parameter that is non-zero (step 3), which results in the choice of model 4.

The first step is to use the VAF method to identify the number of components. This method computes the proportion of VAF for each simultaneous component in each data block (Schouteden et al., 2013). In model 1, in the physical health block, the first two components explain most information (37.14% and 7.64%). In the cognitive function block, the first five components explain most information (27.49%, 21.07%, 1.63%, 11.02% and 8.54%) and in the mental health block, three components explain most information (25.84%, 2.16% and 22.28%). Taking these blocks together, we conclude that we need to retain five components for the analysis for model 1. In model 2, the first two components explain most information in the physical health block (22.67% and 37.37%). In the cognitive function block, the first three components explain most information (56.54%, 9.02% and 12.19%) and in the mental health block, first four components explain most information (24.42%, 10.62%, 1.65% and 30.83%). For model 2, we therefore conclude that we need to retain four components for the analysis.

In step 2, we used cross-validation and the Group Lasso penalty to identify the component structure. We used the block-wise method because we are not sure whether all data sources provide any information at all. The output of step 2 consists of the identified component structure as well as the best fitting Lasso and Group Lasso tuning parameters, which we use to run the models in step 3. In step 3 we again used cross-validation and we use both Group Lasso penalty and Lasso penalty to estimate the final Regularized SCA model. All (Group) Lasso penalties in step 2 and 3 are performed using 10-fold cross-validation with a sequence of 20 Lasso tuning parameters and 20 Group Lasso tuning parameters. We run model 1 with parameters λL = .215, λG = 1.53, and R = 5 and model 2 with parameters λL = .71, λG = 5.84, and R = 4.

To evaluate research question 1, we check for sparseness in both models in the estimated component loading matrix (Table 1 and Table 2). All components are common components, meaning that all components have at least some variables of every source that load on them. However, both model 1 and model 2 show some sparseness, which can be seen by the loadings of zero. This answers our first research question concerning whether there is a component structure underlying the different empirical data sources we include. The answer to this question is yes: although all components are common components, we see that there is some sparsity in the component loading matrix, indicating that there is indeed an underlying component structure in the data.

After finishing this last step of Regularized SCA, the first of two steps within SNACS was finished. We then performed the second and last step, wich was to use the function SNACS to regenerate the data based on the estimated component structure. Regularized partial correlations were estimated using graphical lasso. This results in the multilayer network structure which is ready to be visualized as a network.

Table 1

Grand-mean Standardized Re-estimated Non-shrinkage Component Loading Matrix

Variable	Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5
Amount of Conditions	0.30	7.09	-3.20	1.08	-14.82
Pain level	-0.23	5.28	-4.80	-0.51	-14.76
Difficulties Activities 1	1.17	11.94	-1.75	15.83	-2.71
Difficulties Activities 2	0.14	11.98	-3.80	9.85	-11.57
Frailty	2.07	8.55	-5.52	9.04	-6.69
Does Help Meet Needs	-0.38	8.89	-0.02	12.63	-8.02
Health General	1.39	11.55	-3.82	-0.64	-14.12

Chronic Health Problems	0.78	3.42	-0.72	0.36	-17.57
Limited Activities	0.51	7.38	-2.21	6.00	-16.65
Limited Work	1.27	6.67	-1.11	6.05	-14.45
Hearing	3.40	-13.82	-0.81	3.70	1.47
Reading	0.28	-21.28	0.00	4.41	-0.49
Writing	0.00	-21.32	0.00	3.75	0.00
Oriëntation	-7.83	-8.50	0.00	-12.41	-2.28
Words Memory	-17.66	-11.74	0.52	-0.97	1.19
Numeracy Substraction	-11.74	-9.25	3.76	0.00	0.37
Words Memory Delay	-17.90	-10.91	0.51	0.32	1.34
Mood	-0.58	4.14	-18.05	1.77	-3.03
Perspective	4.01	8.50	-12.22	2.02	0.00
Fitness	0.14	5.55	-13.94	1.65	-8.39
Mental Fitness	1.25	9.27	-12.78	1.61	-1.24
Social	1.59	7.04	-14.07	0.89	-1.52

Table 2

Group-mean Standardized Re-estimated Non-shrinkage Component Loading Matrix

Variable	**Comp. 1**	**Comp. 2**	**Comp. 3**	**Comp. 4**
Amount of Conditions	43.49	41.41	-0.75	-56.29
Pain level	-32.62	33.00	7.83	-29.20
Difficulties Activities 1	5.46	3.77	5.80	-106.51
Difficulties Activities 2	28.89	54.37	-0.42	-140.03
Frailty	-14.10	14.34	5.14	-20.28

Does Help Meet Needs	-45.36	25.75	0.00	-71.87
Health General	115.64	7.82	-6.61	-40.21
Chronic Health Problems	-62.00	17.63	11.32	-10.20
Limited Activities	-56.99	23.51	11.36	-25.92
Limited Work	-78.92	18.06	16.35	-11.74
Hearing	-39.40	0.00	-9.53	19.40
Reading	-19.85	-2.78	-37.00	32.08
Writing	-29.83	-1.57	-36.59	35.88
Oriëntation	0.00	3.09	-11.19	16.73
Words Memory	83.56	5.87	-107.73	52.05
Numeracy Substraction	-43.74	15.27	-91.06	38.49
Words Memory Delay	144.60	-3.39	-124.48	45.22
Mood	-6.66	42.79	15.89	-10.51
Perspective	-55.80	16.39	21.52	-6.17
Fitness	-24.32	27.35	12.34	-16.85
Mental Fitness	-39.79	19.72	21.56	-12.01
Social	24.15	96.96	48.69	-5.95

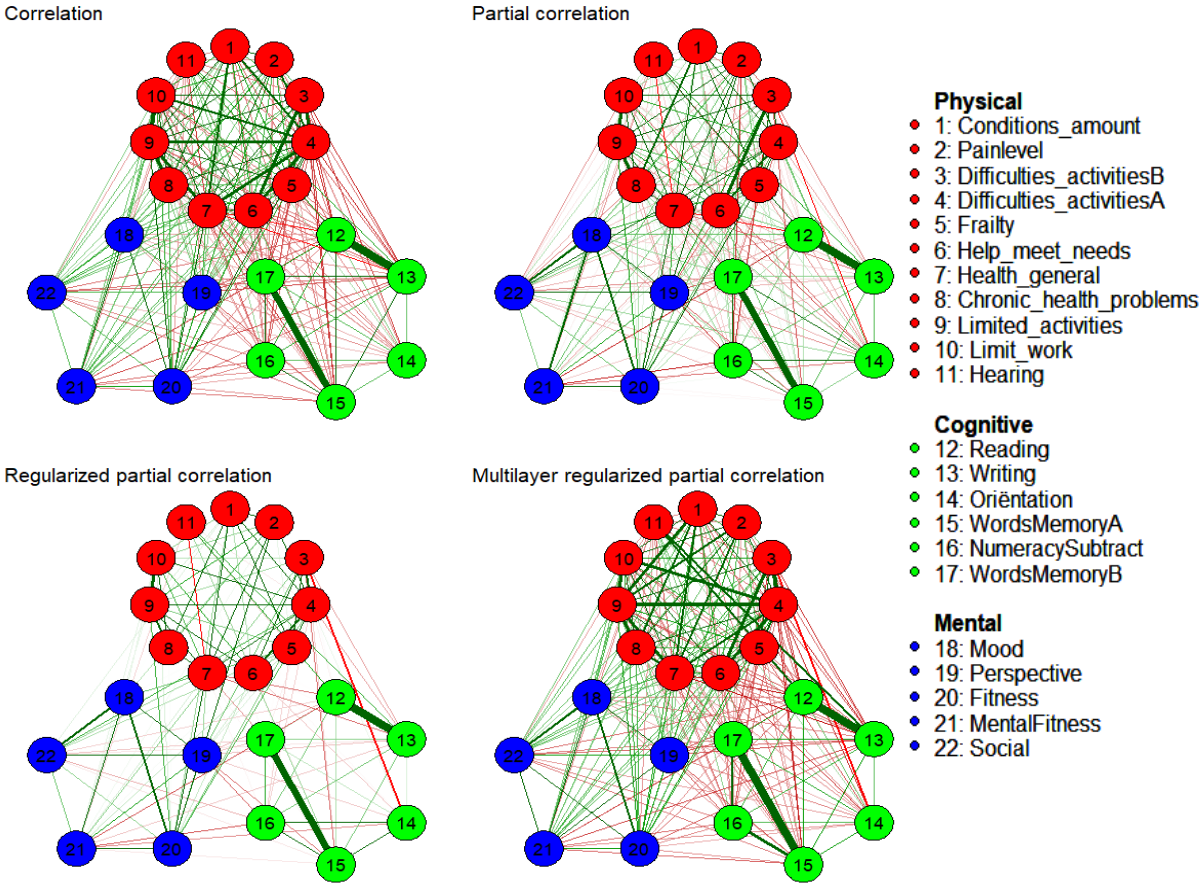
Network Visualization

After finishing all network estimations, we used the package qgraph (Epskamp et al., 2012) to visualize the networks. The matrices with weighted partial correlations as well as the multilayer networks were not positive definite so we forced symmetry (Matrix package; Bates & Maechler, 2021). To facilitate comparison and easier interpretation, we grouped the data together based on source and gave every source a distinctive colour. For the networks created with Grand-mean standardized data, see Figure 7. The networks estimated with the Group-

mean standardized data can be found in Figure 8. Positive associations are green, whereas negative associations are red and the thicker the line, the stronger the association. The figures of the networks indicate that the standardization method has a big impact on the network estimations, especially in the multilayer networks. To evaluate and compare the network structures in more detail, we now compute the centrality indices and clustering coefficients.

Figure 7

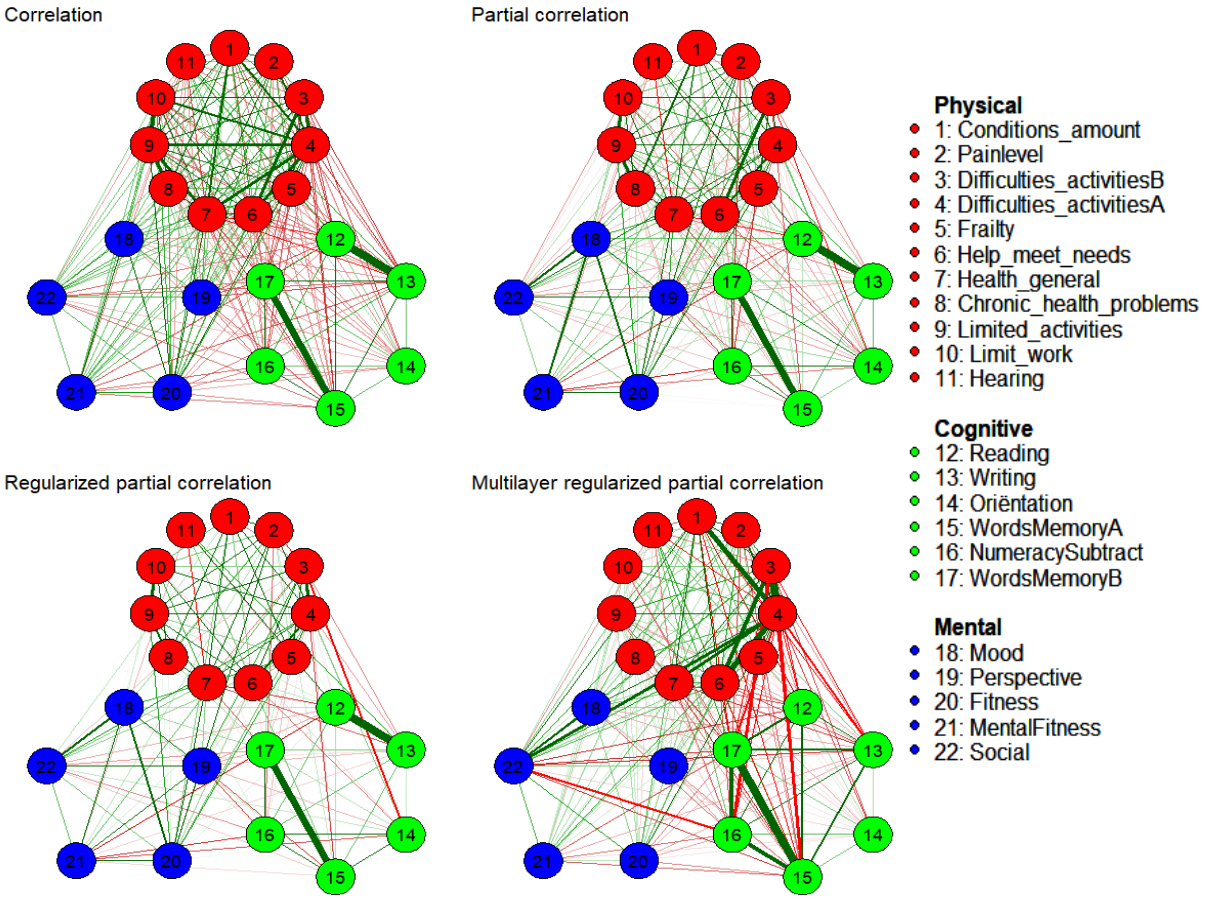
Grand-mean Standardized Networks



Note. The multilayer network looks similar to the correlation network.

Figure 8

Group-mean Standardized Networks



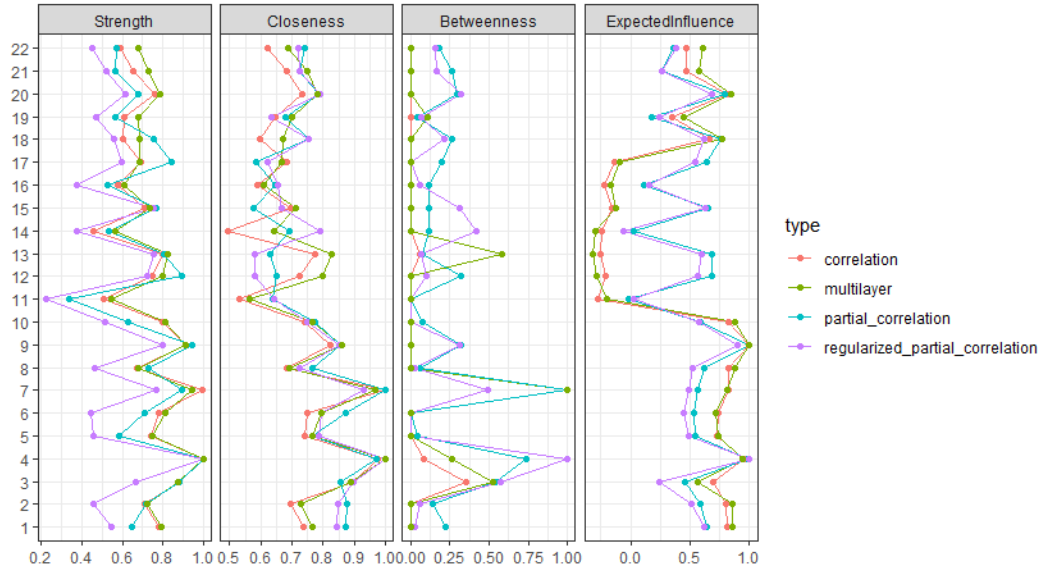
Note. The multilayer network looks different than the other networks.

Centrality Indices and Clustering Coefficient

To compare the different network models, we computed centrality indices and clustering coefficients using the function centralityPlot from package qgraph (Epskamp et al., 2012). In the centrality plot, we focused on the coefficients of Barrat and colleagues (2005) and Onnela and colleagues (2005). All centrality indices and clustering coefficients reflect the importance of all nodes, with higher values indicating that nodes are more important. Figure 9 shows the centrality and Figure 10 the clustering of all four Grand-mean standardized networks. For the Group-mean standardized networks, Figure 11 shows the centrality indices and Figure 12 the clustering coefficients. Before we interpret them, we want to know how stable and accurate the results are.

Figure 9

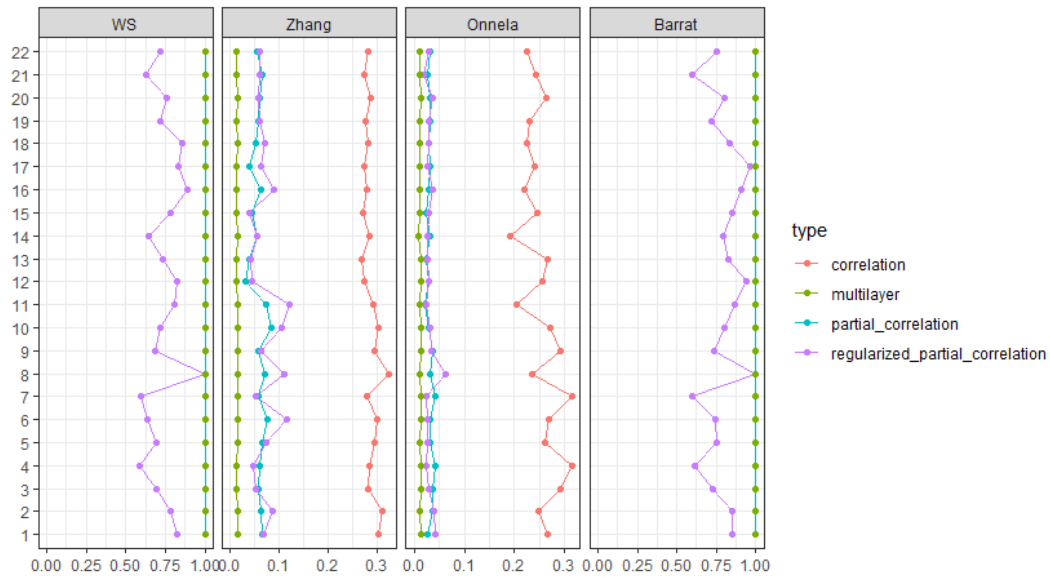
Grand-mean Centrality



Note. Different indices of relative centrality. Note that the centrality of the nodes in the multilayer network are most similar to the correlation network.

Figure 10

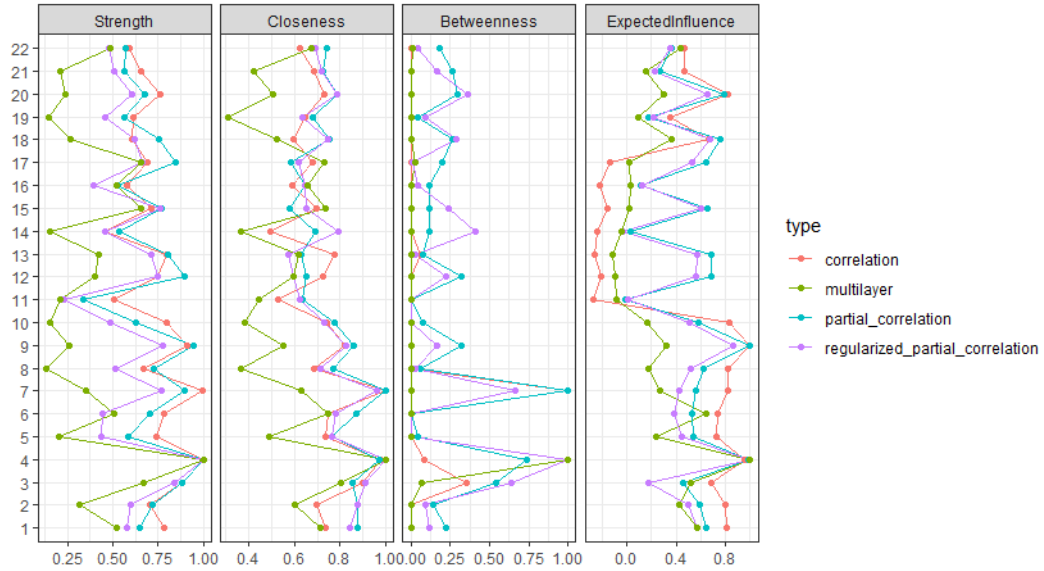
Grand-mean Clustering



Note. Different clustering coefficients. Note that none of them show variation in clustering of all networks and that Onnela and Barrat show almost the opposite results.

Figure 11

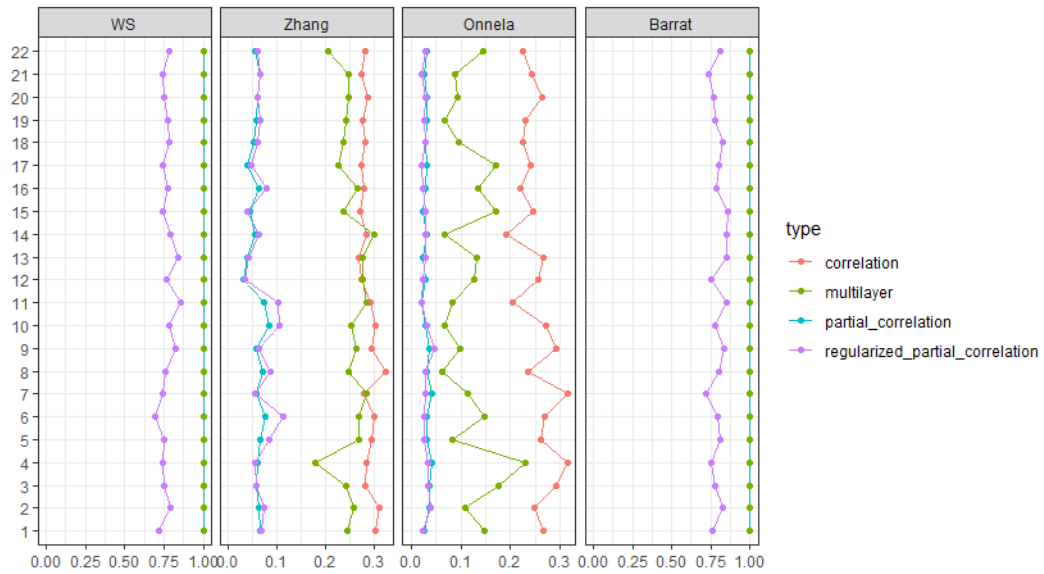
Group-mean Centrality



Note. Different indices of centrality. Note that the central nodes differ between networks, especially in the multilayer network.

Figure 12

Group-mean Clustering



Note. Different clustering coefficients. Note that Onnela and Barrat again show completely different results, making it hard to interpret the clustering of the networks.

Accuracy and Stability

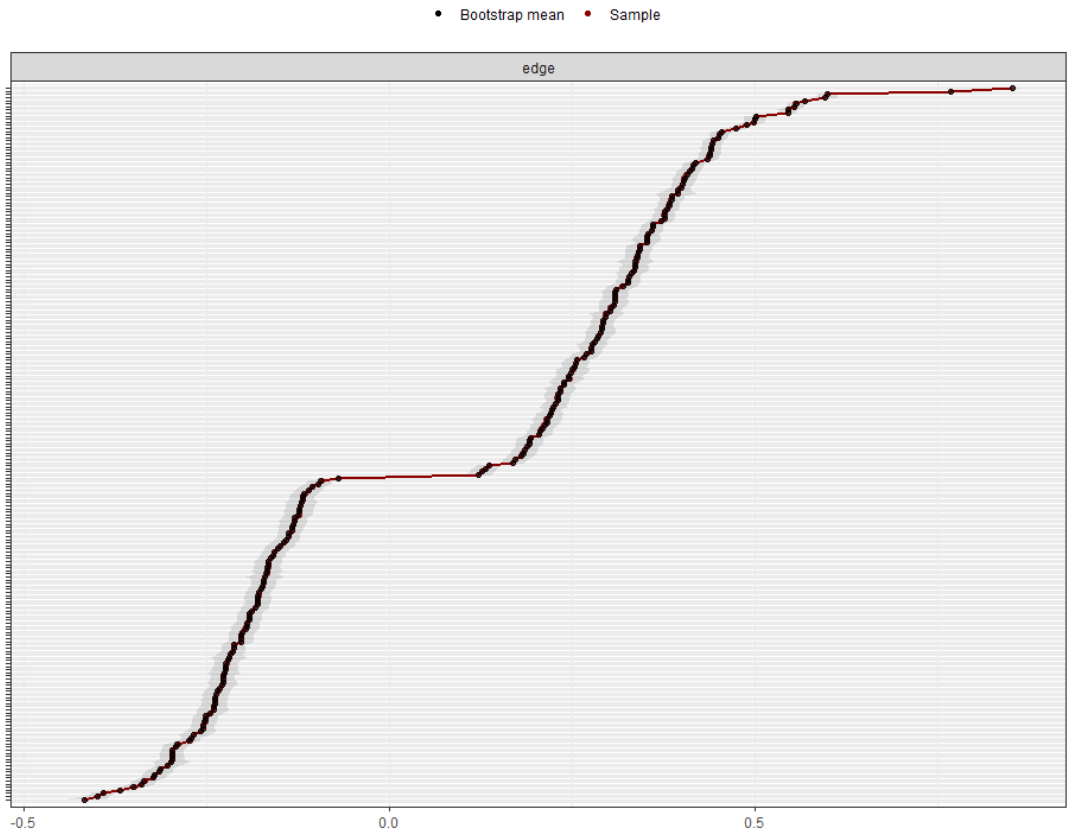
The bootnet function of the bootnet package (Epskamp et al., 2018) was used to perform all accuracy and stability analyses. All plots were extremely similar, if not the same, for both the Grand-mean standardized and the Group-mean standardized data. We therefore only show the plots for the Group-mean standardized data. Unfortunately, our multilayer network matrices are not automatically positive definite. Throughout the network estimations, we forced them to be symmetric, which made the matrices positive definite. The functions for the accuracy plots however do not work with matrices bus with the data. Therefore, we cannot use those functions for the multilayer networks. However, the accuracy and stability are partly relying on the sample size, which is the same for all estimated networks in this study. Therefore, the results of these plots on the correlation matrix might be an indication of the stability and accuracy of the other networks as well.

Edge-weight Accuracy

We assessed edge-weight accuracy using the plot method to show bootstrapped CIs for the estimated edge weights (Figure 13). The y-axis represents all 231 edges. The names of the nodes they connect are left out, because this would have been unreadable and not important for interpretation of the plot. The blue and red lines in the plot represent the estimated edges and the bootstrap means of the edges. Because these are extremely similar, in this plot there is only one line visible. This indicates that the estimation of the edge-weights is very accurate. The grey area around the line in the plot consists of the bootstrapped CIs. Because this line is quite small, this indicates that many edge-weights likely significantly differ from one-another. This implies that the interpretation of the order of most edges will be accurate.

Figure 13

Edge-weight Accuracy



Note. The y-axis represents the edge-weights, ordered from highest to lowest edge-weight. The grey area around the line in the plot consists of the bootstrapped CIs. Non-overlapping CIs indicate two statistics significantly differ at significance level .95.

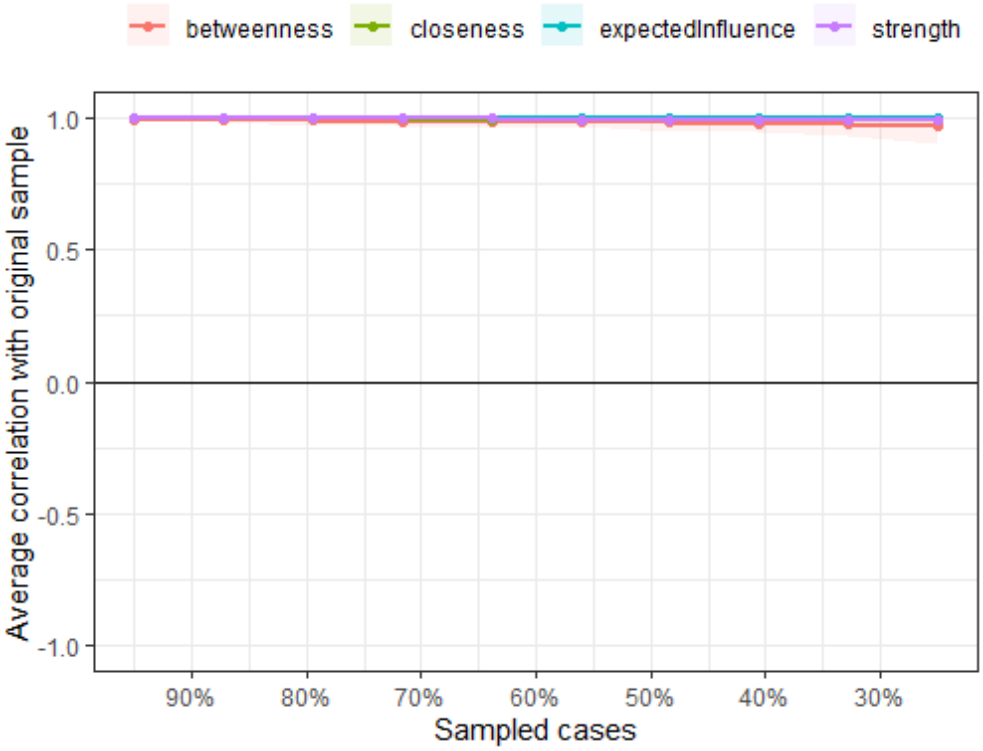
Centrality Stability

To examine the stability of centrality indices we again perform a bootstrap but this time by using subsets of the data. The resulting plot (Figure 14) shows that the lines do not drop, which indicates very high stability of all indices. The stability is quantified using the CS-coefficient, which quantifies the maximum proportion of cases that can be dropped while still retaining a correlation with the original centrality of higher than .7, with a 95% certainty. The cut-off value for stability is .5. The CS-coefficient for betweenness, closeness and

strength are all .75, indicating that they are all stable under subsetting cases. We can therefore conclude that all indices are stable enough to interpret.

Figure 14

Centrality Stability



Note. The areas around the lines indicate the range from the 2.5th quantile to the 97.5th quantile. Note that the lines and the areas around the lines barely drop, indicating the indices are stable.

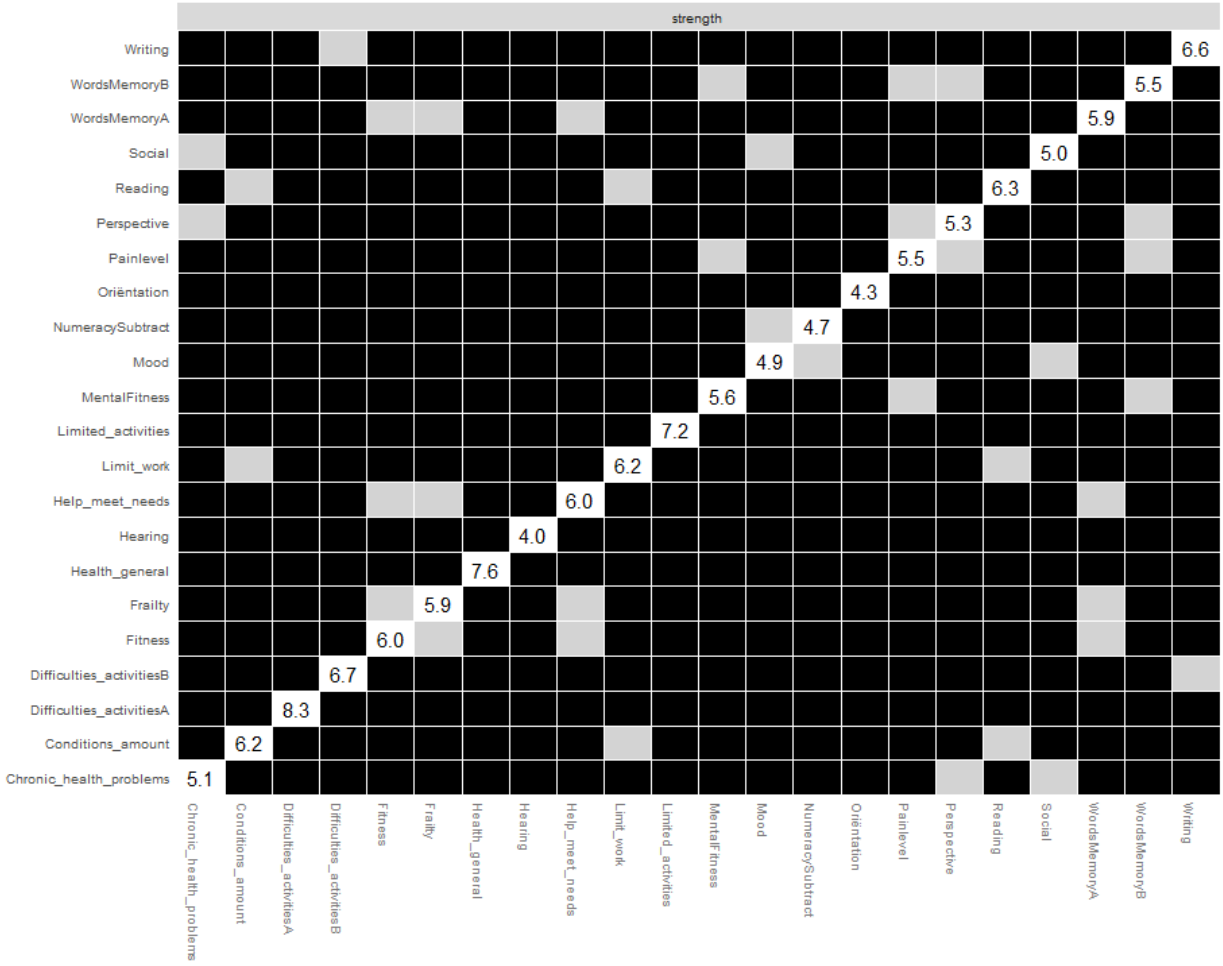
Testing Significant Differences

The difference tests of node strengths (Figure 15) and edge-weights (Figure 16) are plotted using the non-parametric bootstrap results. In both plots, black boxes represent significant differences and grey boxes non-significant differences. The white boxes on the diagonal in Figure 15 show the values of the node strengths. The results show many black

boxes, indicating that most nodes as well as edge-weights significantly differ from one-another.

Figure 15

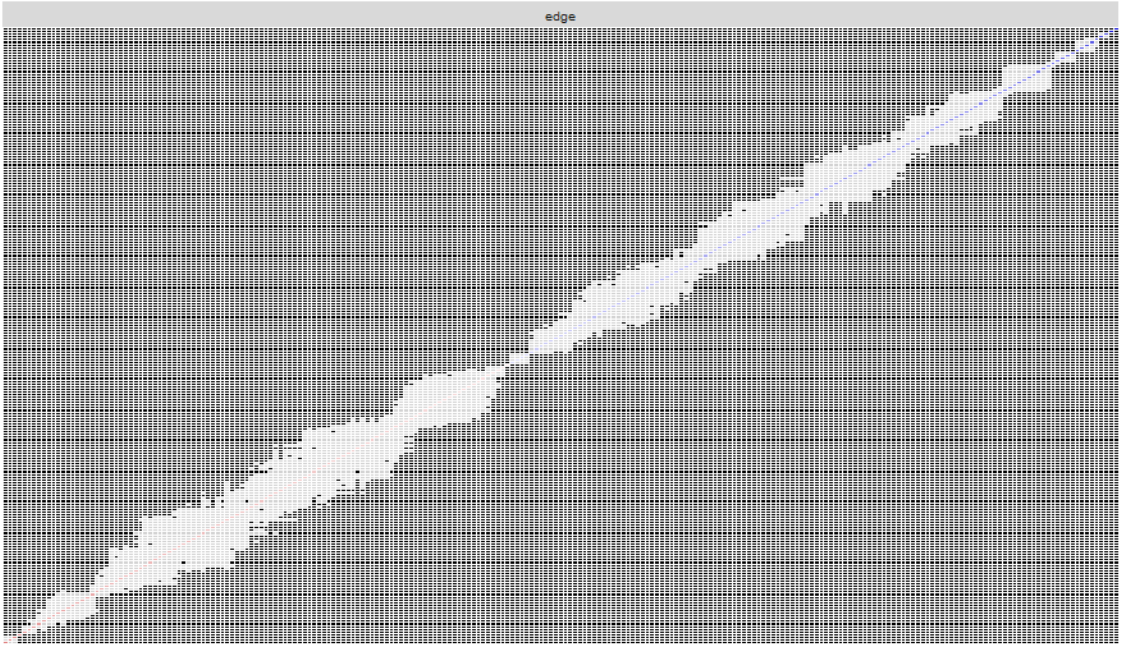
Significant Difference Node-strengths



Note. Bootstrapped difference tests (α = 0.05) between node strengths. White boxes show the value of node strength. Black boxes indicate that nodes are significant different from one-another and gray boxes indicate nodes that do not significantly differ from one-another.

Figure 16

Significant Difference Edge-weigths



Note. Bootstrapped difference tests (α = 0.05) between all non-zero edge-weights. Black boxes indicate that edges are significant different from one-another and gray boxes indicate edges do not significantly differ from one-another.

Interpretation of the Network Estimations

Now that we can (with caution) interpret the results, we can evaluate the results in order to answer the second research question, which entails whether the multilayer networks are different form the monolayer networks in this study. All monolayer network plots look as expected: the correlation network has the most edges, the partial correlation model has less edges and the regularized partial correlation model is even sparser. However, the multilayer network model for the Grand-mean standardized data gave unexpected results. The multilayer network model does not look sparser than any of the monolayer network models and even looks almost exactly the same as the correlation network. The centrality indices show the same pattern: the multilayer network model does not look extremely different from the other

networks and is most similar to the correlation network. Interestingly, the clustering coefficient show another pattern: when looking at Onnela's coefficient, the correlation network seems to stand out while all other networks have a clustering coefficient of (almost) zero for all nodes. Barrat's coefficient shows a different pattern, where all networks except for the regularized partial correlation network show a coefficient of one for all nodes. However, both coefficient do not seem to be able to compare clustering in all different networks at once, which might be the result of large differences in the network structures. Therefore we expect these patterns are not reliable.

In contrast to the Grand-mean multilayer network, the Group-mean standardized multilayer network does show completely different results than the monolayer networks, which aligns with what we expected. This indicates that taking into account the underlying component structure does indeed lead to different network structures, when standardizing the data on Group-mean level. The centrality indices also indicate that all networks have somewhat different structures, where the multilayer network shows the most distinctive patterns. Just as with the Grand-mean standardized data, the clustering coefficients for these Group-mean standardized networks do not seem to be reliable, although Onnela's as well as Zhang's coefficients seem to be able to catch at least some variation in all networks.

In conclusion, the result indicate that the standardization method is an important choice when constructing a multilayer network model. Taking into account the underlying component structure of Group-mean standardized data in a multilayer network format using SNACS, results in a network that is different from monolayer networks (RQ2), while a multilayer network of Grand-mean standardized data is not much different than a monolayer (correlation) network.

Discussion

Now that we primarily focused on the possibilities of multilayer network analysis by providing information on how to design a multilayer network study and by performing a proof-of-concept study, we arrive at the last part of this study: discussing the conceptual and methodological challenges and providing suggestions for future research. We first explore challenges based on the implementation study and then we assess conceptual issues.

Challenges Based on the Implementation Study

An important limitation of the study concerning the methods Regularized SCA and therefore also both SNAC and SNACS, is that we found that outcomes heavily depend on small differences. For instance, differences in one variable in the dataset and even rounding of the lasso tuning and group lasso tuning parameter settings have led to very different outcomes in both lasso parameters (in the case of a change in one variable), component estimation, matrix estimation and therefore also on the network structures. This means that the results are equally unstable and therefore possibly unreliable. It would therefore be valuable if future research evaluates the stability, reliability and suitability of cross-validation and the Regularized SCA method and explore other options such as canonical analysis (Tenenhaus & Tenenhaus, 2014) and the Index of Sparseness (Gajjar et al., 2017; Trendafilov, 2014; Zou et al., 2006). The Index of Sparseness for instance has shown to outperform the cross-validation method when estimating component loadings (Chen & Chen, 2008). It would be interesting to see if these methods are more redundant against rounding in parameter settings and small changes in the data, when estimating the component structures for multilayer models.

Three other methodological limitations of this study impacted the interpretability. First, we needed to concatenate several data items before we could use them for this study, which constrains the interpretability of all individual items and might have altered the results. Furthermore, we could only assess the accuracy and stability of the correlation network model and not the other network models, since these data matrices were not positive definite and

forcing symmetrie was not possible in the function to assess accuracy and stability. Moreover, the clustering coefficients do not seem to be able to reliably compare clustering in all monolayer networks as well as multilayer networks. This might be the result of the large differences between the networks. Hence, further research is needed to develop clustering coefficient that can be used for this purpose and research is needed to know how non-positive definite data can be assessed on stability and accuracy.

Another challenge concerning the interpretablity of the multilayer network is of a conceptual kind, namely how the relationships between the edges should be interpreted. Take for example the plausible strong positive edge between the nodes reading and writing from the Cognitive function domain. This relation is almost zero in the multilayer network of the Group-mean standardized data, while we assume, based on theory and reasoning, that the relationship does exist in the true world (e.g. Fitzgerald & Shanahan, 2000). The question arises how this multilayer network then should be interpreted, and what information should and should not be extracted from the multilayer network. Moreover, this also raises questions about the method SNACS itself. It does account for the underlying component structure but it results in a network that loses some strong relations, making it possibly useless for certain goals of multilayer network analysis, for instance when one simply wants to understand behaviour and relationships from variables between and withing different sources. It would be valuable if follow-up studies on interpretability would not only focus on methodology but also on more conceptual questions such as how (not) to interpret multilayer networks.

Moreover, if the actual goal of a multilayer network model is to increase understanding of relationships within sources as well as to increase understanding of how these sources interact with each other, it might be better to keep their within-layer edges intact. This can be done by only accounting for differences between sources when estimating the between-layer edges but not when estimating the within-layer edges. This way, the edges

within the sources are not disturbed by taking into account the underlying component structure. This is actually very similar to what Tio, Waldorp and Van Deun did with SNAC: combining models from different disciplines, without estimating the edges within the disciplines. This approach might also increase interpretability and it would be interesting to compare it to the current SNACS function, using simulation data. This way, the actual structure of the data is known and models can be evaluated on what best fits the real structure.

Conceptual Challenges

Throughout the article, we assumed that the multisource data should be placed on different layers based on the source they belong to. The reasoning behind this is that data from different sources might be inherently different, which leads to clusters in the data based on the source they belong to. For instance because their constructs/measurement instruments/timescales might be more similar as opposed to those of other sources. The assumed clustering is therefore theory-based. In our proof-of-concept study, it turned out to be true that the data was clustered based on source. However, it might also be the case that clustering in empirical data happens to be cross-source, for instance when data sources are more similar. In those cases, it makes more sense to divide data not based on theory (e.g. the different data sources) but on the actual clustering in the data, which would be a data driven approach. Currently, we do not know if this would lead to different network structures and we do not know which of these two approaches would be better for different goals. To assess these questions, a study that compares the data-driven versus theory-driven approach in the step where nodes are assigned to layers, would benefit the debate on this topic. This can first be done in a simulation study, where different structures can be written in the data (e.g. with the factoextra package; Kassambara & Mundt, 2020).

A more fundamental problem concerns the use of psychological data for network estimations. For instance, many disciplines that work with network models use connections

that exist in the real world, while relationships in psychological research are estimated: they are abstract representations of relationships between concepts that are restricted by how we measure them. Therefore we do not know if a relationship we find in the model truly exists, since it can instead reflect something else, for example the similarity between the operationalization of the psychological concepts. Furthermore, sample sizes in psychology are often limited and yet we do not know how this might affect the results (Epskamp & Fried, 2018). Both problems make it necessary to be cautious when interpreting and reporting network structures and especially when planning to use the results as a base for interventions.

Moreover, in this article we mentioned that the results of a multilayer network model strongly depend on theoretical considerations that inform, amongst others, variable choice and model design. However, these theoretical considerations are not always straightforward when it comes to psychological phenomena, where the concepts we use are not as well defined as in other disciplines that make use of multilayer networks. This also affects decisions concerning on what layer to put the different sources. This again underlines the importance of being cautious when interpreting the network structure.

Summary and Conclusion

In this article, we explored the possibilities of constructing a multilayer network using empirical, multisource data in psychological science. We created an overview of choices that need to be made to construct and evaluate a multilayer network and then performed a proof-of-concept study where we used SNACS model estimation. This method creates a two-step component-graphical model by estimating the underlying component structure and then taking this into account when estimating the network model. Based on this implementation, we documented challenges of this particular implementation as well as multilayer network analysis in psychology in general and provided suggestions for future research.

In line with our hypotheses, the proof-of-concept study showed that an underlying component structure was found in the multisource data and that taking this structure into account, using a multilayer network design, leads to different outcomes compared to monolayer network models where this underlying structure was not taken into account, when the data was Group-mean standardized. This indicates that taking into account the underlying component structure might be very important and maybe even a necessity when examining multisource data. These findings serve as a substantiation for the argument of Tio, Waldorp and Van Deun (2020) that constructing a monolayer network using multisource data might result in biased estimations because of a possible underlying structure that is not being taken into account.

However, as discussed, the results heavily depend on the modelling decisions made, hampering the robustness of the findings. Moreover, the interpretability of the multilayer network structures are difficult and need to be done with caution. Both limitations indicate that the resulting network structures may not be ready to be used for inferences about psychological phenomena yet and instead, further research needs to be done to a) develop methods that are more stable and b) increase interpretability (see discussion for suggestions).

Insights gained in this study are a first step towards understanding the use, possibilities and difficulties of constructing multilayer network models of psychological phenomena and constitute to the theoretical debate on this topic. Furthermore, the documentation in this study can serve as a starting point and guideline for psychological researchers that are planning on using multilayer network models in their studies. Moreover, this line of research is promising for implementation in practice to understand patterns in behaviour or psychopathology (see e.g. Borsboom & Cramer, 2013; Braun et al., 2018; De Boer et al., 2021; Nelson et al., 2017), although we argue based on the present study that we first need to increase our understanding of and provide solutions for the methodological and conceptual challenges explained in this

article. In conclusion, using component analysis to estimate multilayer network models for multisource data seems to be a promising approach to further develop.

Acknowledgements

This paper uses data from SHARE Wave 7 (DOI: 10.6103/SHARE.w7.800), see Börsch-Supan et al. (2013) for methodological details. The SHARE data collection has been funded by the European Commission, DG RTD through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812), FP7 (SHARE-PREP: GA N°211909, SHARE-LEAP: GA N°227822, SHARE M4: GA N°261982, DASISH: GA N°283646) and Horizon 2020 (SHARE-DEV3: GA N°676536, SHARE-COHESION: GA N°870628, SERISS: GA N°654221, SSHOC: GA N°823782, SHARE-COVID19: GA N°101015924) and by DG Employment, Social Affairs & Inclusion through VS 2015/0195, VS 2016/0135, VS 2018/0285, VS 2019/0332, and VS 2020/0313. Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C, RAG052527A) and from various national funding sources is gratefully acknowledged (see www.share-project.org).

References

Aleta, A., Meloni, S., & Moreno, Y. (2017). A Multilayer perspective for the analysis of urban transportation systems. *Scientific Reports*, *7*(July 2016), 1–9. https://doi.org/10.1038/srep44359

Alves, L. G. A., Mangioni, G., Rodrigues, F. A., Panzarasa, P., & Moreno, Y. (2018). Unfolding the complexity of the global value chain: Strength and entropy in the single-layer, multiplex, and multi-layer international trade networks. In *Entropy* (Vol. 20, Issue 12). https://doi.org/10.3390/e20120909

Barabási, A. L. (2016). Network science. In *Network Science*. Cambridge University Press. https://doi.org/10.1109/MNET.2010.5634435

Barrat, A., Barthélemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(11), 3747–3752. https://doi.org/10.1073/pnas.0400087101

Bartel, J., Krumsiek, J., & Theis, F. J. (2013). Statistical methods for the analysis of high-throughput metabolomics data. *Computational and Structural Biotechnology Journal*, *4*(5), 1–9. https://doi.org/10.5936/csbj.201301009

Bates, D., & Maechler, M. (2021). *Matrix: Sparse and Dense Matrix Classes and Methods* (R package version 1.3-2). Article R package version 1.3-2. https://cran.r-project.org/package=Matrix

Benthem de Grave, R., Hasselman, F., & Bijleveld, E. (2020). *From work stress to disease: A computational model Remco*.

Berenstein, A. J., Magariños, M. P., Chernomoretz, A., & Agüero, F. (2016). A Multilayer Network Approach for Guiding Drug Repositioning in Neglected Diseases. *PLoS*

Neglected Tropical Diseases, *10*(1), 1–33. https://doi.org/10.1371/journal.pntd.0004300

Bergmann, M., Scherpenzeel, A., Börsch-Supan, A., & (Eds.). (2019). *Wave 7 Methodology: Panel Innovations and Life Histories.* Munich: MEA, Max Planck Institute for Social Law and Social Policy.

Bianconi, G. (2018). Multilayer Networks. Structure and Function. In *Oxford University Press*.

Blanken, T. F., Bathelt, J., Deserno, M. K., Voge, L., Borsboom, D., & Douw, L. (2021). Connecting brain and behavior in clinical neuroscience: A network approach. *Neuroscience and Biobehavioral Reviews*, *130*(April), 81–90. https://doi.org/10.1016/j.neubiorev.2021.07.027

Boccaletti, S., Bianconi, G., Criado, R., del Genio, C. I., Gómez-Gardeñes, J., Romance, M., Sendiña-Nadal, I., Wang, Z., & Zanin, M. (2014). The structure and dynamics of multilayer networks. *Physics Reports*, *544*(1), 1–122. https://doi.org/10.1016/j.physrep.2014.07.001

Borgatti, S. P. (2005). Centrality and network flow. *Social Networks*, *27*(1), 55–71. https://doi.org/10.1016/j.socnet.2004.11.008

Borsboom, D. (2017). A network theory of mental disorders. In *World Psychiatry* (Vol. 16, Issue 1, pp. 5–13). https://doi.org/10.1016/j.amp.2020.11.014

Borsboom, D., & Cramer, A. O. J. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, *9*, 91–121. https://doi.org/10.1146/annurev-clinpsy-050212-185608

Börsch-Supan, A. (2022). *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 7. Release version: 8.0.0. SHARE-ERIC. Data set.*

https://doi.org/10.6103/SHARE.w7.800

Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S., & Zuber, S. (2013). Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). *International Journal of Epidemiology*, *42*(4), 992–1001. https://doi.org/10.1093/ije/dyt088

Brandes, U. (2001). A Faster Algorithm for Betweenness Centrality. *Journal of Mathematical Sociology*, *25*(2), 163–177. http://perso.ens-lyon.fr/christophe.crespelle/enseignements/GGT/centralite.pdf%5Cnhttp://kops.ub.uni-konstanz.de/bitstream/handle/urn:nbn:de:bsz:352-opus-71888/algorithm.pdf?sequence=1

Braun, U., Schaefer, A., Betzel, R. F., Tost, H., Meyer-Lindenberg, A., & Bassett, D. S. (2018). From Maps to Multi-dimensional Network Mechanisms of Mental Disorders. *Neuron*, *97*(1), 14–31. https://doi.org/10.1016/j.neuron.2017.11.007

Bringmann, L. F., Epskamp, S., Krause, R. W., Schoch, D., Wichers, M., & Wigman, J. T. W. (2020). What do centrality measures measure in psychological networks ? *Journal of Abnormal Psychology*, *128*(8), 892–903.

Brooks, D., Hulst, H. E., de Bruin, L., Glas, G., Geurts, J. J. G., & Douw, L. (2020). The multilayer network approach in the study of personality neuroscience. *Brain Sciences*, *10*(12), 1–15. https://doi.org/10.3390/brainsci10120915

Brummitt, C. D., D'Souza, R. M., & Leicht, E. A. (2012). Suppressing cascades of load in interdependent networks. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(12), 1–22. https://doi.org/10.1073/pnas.1110586109

Buccafurri, F., Foti, V. D., Lax, G., Nocera, A., & Ursino, D. (2013). Bridge analysis in a social internetworking scenario. *Information Sciences*, *224*, 1–18.

https://doi.org/10.1016/j.ins.2012.10.021

Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E., & Havlin, S. (2010). Catastrophic cascade of failures in interdependent networks. *Nature*, *464*(7291), 1025–1028. https://doi.org/10.1038/nature08932

Campi, R., Amparore, D., Checcucci, E., Claps, F., Teoh, J. Y. C., Serni, S., Scarpa, R. M., Porpiglia, F., Carrion, D. M., Rivas, J. G., Loeb, S., Cacciamani, G. E., Esperto, F., Lemos Almeida, J., Fiori, C., Hampson, L. A., Mantica, G., Minervini, A., Olivero, A., … Zhuang, J. (1997). Enhanced Reader.pdf. In *Nature* (Vol. 388, pp. 539–547).

Chemero, A., & Turvey, M. T. (2008). *Autonomy and hypersets*. *91*, 320–330. https://doi.org/10.1016/j.biosystems.2007.05.010

Chen, J., & Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika*, *95*(3), 759–771. https://doi.org/10.1093/biomet/asn034

Costantini, G., Epskamp, S., Borsboom, D., Perugini, M., Mõttus, R., Waldorp, L. J., & Cramer, A. O. J. (2015). State of the aRt personality research: A tutorial on network analysis of personality data in R. *Journal of Research in Personality*, *54*, 13–29. https://doi.org/10.1016/j.jrp.2014.07.003

Criado, R., Flores, J., García Del Amo, A., Gómez-Gardeñes, J., & Romance, M. (2012). A mathematical model for networks with structures in the mesoscale. *International Journal of Computer Mathematics*, *89*(3), 291–309. https://doi.org/10.1080/00207160.2011.577212

Crucitti, P., Latora, V., Marchiori, M., & Rapisarda, A. (2004). Error and attack tolerance of complex networks. *Physica A: Statistical Mechanics and Its Applications*, *340*(1–3), 388–394. https://doi.org/10.1016/j.physa.2004.04.031

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, *Complex Sy*, 1695. https://igraph.org

De Boer, N. S., De Bruin, L. C., Geurts, J. J. G., & Glas, G. (2021). The Network Theory of Psychiatric Disorders: A Critical Assessment of the Inclusion of Environmental Factors. *Frontiers in Psychology*, *12*(February). https://doi.org/10.3389/fpsyg.2021.623970

De Tayrac, M., Lê, S., Aubry, M., Mosser, J., & Husson, F. (2009). Simultaneous analysis of distinct Omics data sets with integration of biological knowledge: Multiple Factor Analysis approach. *BMC Genomics*, *10*(ii), 32. https://doi.org/10.1186/1471-2164-10-32

Donges, J. F., Schultz, H. C. H., Marwan, N., Zou, Y., & Kurths, J. (2011). Investigating the topology of interacting networks: Theory and application to coupled climate subnetworks. *European Physical Journal B*, *84*(4), 635–651. https://doi.org/10.1140/epjb/e2011-10795-8

Dziak, J. J., Coffman, D. L., Lanza, S. T., Li, R., & Jermiin, L. S. (2020). Sensitivity and specificity of information criteria. *Briefings in Bioinformatics*, *21*(2), 553–565. https://doi.org/10.1093/bib/bbz016

Epel, E. S., Crosswell, A. D., Mayer, S. E., Prather, A. A., Slavich, G. M., Puterman, E., & Berry, W. (2018). *Frontiers in Neuroendocrinology More than a feeling : A uni fi ed view of stress measurement for population science*. *49*(December 2017), 146–169. https://doi.org/10.1016/j.yfrne.2018.03.001

Epskamp, S., Borsboom, D., & Fried, E. I. (2017). Estimating psychological networks and their accuracy: a tutorial paper. *Behavior Research Methods*. https://arxiv.org/abs/1604.08462

Epskamp, Sacha, Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and

their accuracy: A tutorial paper. *Behavior Research Methods*, *50*(1), 195–212. https://doi.org/10.3758/s13428-017-0862-1

Epskamp, Sacha, & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, *23*(4), 617–634. https://doi.org/10.1037/met0000167

Epskamp, Sacha, Kramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network Visualizations of Relationships in Psychometric Data. *Journal of Statistical Software*, *48*(4), 1–18. http://www.jstatsoft.org/v48/i04/

Eroglu, D., Marwan, N., Stebich, M., & Kurths, J. (2018). Multiplex recurrence networks. *Physical Review E*, *97*(1), 1–9. https://doi.org/10.1103/PhysRevE.97.012312

Eronen, M. I. (2019). The levels problem in psychopathology. *Psychological Medicine*, *51*(6), 927–933. https://doi.org/10.1017/S0033291719002514

Fitzgerald, J., & Shanahan, T. (2000). Reading and writing relations and their development. *Educational Psychologist*, *35*(1), 39–50. https://doi.org/10.1207/S15326985EP3501_5

Foote, R. (2007). Mathematics and complex systems. *Science*, *318*(5849), 410–412. https://doi.org/10.1126/science.1141754

Foygel, R., & Drton, M. (2010). Extended Bayesian information criteria for Gaussian graphical models. *Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010, NIPS 2010*, 1–14.

Freeman, L. C. (1979). Centrality in social networks. *Social Networks*, *1*(3), 215–239. https://doi.org/10.1016/0378-8733(78)90021-7

Fried, E. I. (2020). Lack of Theory Building and Testing Impedes Progress in The Factor and Network Literature. In *Psychological Inquiry* (Vol. 31, Issue 4, pp. 271–288). https://doi.org/10.1080/1047840X.2020.1853461

Fried, E. I., & Robinaugh, D. J. (2020). Systems all the way down: Embracing complexity in mental health research. *BMC Medicine*, *18*(1), 4–7. https://doi.org/10.1186/s12916-020-01668-w

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*(3), 432–441. https://doi.org/10.1093/biostatistics/kxm045

Friedman, J., Hastie, T., & Tibshirani, R. (2019). *Graphical Lasso: Estimation of Gaussian Graphical Models* (1.11). https://www-stat.stanford.edu/~tibs/glasso

Gajjar, S., Kulahci, M., & Palazoglu, A. (2017). Selection of non-zero loadings in sparse principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, *162*(October 2016), 160–171. https://doi.org/10.1016/j.chemolab.2017.01.018

Gluckman, M. (1995). *The Judicial Process Among the Barotse of Northern Rhodesia*. Manchester University Press.

Gomez, M., Garcia, S., Rajtmajer, S., Grady, C., & Mejia, A. (2020). Fragility of a multilayer network of intranational supply chains. *Applied Network Science*, *5*(1), 1–21. https://doi.org/10.1007/s41109-020-00310-1

Gu, Z., & Van Deun, K. (2018). *RegularizedSCA: Regularized Simultaneous Component Based Data Integration* (R package version 0.5.4). https://cran.r-project.org/package=RegularizedSCA%0AA

Gu, Z., & Van Deun, K. (2019). RegularizedSCA: Regularized simultaneous component analysis of multiblock data in R. *Behavior Research Methods*, *51*(5), 2268–2289. https://doi.org/10.3758/s13428-018-1163-z

Hammoud, Z., & Kramer, F. (2020). Multilayer networks: aspects, implementations, and

application in biomedicine. *Big Data Analytics*, *5*(1), 1–18. https://doi.org/10.1186/s41044-020-00046-0

Hasselman, F. (2022). Early Warning Signals in Phase Space : Geometric Resilience Loss Indicators from Multiplex Cumulative Recurrence Networks. *Frontiers in Physiology*, *13*, 1–30. https://doi.org/10.3389/fphys.2022.859127

Hasselman, F., & Bosman, A. M. T. (2020). Studying complex adaptive systems with internal states: A recurrence network approach to the analysis of multivariate time-series data representing self-reports of human experience. *Frontiers in Applied Mathematics and Statistics*, *6*(April), 1–14. https://doi.org/10.3389/fams.2020.00009

Heino, M. T. J., Knittle, K. P., Noone, C., Hasselman, F., & Hankonen, N. (2020). *Studying behaviour change mechanisms under complexity*. 1–22. https://doi.org/10.31234/osf.io/fxgw4

Hinch, R., Greenstein, J. L., & Winslow, R. L. (2006). Multi-scale models of local control of calcium induced calcium release. *Progress in Biophysics and Molecular Biology*, *90*(1–3), 136–150. https://doi.org/10.1016/j.pbiomolbio.2005.05.014

Ihlen, E. A. F., & Vereijken, B. (2010). Interaction-dominant dynamics in human cognition: Beyond 1/f αfluctuation. *Journal of Experimental Psychology: General*, *139*(3), 436–463. https://doi.org/10.1037/a0019098

Jensen, A. C., Pickett, J. M., Jorgensen-wells, M. A., Andrus, L. E., Leiter, V. K., Graver, H., Meta-analysis, A. M., Jensen, A. C., Pickett, J. M., Jorgensen-wells, M. A., Andrus, L. E., Leiter, V. K., & Graver, H. (2022). The Family Stress Model and Parents ' Differential Treatment of Siblings : A Multilevel Meta-analysis. *Journal of Family Psychology*. https://doi.org/http://dx.doi.org/10.1037/fam0000978

Jeong, H., Mason, S. P., Barabási, A. L., & Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature*, *411*(6833), 1–2. https://doi.org/http://dx.doi.org/10.1038/35075138

Jo, H. H., Baek, S. K., & Moon, H. T. (2006). Immunization dynamics on a two-layer network model. *Physica A: Statistical Mechanics and Its Applications*, *361*(2), 534–542. https://doi.org/10.1016/j.physa.2005.06.074

Kaiser, M., Görner, M., & Hilgetag, C. C. (2007). Criticality of spreading dynamics in hierarchical cluster networks without inhibition. *New Journal of Physics*, *9*. https://doi.org/10.1088/1367-2630/9/5/110

Kalna, G., & Higham, D. J. (2007). A clustering coefficient for weighted networks, with application to gene expression data. *AI Communications*, *20*(4), 263–271.

Kassambara, A., & Mundt, F. (2020). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses* (R package version 1.0.7). https://cran.r-project.org/package=factoextra

Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., & Porter, M. A. (2014). Multilayer networks. *Journal of Complex Networks*, *2*(3), 203–271. https://doi.org/10.1093/comnet/cnu016

Koller, D., & Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press.

Kurant, M., & Thiran, P. (2006). Layered complex networks. *Physical Review Letters*, *96*(13), 1–4. https://doi.org/10.1103/PhysRevLett.96.138701

Kurant, M., Thiran, P., & Hagmann, P. (2007). Error and attack tolerance of layered complex networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *76*(2), 1–

5. https://doi.org/10.1103/PhysRevE.76.026103

Ladyman, J., Lambert, J., & Wiesner, K. (2013). What is a complex system? In *European Journal for Philosophy of Science* (Vol. 3, Issue 1). https://doi.org/10.1007/s13194-012-0056-8

Latora, V., Nicosia, V., & Panzarasa, P. (2013). Social Cohesion, Structural Holes, and a Tale of Two Measures. *Journal of Statistical Physics*, *151*(3–4), 745–764. https://doi.org/10.1007/s10955-013-0722-z

Lauritzen, S. L. (1996). *Graphical models*. Clarendon Press.

Lazega, E., Jourda, M. T., Mounier, L., & Stofer, R. (2008). Catching up with big fish in the big pond? Multi-level network analysis through linked design. *Social Networks*, *30*(2), 159–176. https://doi.org/10.1016/j.socnet.2008.02.001

Lock, E. F., Hoadley, K. A., Marron, J. S., & Nobel, A. B. (2013). Joint and individual variation explained (JIVE) for integrated analysis of multiple data types. In *Annals of Applied Statistics* (Vol. 7, Issue 1, pp. 523–542). https://doi.org/10.1214/12-AOAS597

Maatman, F. O. (2020). *Reformulating the network theory of mental disorders : Folk psychology as a factor , not a fact*. https://doi.org/10.1177/0959354320921464

Magnani, M., & Rossi, L. (2013). Formation of multiple networks. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7812 LNCS*, 257–264. https://doi.org/10.1007/978-3-642-37210-0_28

Masten, A. S., Lucke, C. M., Nelson, K. M., & Stallworthy, I. C. (2021). Resilience in Development and Psychopathology: Multisystem Perspectives. *Annual Review of Clinical Psychology*, *17*, 521–549. https://doi.org/10.1146/annurev-clinpsy-081219-

120307

Miller, G. A. (2010). Mistreating psychology in the decades of the brain. *Perspectives on Psychological Science*, *5*(6), 716–743. https://doi.org/10.1177/1745691610388774

Min, B., & Goh, K. I. (2013). Layer-crossing overheid and information spreading in multiplex social networks. *American Physical Society*, *28*. https://doi.org/10.1038/srep21392

Nelson, B., McGorry, P. D., Wichers, M., Wigman, J. T. W., & Hartmann, J. A. (2017). Moving from static to dynamic models of the onset of mental disorder a review. *JAMA Psychiatry*, *74*(5), 528–534. https://doi.org/10.1001/jamapsychiatry.2017.0001

Newman, M. E. J. (2003). Mixing patterns in networks. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *67*(2), 13. https://doi.org/10.1103/PhysRevE.67.026126

Newman, M. E. J. (2004). Analysis of weighted networks. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *70*(5), 9. https://doi.org/10.1103/PhysRevE.70.056131

Noh, J. D. (2003). Exact scaling properties of a hierarchical network model. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *67*(4), 4. https://doi.org/10.1103/PhysRevE.67.045103

Norman, G. J., DeVries, A. C., Cacioppo, J. T., & Berntson, G. G. (2011). Multilevel Analysis of Stress. In R. J. Contrada & A. Baum (Eds.), *The Handbook of Stress Science: Biology, Psychology and Health* (pp. 619–634). Springer Publisher Company, LLC.

Olthof, M., Hasselman, F., Maatman, F. O., & Bosman, A. M. T. (2021). Complexity theory of psychopathology. *PsyArxiv*, 1–35.

Onnela, J. P., Saramäki, J., Kertész, J., & Kaski, K. (2005). Intensity and coherence of motifs in weighted complex networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *71*(6), 1–4. https://doi.org/10.1103/PhysRevE.71.065103

R Core Team. (2020). *R: A Language and Environment for Statistical Computing* (4.0.3). R Foundation for Statistical Computing. https://www.r-project.org

Ravasz, E., & Barabási, A. L. (2003). Hierarchical organization in complex networks. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *67*(2), 7. https://doi.org/10.1103/PhysRevE.67.026112

Riese, H., & Wichers, M. (2021). Comment on: Eronen MI (2019). The levels problem in psychopathology. *Psychological Medicine*, *51*(3), 525–526. https://doi.org/10.1017/S0033291719003623

Robinaugh, D. J., Hoekstra, R. H. A., Toner, E. R., & Borsboom, D. (2020). The network approach to psychopathology: A review of the literature 2008-2018 and an agenda for future research. *Psychological Medicine*, *50*(3), 353–366. https://doi.org/10.1017/S0033291719003404

Rocklin, M., & Pinar, A. (2013). On clustering on graphs with multiple edge types. *Internet Mathematics*, *9*(1), 82–112. https://doi.org/10.1080/15427951.2012.678191

Rosato, V., Issacharoff, L., Tiriticco, F., Meloni, S., De Porcellinis, S., & Setola, R. (2008). Modelling interdependent infrastructures using interacting dynamical models. *International Journal of Critical Infrastructures*, *4*(1–2), 63–79. https://doi.org/10.1504/IJCIS.2008.016092

Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, *31*(4), 581–603. https://doi.org/10.1007/BF02289527

Saramäki, J., Kivelä, M., Onnela, J. P., Kaski, K., & Kertész, J. (2007). Generalizations of the clustering coefficient to weighted complex networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *75*(2), 2–5. https://doi.org/10.1103/PhysRevE.75.027105

Schmank, C. J., Goring, S. A., Kovacs, K., & Conway, A. R. A. (2021). Investigating the structure of intelligence using latent variable and psychometric network modeling: A commentary and reanalysis. *Journal of Intelligence*, *9*(1), 1–13. https://doi.org/10.3390/jintelligence9010008

Schouteden, M., Van Deun, K., Pattyn, S., & Van Mechelen, I. (2013). SCA with rotation to distinguish common and distinctive information in linked data. *Behavior Research Methods*, *45*(3), 822–833. https://doi.org/10.3758/s13428-012-0295-9

Schuster, S., Fell, D. A., & Dandekar, T. (2000). *Nature Biotechnology 2000 Schuster*. *18*(March), 1–7. papers2://publication/uuid/00A659FC-63C2-438F-880F-F7D7070A1E4B

Silverman, E. K., & Loscalzo, J. (2012). Network medicine approaches to the genetics of complex diseases. *Discovery Medicine*, *14*(75), 143–152.

Simpson-Kent, I. L., Fried, E. I., Akarca, D., Mareva, S., Bullmore, E. T., & Kievit, R. A. (2021). Bridging brain and cognition: A multilayer network analysis of brain structural covariance and general intelligence in a developmental sample of struggling learners. *Journal of Intelligence*, *9*(2). https://doi.org/10.3390/jintelligence9020032

Smilde, A. K., Måge, I., Næs, T., Hankemeier, T., Lips, M. A., Kiers, H. A. L., Acar, E., & Bro, R. (2017). Common and distinct components in data fusion. *Journal of Chemometrics*, *31*(7). https://doi.org/10.1002/cem.2900

Szary, J., Dale, R., Kello, C. T., & Rhodes, T. (2015). Patterns of interaction-dominant dynamics in individual versus collaborative memory foraging. *Cognitive Processing*, *16*(4), 389–399. https://doi.org/10.1007/s10339-015-0731-8

Szell, M., Lambiotte, R., & Thurner, S. (2010). Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(31), 13636–13641. https://doi.org/10.1073/pnas.1004008107

Tanskanen, A. J., Greenstein, J. L., Chen, A., Sun, S. X., & Winslow, R. L. (2007). Protein geometry and placement in the cardiac dyad influence macroscopic properties of calcium-induced calcium release. *Biophysical Journal*, *92*(10), 3379–3396. https://doi.org/10.1529/biophysj.106.089425

Tenenhaus, A., & Tenenhaus, M. (2014). Regularized generalized canonical correlation analysis for multiblock or multigroup data analysis. *European Journal of Operational Research*, *238*(2), 391–403. https://doi.org/10.1016/j.ejor.2014.01.008

Thomas, J. G., & Sharp, P. B. (2019). Mechanistic Science: A New Approach to Comprehensive Psychopathology Research That Relates Psychological and Biological Phenomena. *Clinical Psychological Science*, *7*(2), 196–215. https://doi.org/10.1177/2167702618810223

Tio, P., Waldorp, L., & VanDeun, K. (2020). Constructing graphical models for multi-source data: Sparse network and component analysis. *Studies in Classification, Data Analysis, and Knowledge Organization*, 275–287. https://doi.org/10.1007/978-981-15-3311-2_22

Trendafilov, N. T. (2014). From simple structure to sparse components: A review. *Computational Statistics*, *29*(3–4), 431–454. https://doi.org/10.1007/s00180-013-0434-5

Turvey, M. T. (2007). Action and perception at the level of synergies. *Human Movement Science*, *26*(4), 657–697. https://doi.org/10.1016/j.humov.2007.04.002

Vaiana, M., & Muldoon, S. F. (2020). Multilayer Brain Networks. *Journal of Nonlinear Science*, *30*(5), 2147–2169. https://doi.org/10.1007/s00332-017-9436-8

Valente, T. W. (2012). Network interventions. *Science*, *336*(6090), 49–53. https://doi.org/10.1126/science.1217330

Van Bork, R., Rhemtulla, M., Waldorp, L. J., Kruis, J., Rezvanifar, S., & Borsboom, D. (2021). Latent Variable Models and Networks: Statistical Equivalence and Testability. *Multivariate Behavioral Research*, *56*(2), 175–198. https://doi.org/10.1080/00273171.2019.1672515

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

Van Deun, K., Smilde, A. K., Thorrez, L., Kiers, H. A. L., & Van Mechelen, I. (2013). Identifying common and distinctive processes underlying multiset data. *Chemometrics and Intelligent Laboratory Systems*, *129*, 40–51. https://doi.org/10.1016/j.chemolab.2013.07.005

Van Deun, Katrijn, Smilde, A. K., van der Werf, M. J., Kiers, H. A. L., & Van Mechelen, I. (2009). A structured overview of simultaneous component based data integration. *BMC Bioinformatics*, *10*, 246. https://doi.org/10.1186/1471-2105-10-246

Van Deun, Katrijn, Wilderjans, T. F., Van den Berg, R. A., Antoniadis, A., & Van Mechelen, I. (2011). A flexible framework for sparse simultaneous component based data integration. *BMC Bioinformatics*, *12*(1). https://doi.org/10.1186/1471-2105-12-448

Van Geert, P. L. C. (2019). Dynamic Systems, Process and Development. *Human Development*, *63*(3–4), 153–179. https://doi.org/10.1159/000503825

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General*, *132*(3), 331–350. https://doi.org/10.1037/0096-3445.132.3.331

Van Orden, G., Hollis, G., & Wallot, S. (2012). The blue-collar brain. *Frontiers in Physiology*, *3 JUN*(June), 1–12. https://doi.org/10.3389/fphys.2012.00207

Vazquez, A. (2006). Spreading dynamics on heterogeneous populations: Multitype network approach. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *74*(6), 1–11. https://doi.org/10.1103/PhysRevE.74.066114

Verbrugge, L. M. (1979). Multiplexity in Adult Friendships. *Social Forces*, *57*(4), 1286–1309. https://doi.org/10.1093/sf/57.4.1286

Wallot, S., & Kelty-Stephen, D. G. (2018). Interaction-Dominant Causation in Mind and Brain, and Its Implication for Questions of Generalization and Replication. *Minds and Machines*, *28*(2), 353–374. https://doi.org/10.1007/s11023-017-9455-0

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press.

Watts, D. J., & Strogatz, S. H. (1998). Strogatz - small world network Nature. *Nature*, *393*(June), 440–442. https://www.ncbi.nlm.nih.gov/pubmed/9623998

Wijnants, M. L. (2014). A Review of Theoretical Perspectives in Cognitive Science on the Presence of 1 / f Scaling in Coordinated Physiological and Cognitive Processes . *Journal of Nonlinear Dynamics*, *2014*, 1–17. https://doi.org/10.1155/2014/962043

Wilderjans, T. F., Ceulemans, E., Van Mechelen, I., & Van Den Berg, R. A. (2011).

Simultaneous analysis of coupled data matrices subject to different amounts of noise. *British Journal of Mathematical and Statistical Psychology*, *64*(2), 277–290. https://doi.org/10.1348/000711010X513263

Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology*, *4*(1). https://doi.org/10.2202/1544-6115.1128

Zhou, C., Zemanová, L., Zamora-López, G., Hilgetag, C. C., & Kurths, J. (2007). Structure-function relationship in complex brain networks expressed by hierarchical synchronization. *New Journal of Physics*, *9*. https://doi.org/10.1088/1367-2630/9/6/178

Zhou, C., Zemanová, L., Zamora, G., Hilgetag, C. C., & Kurths, J. (2006). Hierarchical organization unveiled by functional connectivity in complex brain networks. *Physical Review Letters*, *97*(23), 3–6. https://doi.org/10.1103/PhysRevLett.97.238103

Zignani, M., Quadri, C., Gaitto, S., & Rossi, G. P. (2014). *Exploiting all phone media? A multidimensional network analysis of phone users' sociality*. http://arxiv.org/abs/1401.3126

Zou, H., Hastie, T., & Tibshirani, R. (2006). Sparse Principal Component Analysis. *Journal of Computational and Graphical Statistics*, *15*(2), 265–286. https://www.jstor.org/stable/27594179

Appendix A

Data Concatenations and Variable Types

Source	**Variable, Type, Possible Values**	**Concatenated**	**Question**
Physical health	Amount of conditions Count 0-18	Yes; 1 binary item (yes/no)	Please look at card [SHOWCARD]. Has a doctor ever told you that you had/Do you currently have any of the conditions on this card? With this we mean that a doctor has told you that you have this condition, and that you are either currently being treated for or bothered by this condition.
		Yes; 17 binary items (yes/no)	Please tell me the number or numbers of the conditions. 1. A heart attack including myocardial infarction or coronary thrombosis or any other heart problem including congestive heart failure 2. High blood pressure or hypertension 3. High blood cholesterol 4. A stroke or cerebral vascular disease 5. Diabetes or high blood sugar 6. Chronic lung disease such as chronic bronchitis or emphysema

			10. Cancer or malignant tumour, including leukaemia or lymphoma, but excluding minor skin cancers 11. Stomach or duodenal ulcer, peptic ulcer 12. Parkinson disease 13. Cataracts 14. Hip fracture 15. Other fractures 16. Alzheimer's disease, dementia, organic brain syndrome, senility or any other serious memory impairment 18. Other affective or emotional disorders, including anxiety, nervous or psychiatric problems 19. Rheumatoid Arthritis 20. Osteoarthritis, or other rheumatism 21. Chronic kidney disease
		Yes; 1 item (binary; yes/no)	What other conditions have you had?
	Pain level	Yes; 1 item (binary; yes/no)	Are you troubled with pain?
	Ordinal 0-3 0: none 1: mild	Yes; 1 item (ordinal)	How bad is the pain most of the time? 1. mild 2. moderate 3. severe

	2: moderate 3: severe		
	Difficulties activities 1 Count 0-15	Yes; 15 items (binary; yes/no)	Please look at card [SHOWCARD]. Please tell me if you have any difficulty with these activities because of a physical, mental, emotional or memory problem. Again exclude any difficulties you expect to last less than three months. 1. Dressing, including putting on shoes and socks 2. Walking across a room 3. Bathing or showering 4. Eating, such as cutting up your food 5. Getting in or out of bed 6. Using the toilet, including getting up or down 7. Using a map to figure out how to get around in a strange place 8. Preparing a hot meal 9. Shopping for groceries 10. Making telephone calls 11. Taking medications 12. Doing work around the house or garden 13. Managing money, such as paying bills and keeping track of expenses

			14. Leaving the house independently and accessing transportation services 15. Doing personal laundry 0. None of these
	Difficulties Activities 2 Count 0-11	Yes; 11 items (binary; yes/no)	Please look at card [SHOWCARD]. Please tell me whether you have any difficulty doing each of the everyday activities on this card. Exclude any difficulties that you expect to last less than three months. 1. Walking 100 metres 2. Sitting for about two hours 3. Getting up from a chair after sitting for long periods 4. Climbing several flights of stairs without resting 5. Climbing one flight of stairs without resting 6. Stooping, kneeling, or crouching 7. Reaching or extending your arms above shoulder level 8. Pulling or pushing large objects like a living room chair 9. Lifting or carrying weights over 10 pounds/5 kilos, like a heavy bag of groceries 10. Picking up a small coin from a table 0. None of these

	Frailty Count 0-4	Yes; 5 items (binary; yes/no)	For the past six months at least, have you been bothered by any of the health conditions on this card? Please tell me the number or numbers. 1. Falling down 2. Fear of falling down 3. Dizziness, faints or blackouts 4. Fatigue 0. None
	Help meet needs Ordinal 0-4	Yes; 2 items (binary + ordinal)	Thinking about the activities that you have problems with, does anyone ever help you with these activities? (binary)
			Would you say that the help you receive meets your needs? 1. All the time 2. Usually 3. Sometimes 4. Hardly ever
	Health in General Ordinal 1-5	No	Would you say your health is... 1. Poor 2. Fair 3. Good 4. Very good 5. Excellent
	Chronic health problems	No	Some people suffer from chronic or long-term health problems. By chronic or long-term we

	Binary 0-1		mean it has troubled you over a period of time or is likely to affect you over a period of time. Do you have any such health problems, illness, disability or infirmity? Yes/no
	Limited activities Ordinal 0-2	No	For the past six months at least, to what extent have you been limited because of a health problem in activities people usually do? 0. Not limited 1. Limited, but not severely 2. Severely limited
	Limited Work Binary 0-1	No	Do you have any health problem or disability that limits the kind or amount of paid work you can do? Yes/no
	Hearing Ordinal 1-5	No	Would you say your hearing (using a hearing aid as usual) is.. 1. Poor 2. Fair 3. Good 4. Very good 5. Excellent
Cognitive Function	Reading Ordinal	No	Now I would like to ask some questions about your reading and writing skills. How would

	1-5		you rate your reading skills needed in your daily life? Would you say they are... 1. Poor 2. Fair 3. Good 4. Very good 5. Excellent
	Writing Ordinal 1-5	No	How would you rate your writing skills needed in your daily life? Would you say they are... 1. Poor 2. Fair 3. Good 4. Very good 5. Excellent
	Orientation Count 0-4	Yes; 4 items (binary; yes/no)	Part of this study is concerned with people's memory AND ability to think about things. First, I am going to ask about today's date. Which day of the month is it?
			Which month is it?
			Which year is it?
			Can you tell me what day of the week it is?
	Words memory A Count	Yes; 4 items (count; 0-10)	Now, I am going to read a list of words from my computer screen. We have purposely made the list long so it will be difficult for anyone to recall all the words. Most people recall just a

	0-41		few. Please listen carefully, as the set of words cannot be repeated. When I have finished, I will ask you to recall aloud as many of the words as you can, in any order. Is this clear? Now please tell me all the words you can recall. 10 words, first trial
			10 words, second trial
			10 words, third trial
			10 words, fourth trial
	Words Memory B Count 0-41	Yes; 4 items (count; 0-10)	A little while ago, I read you a list of words and you repeated the ones you could remember. Please tell me any of the words that you can remember now? 10 words, trial 1 delay
			10 words, trial 2 delay
			10 words, trial 3 delay
			10 words, trial 4 delay
	Numeracy Subtraction Count 0-5	Yes; 5 items (binary; yes/no)	Now let's try some subtraction of numbers. One hundred minus 7 equals what?
			And 7 from that
			And 7 from that
			And 7 from that
			And 7 from that

Mental Health	Mood Count 0-6	Yes; 5 items (binary + ordinal)	n the last month, have you been sad or depressed? Binary (yes/no)
			In the last month, have you felt that you would rather be dead? Binary (yes/no)
			Do you tend to blame yourself or feel guilty about anything? 0. No such feelings 1. Mentions guilt or self-blame, but it is unclear if these constitute obvious or excessive guilt or self-blame 2. Obvious excessive guilt or self-blame
			What have you enjoyed doing recently? Binary (yes/no)
			In the last month, have you cried at all? Binary (yes/no)
	Perspective Count 0-2	Yes; 2 items (binary; yes/no)	What are your hopes for the future? 0 Any hopes mentioned 1 No hopes mentioned
			0. Less interest than usual mentioned 0. Non-specific or uncodeable response 1. No mention of loss of interest
	Fitness		Have you had trouble sleeping recently?

	Count 0-3	Yes; 3 items (binary; yes/no)	In the last month, have you had too little energy to do the things you wanted to do?
			In the last month, have you had too little energy to do the things you wanted to do?
	Mental Fitness Count 0-3	Yes; 3 items (binary; yes/no)	How is your concentration? For example, can you concentrate on a television programme, film or radio programme? 1. Difficulty in concentrating on entertainment 2. No such difficulty mentioned
			Can you concentrate on something you read? 1. Difficulty in concentrating on reading 2. No such difficulty mentioned
			Have you been irritable recently? 1. Yes 5. No
	Social Count 0-8	Yes; 4 items (ordinal; 0-2)	How much of the time do you feel you lack companionship? 0. Hardly ever or never 1. Some of the time 2. Often
			How much of the time do you feel left out? 0. Hardly ever or never 1. Some of the time 2. Often

			How much of the time do you feel isolated from others? 0. Hardly ever or never 1. Some of the time 2. Often
			How much of the time do you feel lonely? 0. Hardly ever or never 1. Some of the time 2. Often

Appendix B

Script SNACS

```
#' SNACs

#' Sparse Network And Component (supplemental) analysis

#' Adapted from original SNAC function 06-05-2022

#'

#'

#' @param data Data matrix, ordered according to groups, variables group normalized

#' @param Jk Numeric vector of length equal to number of sources in the data, each value
indicating the number of vars belonging to each source (block)

#' @param preProcess Should the data be pre processed according to
[RegularizedSCA::pre_process]

#' @param sourceStandardize Standardize variables by group (source/block)

#' @param R Number of components to consider...

#' @param graphLambda Lambda for call to [glasso::glasso]

#' @param nfolds Perform n-fold cross validation [RegularizedSCA::cv_sparseSCA]

#' @param NStart Multi starts see [RegularizedSCA::sparseSCA]

#' @param method How to calculate lasso penalty see [RegularizedSCA::sparseSCA]

#' @param MaxIter Max iterations see [RegularizedSCA::sparseSCA]
```

```
#' @param estLambdas If `TRUE` and `crossval` is also `TRUE`, recommended Lambda's will be estimated from cross-validation [RegularizedSCA::cv_sparseSCA], this takes a lot of time. If `TRUE` and `crossval` is  `FALSE`, function [RegularizedSCA::maxLGlasso] will be used, which is fast, but likely less optimal than cross validation. If `estLambdas` if `FALSE`, provide the values in `SCALambda` and `SCAgroupLambda`

#' @param crossval If `TRUE` and `estLambdas` is `TRUE` recommended Lambda's will be estimated from cross-validation [RegularizedSCA::cv_sparseSCA], this takes a lot of time.

#' @param crossvalMethod How to calculate lasso penalty in cross validation see [RegularizedSCA::cv_sparseSCA]

#' @param SCALambda Recommended Lambda for [RegularizedSCA::sparseSCA]

#' @param SCAgroupLambda Recommended Group Lambda for [RegularizedSCA::sparseSCA]

#'

#' @return

#'

#' @export

#'

#' @examples

#'

SNACS <- function(data,

                  Jk,
```

```
            preProcess = FALSE,
            sourceStandardize = FALSE,
            R,
            graphLambda = 0.002,
            NStart = 10,
            method = c("component","datablock")[2],
            nfolds = 10,
            MaxIter = 100,
            estLambdas = TRUE,
            crossval = FALSE,
            crossvalMethod = c("component","datablock")[1],
            SCALambda = NA,
            SCAgroupLambda = NA){


require(RegularizedSCA)
require(glasso)


if(is.numeric(Jk)&(length(Jk)>=2)){
  if(!sum(Jk)==NCOL(data)){
    stop("Number of variables on data not equal to blocks in Jk")
```

```
  }
}

JkID <- c(1, cumsum(Jk))

if(preProcess){
  data <- RegularizedSCA::pre_process(data)
}

if(sourceStandardize){
  for(j in 2:length(JkID)){
    M  <- mean(unlist(data[,JkID[j-1]:JkID[j]]), na.rm = TRUE)
    SD <- sd(unlist(data[,JkID[j-1]:JkID[j]]), na.rm = TRUE)
    data[,JkID[j-1]:JkID[j]] <- (data[,JkID[j-1]:JkID[j]]-M)/SD
  }
}

#Find lambda values for lasso and grouplasso using cross-validation
if(estLambdas){
  if(crossval){
```

```
    para <- RegularizedSCA::cv_sparseSCA(DATA = data, Jk = Jk, R = R, nfolds = nfolds,
method = method, MaxIter = MaxIter)
    SCALambda      <- para$RecommendedLambda[1]
    SCAgroupLambda <- para$RecommendedLambda[2]
  } else {
    para <- RegularizedSCA::maxLGlasso(DATA = data, Jk = Jk, R = R)
    SCALambda      <- para$Lasso
    SCAgroupLambda <- para$Glasso
  }
} else {
  if(any(is.na(SCALambda),is.na(SCAgroupLambda))){
    stop("Need values for SCALambda and SCAgroupLambda")
  } else {
    message("Skipping cross-validation...")
  }
}


SCALambda      <- round(SCALambda, digits = 6)
SCAgroupLambda <- round(SCAgroupLambda, digits = 6)
```

```
#Run SCA

results <- RegularizedSCA::sparseSCA(DATA = data,

                                     Jk = Jk,

                                     R = R,

                                     LASSO = SCALambda,

                                     GROUPLASSO = SCAgroupLambda,

                                     NRSTART = NStart,

                                     method  = method,

                                     MaxIter = MaxIter)



#Undo shrinkage done on the P and T matrices

final   <- RegularizedSCA::undoShrinkage(DATA = data, R = R, Phat = results$Pmatrix)

Pmatrix <- final$Pmatrix

Tmatrix <- final$Tmatrix



#combis <- data.frame(t(utils::combn(seq_along(Jk), 2L)), stringsAsFactors = FALSE)



#Find common components

common   <- matrix(NA, nrow = ncol(final$Pmatrix), ncol = length(Jk))

comdist  <- character(length = R)
```

```
  sources  <- paste0("s",seq_along(Jk))


  for (c in 1:ncol(final$Pmatrix)){# look per component whether it is a common or distinctive
one
    d <- NULL


    for (j in 2:length(JkID)){
      d[j-1] <- sum(abs(final$Pmatrix[JkID[j-1]:JkID[j],c]))
      if(d[j-1]!=0){d[j-1] = 1}
    }


    common[c,] <- d


    if(sum(d)>=2){
      comdist[c] <- paste("common:", paste(sources[d==1], collapse = " & "))
    } else {
      if(sum(d)==1){
       comdist[c] <- paste("distinct:", sources[d==1])
      } else {
        comdist[c] <- "NA"
```

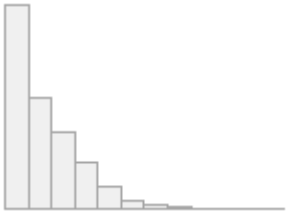
```
      }
    }
  }

  common <- data.frame(common, type = comdist, row.names = paste0("comp",1:R))
  colnames(common) <- c(paste0("source",1:length(Jk)),"type")

    print(common)
    cat("\nWhich components (in rows) do you want to keep?\n1. All\n2. Some\n3. None
(quit)")
    CHOICE <- readline()
    comp <- try(switch(CHOICE,
                "1" = paste(1:R, collapse=","),
                "2" = readline("Enter component numbers, seperated by a comma: "),
                "3" = NA))

  if(any(is.na(comp))){
    warning("No components selected!\nReturning unshrunken results.")
    return(final)
  } else {
```

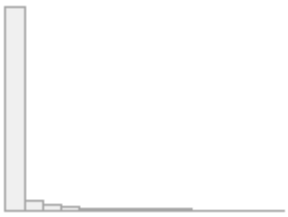
```
    comp <- strsplit(comp, ",")

    comp <- as.integer(comp[[1]])

    comp <- na.exclude(comp)

  }


  TmatCommon = matrix(NA, nrow = nrow(data), ncol = length(comp)) #, dimnames =
list(colnames(data),colnames(data)))

  PmatCommon = matrix(NA, nrow = ncol(data), ncol = length(comp)) #, dimnames =
list(colnames(data),colnames(data)))


  for (i in 1:length(comp)){

    TmatCommon[,i] = final$Tmatrix[,comp[i]]

    PmatCommon[,i] = final$Pmatrix[,comp[i]]

  }


  #Create data set based on selected components

  Xcom <- TmatCommon%*%t(PmatCommon)


  #Graphical lasso
```

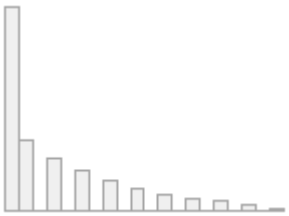
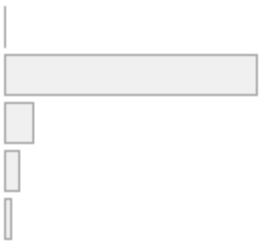
```
  SigmaINVC_hat <- glasso::glasso(s = cov(Xcom), rho = graphLambda, penalize.diagonal =
FALSE)$w

  attr(SigmaINVC_hat,"scaLambdas") <- c(LASSO = SCALambda, GROUPLASSO =
SCAgroupLambda)

  if(estLambdas){

    attr(SigmaINVC_hat,"crossval") <- para

  }



  colnames(Xcom) <- colnames(data)

  #dimnames(SigmaINVC_hat) <- colnames(data)

  return(list(compVsource = common,

              comp = comp,

              componentData = Xcom,

              glasso = SigmaINVC_hat)

         )

}
```

Appendix C

Descriptive Statistics

No	Variable	Stats / Values	Freqs (% of Valid)	Graph
1	Conditions_amount [numeric]	Mean (sd) : 2 (1.6) min ≤ med ≤ max: 0 ≤ 2 ≤ 12 IQR (CV) : 2 (0.8)	13 distinct values	
2	Painlevel [numeric]	Mean (sd) : 0.9 (1.1) min ≤ med ≤ max:	0 : 7181 (52.9%) 1 : 1444 (10.6%) 2 : 3448 (25.4%) 3 : 1504 (11.1%)	

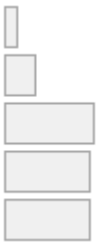
No	Variable	Stats / Values	Freqs (% of Valid)	Graph
		0 ≤ 0 ≤ 3 IQR (CV) : 2 (1.2)		
3	Difficulties_activitiesB [numeric]	Mean (sd) : 1.6 (1.9) min ≤ med ≤ max: 0 ≤ 1 ≤ 15 IQR (CV) : 0 (1.2)	16 distinct values	

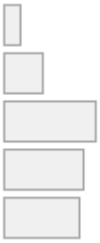
No	Variable	Stats / Values	Freqs (% of Valid)	Graph
4	Difficulties_activitiesA [numeric]	Mean (sd) : 1.9 (2.4) min ≤ med ≤ max: 0 ≤ 1 ≤ 10 IQR (CV) : 3 (1.3)	11 distinct values	
5	Frailty [numeric]	Mean (sd) : 1.3 (0.6) min ≤ med ≤ max:	0 : 2 (0.0%) 1 : 11366 (83.7%) 2 : 1305 (9.6%) 3 : 610 (4.5%) 4 : 294 (2.2%)	

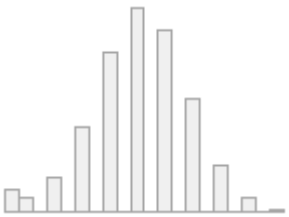
No	Variable	Stats / Values	Freqs (% of Valid)	Graph
		0 ≤ 1 ≤ 4 IQR (CV) : 0 (0.5)		
6	Help_meet_needs [numeric]	Mean (sd) : 0.9 (1.6) min ≤ med ≤ max: 0 ≤ 0 ≤ 4 IQR (CV) : 0 (1.8)	0 : 10200 (75.1%) 1 : 36 (0.3%) 2 : 208 (1.5%) 3 : 829 (6.1%) 4 : 2304 (17.0%)	

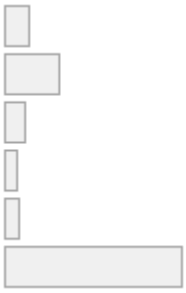
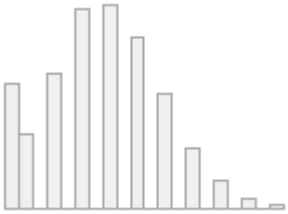
No	Variable	Stats / Values	Freqs (% of Valid)	Graph
7	Health_general [numeric]	Mean (sd) : 3.2 (1) min ≤ med ≤ max: 1 ≤ 3 ≤ 5 IQR (CV) : 1 (0.3)	1 : 756 (5.6%) 2 : 2404 (17.7%) 3 : 5309 (39.1%) 4 : 3832 (28.2%) 5 : 1276 (9.4%)	
8	Chronic_health_problems [numeric]	Min : 0 Mean : 0.5 Max : 1	0 : 6568 (48.4%) 1 : 7009 (51.6%)	
9	Limited_activities [numeric]	Mean (sd) : 0.6 (0.7)	0 : 7336 (54.0%) 1 : 4327 (31.9%) 2 : 1914 (14.1%)	

No	Variable	Stats / Values	Freqs (% of Valid)	Graph
		min ≤ med ≤ max: 0 ≤ 0 ≤ 2 IQR (CV) : 1 (1.2)		
10	Limit_work [numeric]	Min : 0 Mean : 0.3 Max : 1	0 : 10107 (74.4%) 1 : 3470 (25.6%)	
11	Hearing [numeric]	Mean (sd) : 3.3 (1) min ≤ med ≤ max:	1 : 469 (3.5%) 2 : 2247 (16.6%) 3 : 5553 (40.9%) 4 : 3621 (26.7%) 5 : 1687 (12.4%)	

No	Variable	Stats / Values	Freqs (% of Valid)	Graph
		1 ≤ 3 ≤ 5 IQR (CV) : 1 (0.3)		
12	Reading [numeric]	Mean (sd) : 3.7 (1.1) min ≤ med ≤ max: 1 ≤ 4 ≤ 5 IQR (CV) : 2 (0.3)	1 : 598 (4.4%) 2 : 1352 (10.0%) 3 : 3975 (29.3%) 4 : 3804 (28.0%) 5 : 3848 (28.3%)	

No	Variable	Stats / Values	Freqs (% of Valid)	Graph
13	Writing [numeric]	Mean (sd) : 3.5 (1.2) min ≤ med ≤ max: 1 ≤ 4 ≤ 5 IQR (CV) : 1 (0.3)	1 : 791 (5.8%) 2 : 1756 (12.9%) 3 : 4130 (30.4%) 4 : 3538 (26.1%) 5 : 3362 (24.8%)	
14	Oriëntation [numeric]	Mean (sd) : 3.8 (0.6) min ≤ med ≤ max:	0 : 159 (1.2%) 1 : 114 (0.8%) 2 : 287 (2.1%) 3 : 1357 (10.0%) 4 : 11660 (85.9%)	

No	Variable	Stats / Values	Freqs (% of Valid)	Graph
		0 ≤ 4 ≤ 4 IQR (CV) : 0 (0.2)		
15	WordsMemoryA [numeric]	Mean (sd) : 5 (1.8) min ≤ med ≤ max: 0 ≤ 5 ≤ 10 IQR (CV) : 2 (0.4)	11 distinct values	

No	Variable	Stats / Values	Freqs (% of Valid)	Graph
16	NumeracySubtract [numeric]	Mean (sd) : 3.6 (1.9) min ≤ med ≤ max: 0 ≤ 5 ≤ 5 IQR (CV) : 4 (0.5)	0 : 1070 (7.9%) 1 : 2457 (18.1%) 2 : 895 (6.6%) 3 : 544 (4.0%) 4 : 615 (4.5%) 5 : 7996 (58.9%)	
17	WordsMemoryB [numeric]	Mean (sd) : 3.6 (2.2) min ≤ med ≤ max:	11 distinct values	

No	Variable	Stats / Values	Freqs (% of Valid)	Graph
		0 ≤ 4 ≤ 10 IQR (CV) : 3 (0.6)		
18	Mood [numeric]	Mean (sd) : 1.1 (1.2) min ≤ med ≤ max: 0 ≤ 1 ≤ 6 IQR (CV) : 2 (1.2)	0 : 6041 (44.5%) 1 : 3382 (24.9%) 2 : 2334 (17.2%) 3 : 1225 (9.0%) 4 : 418 (3.1%) 5 : 137 (1.0%) 6 : 40 (0.3%)	

No	Variable	Stats / Values	Freqs (% of Valid)	Graph
19	Perspective [numeric]	Mean (sd) : 0.3 (0.5) min ≤ med ≤ max: 0 ≤ 0 ≤ 2 IQR (CV) : 0 (2)	0 : 10474 (77.1%) 1 : 2474 (18.2%) 2 : 629 (4.6%)	
20	Fitness [numeric]	Mean (sd) : 0.8 (0.9) min ≤ med ≤ max:	0 : 6234 (45.9%) 1 : 4355 (32.1%) 2 : 2308 (17.0%) 3 : 680 (5.0%)	

No	Variable	Stats / Values	Freqs (% of Valid)	Graph
		0 ≤ 1 ≤ 3 IQR (CV) : 1 (1.1)		
21	MentalFitness [numeric]	Mean (sd) : 0.5 (0.8) min ≤ med ≤ max: 0 ≤ 0 ≤ 3 IQR (CV) : 1 (1.5)	0 : 8438 (62.1%) 1 : 3457 (25.5%) 2 : 1191 (8.8%) 3 : 491 (3.6%)	

No	Variable	Stats / Values	Freqs (% of Valid)	Graph
22	Social [numeric]	Mean (sd) : 1.3 (2) min ≤ med ≤ max: 0 ≤ 0 ≤ 8 IQR (CV) : 2 (1.5)	0 : 7174 (52.8%) 1 : 2104 (15.5%) 2 : 1432 (10.5%) 3 : 905 (6.7%) 4 : 832 (6.1%) 5 : 354 (2.6%) 6 : 327 (2.4%) 7 : 180 (1.3%) 8 : 269 (2.0%)