Network connectivity as a prognostic marker for depression severity increase

Elsing, Tim

Network Connectivity as a Prognostic Marker for Depression Severity Increase

Student: Tim Elsing, s3632504

Faculty of Social and Behavioural Sciences, Leiden University

Master of Science in Clinical Psychology

Master Thesis

Supervisor: Carlotta L. Rieble

April 9, 2024

Abstract

This thesis aimed to investigate the potential role of overall network connectivity in predicting changes in depression levels over time. It was hypothesized that higher baseline network connectivity would be observed in the group of individuals who experienced an increase in depression severity during the study. Methodologically, a prospective cross-sectional design was chosen. A sample of 489 participants (aged 18 to 48) was assessed at baseline and follow-up with the Patient Health Questionnaire-15 (PHQ-15), which measured depression symptoms and severity for each participant. Based on the PHQ-15 data, statistical analyses were performed, which involved network estimation, centrality index computations, and, as the primary analysis, the Network Comparison Test. Results revealed no significant differences in baseline network structure or connectivity strength between groups of individuals whose depression levels either worsened, improved, or remained stable from baseline to follow-up. Accordingly, contrary to the main hypothesis, baseline network connectivity did not appear to be associated with subsequent changes in depression severity. However, several methodological limitations were identified, which include underpowered sample sizes, potential confounding variables, and homogeneity of the participant sample. This limits the interpretability and generalizability of findings. Recommendations for future research include replication with larger sample sizes, exclusion of participants with previous depressive episodes, and control for baseline variance. Moreover, time-series within-subjects data studies are recommended for studies on clinical utility. In conclusion, while no connection between high network connectivity and subsequent elevation of depression levels could be observed, the findings have to be interpreted with caution due to low power.

Keywords: network connectivity, network approach, major depression

Layman's Abstract

In this thesis, we examined whether people whose depression levels heightened throughout the study have symptoms that, at the beginning of the study, change simultaneously and in the same intensity. We analysed data from a group of individuals, some of whom experienced heightened depression symptoms while others showed improved or stable depression levels. We found that symptom behaviour was similar in all groups, challenging previous beliefs. However, our study had limitations mainly because not enough participants were gathered, which decreased the likelihood of detecting a difference between the groups. For a definite conclusion, future research with more participants is needed in order to confirm or negate our results.

Network Connectivity as a Prognostic Marker for Depression Severity Increase

Major Depressive Disorder (MDD) is one of the most prevalent and fatal mental disorders in modern society (Cuijpers et al., 2012). Despite prolonged scientific efforts to generate data on the construct of depression, it is uncertain which factors are involved in the development of MDD (Bekhuis et al., 2019). Furthermore, the current treatments for depression are not sufficiently effective in reducing depression (Bekhuis et al., 2019). Given the lack of treatment efficacy for MDD, prevention techniques may offer a promising opportunity to reduce depression rates by hindering the disorder from developing in the first place (Bekhuis et al., 2019). In prevention research, vulnerability markers have to be identified. In detail, vulnerability markers are factors that increase the likelihood that a given disorder occurs (Coie et al., 1993; Siddique & Hanif, 2021). For MDD, such a potential marker could potentially be found in cross-sectional network approach studies (Cramer et al., 2016; Lee et al., 2023; van Borkulo et al., 2015). In such studies, symptoms are modelled as nodes in an inter-connected network. The overall strength of connections between symptom nodes of a given network is described as connectivity (van Borkulo et al., 2015).

Although respective results suggest that there may be a link between vulnerability towards depression persistence and network connectivity (van Borkulo et al., 2015), no studies investigated a possible link between depression elevation and network connectivity (see Wichers et al., 2021). Since prevention is about detecting a disorder before clinical levels are developed (Coie et al., 1993), it may be meaningful to investigate whether heightened symptom connectivity levels are associated with depression elevation in the future. Importantly, this would be the first sign that network connectivity could be a vulnerability marker for the development of MDD.

In this introduction section, we will first review the current scientific literature on the network approach. Thereby, we will discuss the role of network connectivity and how it may act as a vulnerability marker. Ultimately, the research objective of this study is explained.

The Network Approach to Psychopathology

Over the last decades, the common cause model (CCM) was the main psychodiagnostic framework to describe how depression is developed and maintained. This specific model explains symptoms as a result of an underlying cause – this underlying cause being the disorder itself: MDD (Bringmann & Eronen, 2018). However, there are several critiques which question the validity of the CCM. In detail, the CCM cannot answer why there are significant differences between individuals concerning the expression of symptoms, the context in which MDD develops, and which comorbidities are present (Bekhuis et al., 2019). Further critique of the CCM entails that it does not investigate how the individual depressive symptoms relate to each other. In other words, covariation between symptoms is solely regarded to be caused by the underlying disorder MDD. Thereby, the interaction between symptoms is omitted (Fried & Nesse, 2015). Additionally, it has been shown that depression is often misdiagnosed and, therefore, mistreated (Bekhuis et al., 2019). State-of-the-art depression diagnosis is based on a conceptual understanding of MDD as defined in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) (Bekhuis et al., 2019). In summation, research suggests that the CCM does not accurately describe depression. And since the DSM uses the theoretical framework of the CCM, it may be argued that current treatments are ineffective due to false diagnostics.

However, there is an alternative model which takes into account the interaction of symptoms. This promising framework is the Network Theory of Psychopathology (NTP), which models MDD as a network of inter-connected symptoms and typically uses between-subjects data (see Robinaugh et al., 2020; Schweren et al., 2018; see Wichers et al., 2021). In

general, a network represents how a specific system is made up of interconnected elements. In detail, the elements that constitute a system are called *nodes*, and the connections that connect the nodes with each other are called *edges* (Bringmann & Eronen, 2018). An everyday example would be to represent a social group as a network – every individual representing a node, and the relationship between every two individuals modelled as an edge. It has been exemplified that representing systems as networks can lead to inferences about how a specific system is functioning in a natural setting (Bringmann & Eronen, 2018).

Applied to the construct of MDD, this *network approach* offers an alternative theory on the working mechanism of depression, which diverges from the CCM model (van Borkulo et al., 2015). In the network approach, instead of the symptoms being the result of the underlying disorder, the disorder is understood as an emergent property of the causal interactions between symptoms. Conceptually, the symptoms that are associated with MDD are modelled as nodes of the network (van Borkulo et al., 2015). Contrary to the mentioned social networks, edges that connect the nodes in an MDD network are statistically estimated since they cannot be overtly observed (Epskamp et al., 2018). Those estimated edges indicate a partial correlation between two nodes while controlling for the influence of all the other nodes of the network (Schweren et al., 2018). Additionally, the edges of the network can be characterized by their weight, which provides information about the strength of the correlation between two nodes (Cramer et al., 2016; van Borkulo et al., 2015). Ultimately, networks are analysed based on correlations between nodes (indicated by the edge-weight) and not based on the severity of symptoms.

Connectivity as a Vulnerability Marker

The following section will discuss the scientific body concerning connectivity in the context of networks. In the first paragraph, we discuss what connectivity is. Then, it is

explained why connectivity may act as a vulnerability marker for depression. Ultimately, empirical findings are reviewed.

Overall network connectivity refers to the sum of absolute edges in a network (see Wichers et al., 2021). In other words, a network in which many nodes are connected with strong edges exhibits a high overall connectivity, whereas a network in which the nodes are sparsely linked will have a low overall connectivity. Concretely, connectivity describes the degree to which the nodes of the network are correlated with each other (see Robinaugh et al., 2020).

Researchers have hypothesized that high overall network connectivity may be linked to MDD vulnerability (Cramer et al., 2016). This hypothesis is based on the assumption that strongly interlinked nodes would create vicious cycles in which a surge in activity in one node would naturally create a surge in activity throughout the entire network, which ultimately would result in a switch to a depressed state (Cramer et al., 2016). As a consequence of this vicious cycle, highly connected networks would be more vulnerable to symptom activation, even without significant external stressors (see Robinaugh et al., 2020). This idea that depression vulnerability is linked to high network connectivity is called the *connectivity hypothesis* (see Wichers et al., 2021).

In this paragraph, three studies are reviewed that have examined the potential of network connectivity as a vulnerability factor for depression persistence. These specific studies are chosen due to their methodological design, which closely resembles this thesis' design. Van Borkulo et al. (2015) investigated a sample of clinically depressed adults. The participants were followed over two years, and depression levels were assessed at baseline and at a 2-year follow-up. Van Borkulo et al. (2015) aimed to discover whether high baseline connectivity levels would be linked to persistent depression levels and low baseline levels to a remittal of depression. The study found that participants whose symptoms did not alleviate

after two years had significantly higher overall connected networks compared to participants whose depression severity improved. Although another study could not replicate this significant finding of van Borkulo et al. (2015), a trend of higher connectivity in the group with persistent depression was observed (Schweren et al., 2018). This second study investigated treatment resistance in adolescent participants. Apart from the participants, the methods and sample size were similar to the original van Borkulo et al. (2015) study (Schweren et al., 2018). A third study by Lee et al. (2023) examined treatment response and its relation to network connectivity. Importantly, Lee et al. (2023) found important differences compared to Schweren et al. (2018) and van Borkulo et al. (2015). The sample size was 100 times larger than in either of the other two studies, which enabled sufficient power. Crucially, the findings show initial support for the connectivity hypothesis. In the context of treatment response, this means that higher connectivity was shown to be associated with treatment resistance and, therefore, depression persistence. However, the effect size was shown to be very small – a sample size of $n = 750$ per group would be needed in order to have a chance of detecting an effect. More importantly, once the researchers controlled for baseline variance, the association between network connectivity and treatment resistance disappeared. In other words, the described study did not support the connectivity hypothesis since the detected effect was ultimately attributed to baseline variance. It can be concluded that there are ambiguous results for network connectivity as a prognostic marker of depression persistence.

Research Objective

This thesis is the first study that investigates whether there is an association between high overall network connectivity and a subsequent elevation of depression severity in a homogenous student population. In detail, the following research question is examined: Does the overall baseline connectivity of depressive symptoms differ for a group of individuals who

develop higher MDD symptom levels over time compared to groups of individuals whose MDD symptoms either improve or stay the same?

A significant result of this thesis would indicate that high connectivity is associated with depression elevation at a later time point. This would support the connectivity hypothesis of the NTP with a divergent study design and a new participant population. If this thesis does not show significant results, it does not support the connectivity hypothesis. It would diversify the study design and provide insight into limitations that have to be regarded by future research in order to have meaningful data.

In general, it is important to note that the relevance of this thesis is strictly research-related, and therefore, one cannot draw any clinical implications based on the results of this thesis. The reason lies in the state of the research field of the NTP (cross-sectional between-subjects network studies on depression) (see Wichers et al., 2021). In general, this research field is still in its beginning phase. This means that the focus is on validating whether the theoretical framework of the NTP is valid. If this framework were validated, empirical research could begin to examine tools for use in clinical practice (see Wichers et al., 2021). However, since findings concerning the NTP are currently ambiguous, this thesis is not focused on investigating clinical utility.

Methods

Design

This study used a prospective cross-sectional between-subjects design. The used data came from the ongoing WARN-D project, which aims to develop a system for detecting early warning signals for the onset of MDD (Fried et al., 2022).

Participants

Data has been collected in cohorts that differed solely in temporal starting point (6 months between the beginning of the study for cohorts one and two). Initially, we retrieved

data from 905 possible participants. Following the *Exclusion Criteria* and the *Participant Selection* specified in the procedures section, 489 participants remained eligible for the study. The participants who were ultimately included were aged 18 to 48 (M = 22.55, SD = 3.60). 82.4% of the participants identified as female, 13.7% as male, and 3.9% indicated another gender. One-third of the participants (n = 163) developed worse overall symptom levels throughout the study, one-third did not change their symptom levels from baseline to follow-up, and one-third improved in their respective depression severity levels from baseline to follow-up.

Measures

This study used an adapted version of the Patient Health Questionnaire-9 (PHQ-9; Kroenke & Spitzer, 2002), which assesses overall depression levels by assessing each of the nine DSM-5 depressive symptoms and overall functional impairment over the past two weeks. This adapted version is called PHQ-15 and assesses 14 symptoms plus the degree of impairment the symptoms cause altogether. The PHQ-15 differs from the PHQ-9 in two ways. First, four items of the PHQ-9 are disaggregated in the PHQ-15. Secondly, two items are added to the PHQ-15, which previously did not exist in the PHQ-9. An example item looks as follows: "Over the last 2 weeks, how often have you been bothered by any of the following problems? 1. Little interest or pleasure in doing things" (Kroenke & Spitzer, 2002, p.6). The item can be given a self-rating score between the anchor points 0 (*not at all*) and 3 (*nearly every day*) (Kroenke & Spitzer, 2002). For a full overview of all items and a detailed description of the differences between both questionnaires, see Appendix A.

Procedure

Participants were recruited via posters, social media, e-mail newsletters, and word-of-mouth. People interested in participating could indicate their e-mail address in an online survey and were then invited to online surveys assessing inclusion and exclusion criteria and

asking for their informed consent. Once participants met the criteria, they were sent and asked to fill out the WARN-D baseline battery, which included the depression severity questionnaire PHQ-15. The total baseline battery took about 75 minutes to complete. After ~4 months, the participants answered a follow-up survey reassessing the most important constructs from the baseline assessment, including the PHQ-15. Participants were reimbursed for their time and effort (7.50€ for completing the baseline battery and another 3.75€ for the follow-up survey; Fried et al., 2022).

Inclusion Criteria

Participants needed to be at least 18 years old and study at a Dutch higher education facility (pursuing an MBO, HBO, or WO degree). In addition, participants had to be fluent in either Dutch or English, own a smartphone with an Android/iOS operating system, and have a European bank account with an IBAN (Fried et al., 2022).

Exclusion Criteria

Based on ethical and research concerns, the exclusion criteria were the following: 1) current moderate to severe depression and 2) current schizophrenia, psychosis, or thought disorder. 3) (hypo)mania or bipolar disorder. 4) primary substance use disorder. 5) moderate or severe suicidal ideation. 6) participants were excluded who found it distressing to see an estimate of daily calories burnt on their smartwatches used in a part of the study that is not relevant to this thesis. Participants were asked if they were currently waiting for or are in treatment by a licensed psychologist/psychiatrist for criteria 1 through 4. Participants indicating yes were excluded. Then, validated self-report screeners were used to exclude potential participants meeting exclusion criteria 1 through 5 (Fried et al., 2022).

Statistical Analyses

All estimations, visual representations, and statistical tests were performed in JASP (v0.18.1.0; JASP Team, 2022) with additional elements performed in R (v4.3.2; R Core Team, 2022).

Participant Selection

Based on the change in overall depression severity from baseline to follow-up (assessed by the PHQ-15), groups were constructed. First, based on the PHQ-9 scores which were reconstructed from the PHQ-15 data, individuals were categorized as having "minimal or none" depression levels (score: 0-4), "mild" depression levels (score: 5-9), "moderate" depression levels (score: 10-14), "moderately severe" depression levels (score: 15-19), or "severe" depression levels (score: 20-25) (PHQ-9; Kroenke & Spitzer, 2002). See Appendix A for a detailed description of how the PHQ-9 scores were reconstructed from the PHQ-15 data. Secondly, if depression levels worsened from baseline to follow-up, the respective participant was categorized into the *Worse* group. A worsening of depression level means that, at follow-up, the participant would reach a higher depression score compared to the baseline score, which would also put them in a different depression level categorization. On the contrary, if the participant's depression score was lower at follow-up and could be assigned to a different depression level compared to the baseline score, the participant was assigned to the *Improved* group. Lastly, if the depression level remained constant from baseline to follow-up, individuals were classified into the *same* group.

Based on this initial method to construct groups, significantly more individuals were allocated to the *Same* (n=367) and *Improved* group (n=216) compared to the *Worse* group (n=163). However, a simulation study shows that the NCT has increased power when group sizes are equal (van Borkulo et al., 2023). Empirically, this was verified by an experimental study, which found that unequal subgroup sizes are associated with decreased power and low

sensitivity (Steen et al., 2021). Accordingly, we equalized the sample size for all groups (Lee et al., 2023). This was done by, first, randomly deleting participants until all groups had the same sub-sample size*, n*. Secondly, we controlled for fundamental changes in mean and standard deviation, possibly induced by the random deletion of participants. If, after the deletion, a group with removed participants would stay robust in mean and standard deviation of the included PHQ-15 items compared to itself before the deletion was performed, it could be inferred that this group would not be significantly altered by the deletion process. In that case, subsequent statistical analyses would not be affected (Fields, 2013; Kazdin, 2021). For a detailed description of how sub-sample sizes were equalized, see Appendix C. Importantly, the overall group values and general scores concerning the PHQ-items remained stable irrespective of whether participants were deleted or not. As a consequence, the statistical analyses were carried out with equal group sizes (n =163). It follows that in the remaining main text of this thesis, all reported results are based on the design with equal group sizes. The results based on unequal group sizes can be found in Appendix F.

Network Estimation and Analysis

Following the construction of the groups, one symptom network was estimated for each group separately. Based on the depression symptom scores at baseline, the respective network structures for participants of groups *Worse, Improved,* and *Same* were estimated with EBIC gLASSO-regularized partial correlations (Burger et al., 2022; Epskamp et al., 2018). The EBIC gLASSO regularization technique was used to control for spurious edges (Epskamp & Fried, 2018). Ultimately, the estimated networks were then visualized via a Gaussian Graphical Model (Foygel & Drton, 2010).

In general, networks can only be adequately estimated if there are enough participants in each group. This ensures that a significant effect can be detected if there is one in the actual population. The degree to which an effect that exists in the population is statistically

detectable is indicated by the power of a given statistical analysis. Regarding this thesis, to obtain a sufficient power of at least 80% (when the significance level is .05 and the effect size is .2), a network with a maximum of six nodes (15 parameters) could be estimated (Epskamp et al., 2018). See Appendix B for more information on the formula. Accordingly, this study estimated networks including six nodes. Following prior research, we included the symptoms of *loss of interest, depressed mood, hopelessness, fatigue, worthlessness*, and *concentration problems* as nodes for each group network (Schweren et al., 2018; van Borkulo et al., 2015).

After the network estimation, non-parametric bootstrapping of confidence intervals for both groups was performed with the R-package *bootnet* (v1.6; Epskamp et al., 2018) to check for the accuracy of the estimated edges.

Centrality Indices

After the networks for each group were estimated, centrality indices were computed, which describe the nodes and edges of each network in detail. Since it is assumed that highly connected nodes are more important than less connected ones, centrality indices measure the degree to which a given node has connections to all other nodes of the network (Santos et al., 2017). In our study, the connectivity of the nodes was examined via the centrality indices *of node strength, closeness*, and *betweenness*. *Node strength* indicates the overall connectivity a given node had with all other symptom nodes of the network (McNally, 2021). *Closeness* examines how close the respective node was to other nodes in the network structure spatially. Lastly, *betweenness* investigates the level with which a specific node acted as an intermediary along the shortest paths between other pairs of nodes in a network (van Borkulo et al., 2015). Despite the information centrality indices provided on the included nodes of a network, they are not suitable for making causal inferences (Bringmann et al., 2019). The centrality indices purely function as descriptive information for the respective networks. After the centrality

indices were computed, we performed case-dropping bootstrapping with *bootnet* to check for their accuracy (Epskamp et al., 2018).

Network Comparison

Once all networks were estimated and described, the structure and connectivity of each network was compared with each other. These network comparisons were the main analyses of this thesis since they answered the research question of whether connectivity would be significantly higher for group *Worse* compared to groups *Same* and *Improved*. The estimated group networks were compared via the Network Comparison Test (NCT; van Borkulo et al., 2015), which consists of two main tests. First, the Global Strength Invariance Test of the NCT tested the null hypothesis that the depression networks of the compared groups are equal in overall connectivity. Secondly, the Network Invariance Test assessed whether the compared groups would differ in network structure. In total, three comparisons were performed (*Improved* vs. *Worse, Improved* vs. *Same*, *Worse* vs. *Same*) via both the Global Strength Invariance Test and the Network Invariance Test.

Results

Network Estimation

In this section, we provide an overview of the key characteristics of the respective estimated networks. First, we describe whether groups differ in how the PHQ-items were answered. Then, the zero-order correlations of edges of the estimated networks are described. Further, following the assessment of the networks' edge accuracy, the estimated networks are shown. Additionally, the networks are compared with the zero-order correlations. Ultimately, the centrality indices are examined, given that they are sufficiently stable.

In Table 1, the groups were compared based on the results of the respective PHQ-items that are included in the network estimations. For all six items, the *Improved* group showed a higher mean value compared to the other groups. However, this difference is not

tested for significance via a statistical test. The standard deviations varied across groups and items, with a maximum value of 1.0 and a minimum value of 0.6. This means that the scores vary moderately within the groups.

Table 1

Descriptive Statistics

	Loss of Interest			Depressed Mood			Hopelessness			Fatigue			Worthlessness			Concentration Problems		
Groups	*Imp*	*Worse*	*Same*	*Imp*	*Worse*	*Same*	*Imp*	*Worse*	*Same*	*Imp*	*Worse*	*Same*	*Imp*	*Worse*	*Same*	*Imp*	*Worse*	*Same*
M	1.1	0.5	0.6	1.1	0.5	0.5	1.0	0.4	0.4	2.0	1.4	1.3	1.1	0.6	0.5	1.6	0.8	0.9
SD	0.8	0.6	0.6	0.8	0.6	0.7	0.9	0.6	0.6	0.9	0.8	0.8	0.9	0.7	0.7	1.0	0.8	0.9
n	163	163	163	163	163	163	163	163	163	163	163	163	163	163	163	163	163	163

Note. In this Table, the groups Improved, Worse, and Same were compared for each of the PHQ-15 items that are included in the network analysis.

Tables 2 to 4 display the zero-order correlations for all edges that theoretically could be included in the networks for each group separately. Across groups, the strongest symptom-symptom correlations consistently were *Depressed mood-Hopelessness* (*Improved:* .65; *Worse:* .55; *Same:* .63) and *Depressed Mood-Loss of Interest* (*Improved:* .56; *Worse:* .46; *Same:* .57). The third strongest symptom-symptom correlation differed depending on the group (*Improved: Depressed Mood-Worthlessness* = .53; *Worse: Loss of interest-Hopelessness* = .41; *Same: Hopelessness-Worthlessness* = .56). In general, the *Worse* groups exhibited slightly less strong correlations compared to the *Same* and *Improved* groups' correlations.

The edge-weights of the EBIC gLASSO-regularized partial correlations network showed a similar picture. The *Depressed Mood-Hopelessness* edge was the strongest for all groups (*Improved:* .43; *Worse*: .37; *Same*: .36). The D*epressed Mood-Loss of Interest* was the second strongest edge in groups *Improved* (.35) and *Worse* (.24), and the third strongest in group *Same* (.28). Importantly, no difference test was computed which would have clarified whether the zero-order correlations/ edge-weights would be significantly stronger than other correlations/ edge-weights from the same group.

Table 2

Zero-Order Correlations of Group Improved

Variable	Loss of Interest	Depressed Mood	Hopelessness	Fatigue	Worthlessness	Concentration Problems
Loss of Interest	/					
Depressed Mood	.56	/				
Hopelessness	.45	.65	/			
Fatigue	.27	.20	.24	/		
Worthlessness	.36	.53	.50	.19	/	
Concentration Problems	.21	.17	.26	.39	.27	/

Note. This Table shows the zero-order correlations of each edge that could be included in the *Improved* group network.

Table 3

Zero-Order Correlations of Group Worse

Variable	Loss of Interest	Depressed Mood	Hopelessness	Fatigue	Worthlessness	Concentration Problems
Loss of Interest	/					
Depressed Mood	.46	/				
Hopelessness	.41	.55	/			
Fatigue	.34	.35	.30	/		
Worthlessness	.20	.34	.37	.20	/	
Concentration Problems	.27	.28	.24	.22	.17	/

Note. This Table shows the zero-order correlations of each edge which could be included in the *Worse* group network.

Table 4

Zero-order Correlations of Group Same

Variable	Loss of Interest	Depressed Mood	Hopelessness	Fatigue	Worthlessness	Concentration Problems
Loss of Interest	/					
Depressed Mood	.57	/				
Hopelessness	.47	.63	/			
Fatigue	.42	.28	.25	/		
Worthlessness	.44	.51	.56	.13	/	
Concentration Problems	.43	.35	.33	.40	.33	/

Note. This Table shows the zero-order correlations of each edge, which could be included in the *Same* group network.

Description of Estimated Networks

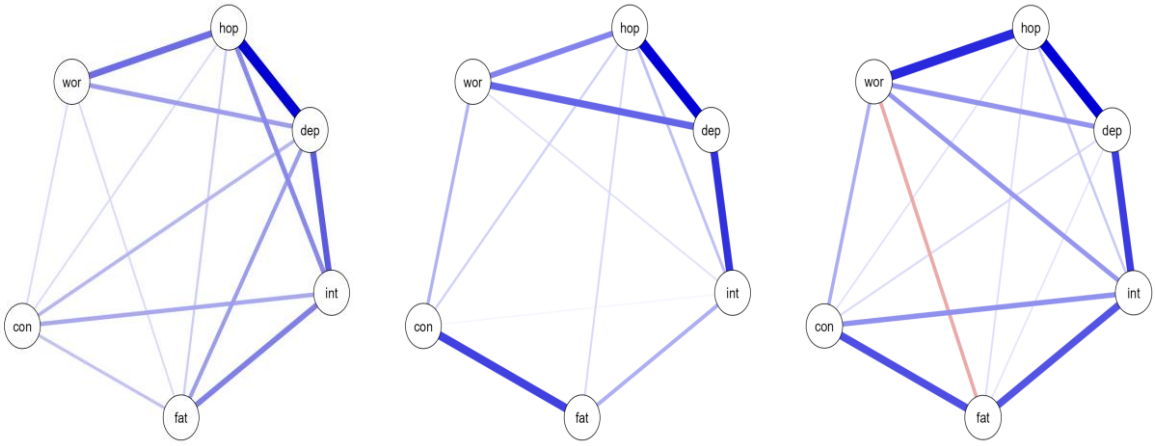
In Figure 1, the group networks based on partial correlations are displayed. The networks were visualized via the Gaussian Graphical Model (Foygel & Drton, 2010). In Appendix D, the non-parametric bootstrapped 95% confidence intervals (CI) are displayed, which assessed the interpretability of the networks. Based on the information of the bootstrapped confidence intervals, it can be concluded that the networks were accurately

modelled, and their data was interpretable (Epskamp et al., 2018). More detailed information about the bootstrapped confidence intervals can be found in Appendix D.

The density of each network was assessed with the *degree of sparsity.* It shows how many edges were included in a given network relative to the maximum amount of possible edges – a value of 0 indicates that no edges were removed, and the closer the value gets to 1, the more edges were removed as a consequence of weak correlation between the respective nodes (Liu et al., 2008). With regards to our group networks, the *Same* network was the densest network (*degree of sparsity:* 0.0), followed by the *Worse* network (*degree of sparsity:* 0.07) and the *Same* network (*degree of sparsity:* 0.2). *Degree of sparsity* was calculated by dividing edges which have a correlation of zero with the total amount of possible edges of a network. Because edges with the value of 0 were removed in our EBIC-gLASSO regularized network, the fewer edges were zero, the fewer were removed, the higher was the density of a network (Epskamp & Fried, 2018).

Figure 1

Visualized Networks of the Improved, Worse, and Same Groups



Note. Network structures of the I*mproved* (left), *Worse* (middle), and *Same* (right) groups based on baseline data. Node names are shortened in order to fit in the figure: int = loss of interest; dep = depressed mood; hop = hopelessness; fat = fatigue; wor = worthlessness; con = concentration problems. Edges were visualized as lines which connect the lines. The colour blue indicates a positive association between nodes. The colour red indicates a negative association.

Centrality Indices

Before the centrality indices can be reported, the stability of the centrality indices is reported. Only the *Worse* group exhibited stable indices, whereas the two remaining groups were unstable. As a consequence, the groups were not compared to each other regarding the centrality indices and only the results of the *Worse* group are displayed in this thesis. More specifically, the centrality index *strength* was the most stable out of the computed centrality indices in group Worse. Accordingly, only the centrality index strength is reported. The centrality stability tests and a detailed description of their implications can be found in Appendix E. Concerning the results, the node *depressed mood* exhibited the highest edge

strength in group *Worse* (1.40). The node *Hopelessness* (0.88) was the second strongest node (see Table 5).

Table 5

Strength Centrality Values of Group Worse

Variable	Centrality Index *Strength*
Loss of Interest	0.21
Depressed Mood	1.40
Hopelessness	0.88
Fatigue	-0.55
Worthlessness	-0.90
Concentration Problems	-1.05

Note. This Table shows the values for the centrality index s*trength* for every node of the W*orse* group network.

Network Comparison

The NCT's results of the Global Strength Test (comparison of network connectivity) and the Network Invariance Test (comparison of network structure) are displayed in this section. First, the Global Strength Test did not indicate significant differences for any of the group comparisons (Global Strength Test: *Improved/ Worse: p* = .247; *Improved/ Same: p* = .960; *Worse/ Same: p* = .545). This suggests that the group networks did not differ regarding their overall connectivity. Secondly, the Network Invariance Test also did not show any significant differences between groups (Network Invariance Tests: *Improved/ Worse: p* = .287; *Improved/ Same: p* = .782; *Worse/ Same: p* = .950). This implies that no evidence was found which would indicate a difference between groups regarding the structure of the networks.

Discussion

Summary

Although Major Depressive Disorder (MDD) is one of the most prevalent mental disorders (Cuijpers et al., 2012), there are conflicting ideas as to which elements increase the likelihood of developing or maintaining MDD (Bekhuis et al., 2019). Studies on the Network Theory of Psychopathology (NTP) have suggested that network connectivity may be a potential marker for depression persistence (Cramer et al., 2016; van Borkulo et al., 2015). However, further studies could not verify an association between depression and network connectivity (Lee et al., 2023; Schweren et al., 2018). Whilst studies on the NTP grew over the last decade, a critical scientific gap remained: It is uncertain whether network connectivity can serve as a prognostic marker for the elevation of depression levels in the future (see Wichers et al., 2021). To address this gap, we aimed to investigate whether the overall baseline connectivity of a depression network differs for a group of individuals who experience an increase in MDD symptoms over time compared to a group of individuals whose symptoms either improve or remain stable.

Our results could not support the hypothesis that higher baseline network connectivity would be associated with a subsequent increase in depression severity over time. In the main analysis, we did not find any significant differences between groups in terms of network structure or connectivity strength. Furthermore, the centrality indices were not comparable between groups since the indices of groups *Improved* and *Same* were not stable enough to be interpretable. However, the stable centrality index *strength* of group *Worse* suggested that *Depressed Mood* had the highest overall connectivity with all other nodes of the network, which would make it an important hub node. Ultimately, the results of the zero-order correlations implied that the symptom-symptom correlation *Depressed Mood-Hopelessness* was consistently the strongest correlation within groups, which was also observed with the edge-weights of the estimated networks.

Comparison With Previous Research

Of the between-subject NTP studies, van Borkulo et al. (2015), Schweren et al. (2018), and Lee et al. (2023) are the most comparable with this thesis due to similar study design. Although these studies all reported either a significant effect of network connectivity or a non-significant trend towards significance for an association between depression persistence and network connectivity, this thesis did not observe a significant finding nor a trend towards it. Importantly, whereas all mentioned studies were similar to our study in their method of estimating and comparing networks, our study investigated depression elevation instead of depression persistence (Lee et al., 2023; Schweren et al., 2018; van Borkulo et al., 2015). This difference in the dependent variable may have led to a different outcome. Generally, in the following section, possible reasons for the observed null finding of this thesis are discussed.

Interpretation of Results

Network Comparison

The null finding that prospective network structure and connectivity are not related to subsequent depression elevation may have several possible reasons, which can be found in the methodological setup of our study and in the theoretical framework of the NTP. Regarding the methodology of this study, previous research has shown that network studies require a minimum $n = 750$ per group in order to ensure sufficient power (Lee et al., 2023). However, this study merely entails $n = 163$ for every group. Due to this experiment having 487 too few participants for each group, it can be concluded that our thesis is severely underpowered. Accordingly, a null finding is coherent with the high likelihood of a Type-II error (false negative), which is the result of a lack of power (Field, 2013).

Concerning the theoretical framework of this thesis, it can be categorized as belonging to the Network Theory of Psychopathology (NTP), which typically uses cross-sectional between-subjects data to construct groups (see Wichers et al., 2021). A rivaling framework within the network approach is the Complementary Momentary Affective Network Theory (MAD), which typically uses time-series within-subjects data to construct networks of individuals (see Wichers et al., 2021). Already existing research shows that studies based on the NTP tend to exhibit more null findings compared to studies based on the MAD theory (see Wichers et al., 2021). One can assume that connectivity, as defined by the NTP, may simply have no effect on depression elevation in the actual population.

Furthermore, the usage of cross-sectional between-subjects data and its underlying assumptions are generally controversial for multiple reasons. First, the assumption that negative feedback loops are the reason for an effect of network connectivity on depression onset, elevation, or persistence is not scientifically proven by individual networks (see

Wichers et al., 2021). This is problematic because it means that our study draws inferences about individual network behaviour based on a method that is not scientifically validated for that purpose. Secondly, this thesis assumes that network parameters remain constant over time, which is in line with NTP theory. However, studies based on time-series data show that network parameters of individual networks change prior to a shift into a depressed state (Nemesure et al., 2024).

All in all, the methodological setup and theoretical assumptions of this thesis may be inadequate to investigate the effect of network connectivity on depression elevation. As a consequence, the observed null finding may have occurred due to the methodological and theoretical setup of the study.

Network Descriptive Statistics

Centrality Indices. The centrality index *strength* of the only stable group, *Worse*, showed *Depressed Mood* and *Hopelessness* as the strongest nodes of the network. This is consistent with previous literature, which mainly identifies *Depressed mood* as the most important hub node of a given network (Schweren et al., 2018; van Borkulo et al., 2015). As a consequence, it can be inferred that *Depressed Mood* is an important node for a given group depression network. However, centrality indices have to be interpreted with caution since their scores do not substitute a significance test (see Wichers et al., 2021).

Zero-Order Correlations and Edge-Weights. *Depressed Mood-Hopelessness* consistently appears to have the strongest connection based on both the zero-order correlations and the edge-weights. This is not in line with prior research, which often found *Depressed Mood-Loss of Interest* to be of high importance (Lee et al., 2023; van Borkulo et al., 2015). Our result can be explained by the disaggregation of the item *Depressed Mood* in the PHQ-15. In previous studies, depressed mood and feeling hopeless were assessed together

in one item, whereas in the PHQ-15 questionnaire that we used, they were disaggregated (Fried et al., 2022). Accordingly, one can assume that *Depressed Mood-Hopelessness* would remain closely correlated after disaggregation. Furthermore, *Depressed Mood-Loss of Interest* was shown to be the second strongest symptom-symptom correlation and second or third strongest edge-weight in our study, which is in line with previous research (Lee et al., 2023; van Borkulo et al., 2015).

Limitations

In this section, the limitations of this thesis are discussed with regard to the design, the sample group, the included measures, and the statistical analysis. We start with the general experimental setup of this study, which is the first to compare network connectivity between groups of individuals whose depression levels worsened, improved, or stayed the same during the study. In our study, there are five distinct levels of depression severity – none or minimal, mild, moderate, moderately severe, and severe. Since this study only took into account change over time, a participant who changed from severe to moderately severe depression may end up in the same group as someone who changed from severe to mild depression. In other words, the magnitude of change is not considered in this thesis' design. Secondly, a participant who stayed severely depressed would end up in the *Same* group, whereas an individual who changed from none to mild depression would be categorized in the *Worse* group. This means that the actual depression level was not accounted for. One may argue that the omission of the actual depression levels may have distorted the data.

Alternatively, instead of controlling for change in depression severity over time, we could have excluded participants who had prior MDD episodes or MDD at baseline – participants with depression severity of moderate or higher would be classified as having MDD (Kroenke & Spitzer, 2002). Insights as to why it could be meaningful to exclude participants with prior or current depression experiences come from comparisons with the

brain. The brain is the most prominent biological system, which can be modelled as a network (Bringmann & Eronen, 2018). Conceptually, the nodes would represent nuclei and edges axons. In the brain, Hebbian learning is a key mechanism of how cognitive and behavioural patterns are learned. In detail, Hebbian learning is the principle that neurons that fire together will have a stronger connection in the future (Munakata & Pfaffly, 2004). It can be deduced that the connection between nodes in an MDD network may operate in a similar way. In other words, when an individual experiences a depressive episode, various symptoms of the MDD network would occur together, and thus, the connection between those symptoms may increase. Consequently, high overall connectivity in edges between nodes could be the consequence of illness duration rather than there being differences in groups of individuals by default (Schweren et al., 2018). Accordingly, an exclusion of participants with baseline or prior MDD would be meaningful. However, in our thesis, this was not possible because it would have required the exclusion of too many participants, which would have rendered the already low power too small to have the possibility of detecting a significant outcome (Field, 2013).

Regarding the study sample of this thesis, the selected participants are predominantly Dutch students. As a consequence, the sample represents a very specific sub-group of the population. Therefore, the results cannot be generalized to the general public, which is made up of humans who do not study in the Netherlands. In other words, the external validity of the thesis is low due to a largely homogeneous sample group (Field, 2013).

Furthermore, concerning the measures, this thesis used a modified version of the renowned PHQ-9 questionnaire, which is a valid questionnaire with high inter-rater reliability (Kroenke & Spitzer, 2002). The modified version PHQ-15, which was used in this version, has a high correlation with the original PHQ-9 questionnaire and also exhibits high inter-rater reliability (Fried et al., 2022). However, researchers criticize the use of questionnaires to

construct partial-correlation networks (see Wichers et al., 2021). In detail, node selection based on questionnaire data is not optimal for constructing networks because it does not control for possible underlying causes of the depression symptoms, which are directly measured with the questionnaire.

The main limitation of the statistical analysis and of this thesis, in general, was the lack of sufficient power. Research shows that $n = 750$ per group is required for adequate power in network studies (Lee et al., 2023). This study had less than half of the required participants per group ($n = 163$). This severely limits the interpretation of the observed null finding because the likelihood of finding no significant effect was extraordinarily high by default.

Furthermore, the groups in this thesis exhibited a considerable amount of variance, as indicated by the *SD*. In addition, SD scores differed between groups (see Table 1). Previous research shows that uncontrolled baseline variance may confound the results (Lee et al., 2023). Despite the risk of confounding results if the variance is not controlled, it would have required the exclusion of participants, which would have limited the power of this thesis even more (Lee et al., 2023). Therefore, no control of variance was included. Lastly, in this thesis, centrality indices were not tested for their ordering via bootstrapping. This limits the interpretability since we cannot say whether *Depressed Mood* is significantly more central than other nodes of the *Worse* network (see Wichers et al., 2021).

Implications for Future Research and Clinical Practice

The null finding of this study implies that connectivity is not related to subsequent depression elevation. However, as discussed in the interpretation of the results, low power increases the likelihood of a Type-II error – not detecting an effect that is actually present in the population. Moving forward, future research should first aim to ensure sufficient power by

establishing a sample size of $n = 750$ (Lee et al., 2023). If power is not restricted, future studies may use an alternative experimental setup to investigate changes in depression level prospectively. In this alternative design, individuals with current or previous MDD (indicated by depression severity of moderate or higher) should be excluded in order to control for the influence of prior MDD experiences on the participants' network structure (Munakata & Pfaffly, 2004; Schweren et al., 2018). Then, two groups can be formed based on whether or not a switch to MDD occurs throughout the study. In other words, one group with a minimum of $n = 750$ would not develop MDD throughout the study, and one study with at least $n = 750$ would develop MDD. In this way, everyone would be on the same depression level at baseline (less than moderate). Further, group 1 would consist of everyone who switched to moderate or higher depression levels, and everyone who stayed the same would join Group 2. Regarding the statistical analysis, it is important for future studies to control for baseline variance since it could otherwise confound the results (Lee et al., 2023). Moreover, EBIC gLASSO regularized partial correlation networks should be estimated and compared with the NCT, as depicted in this thesis (Epskamp & Fried, 2018). In addition, it should be assessed whether the estimated networks can be interpreted via edge accuracy tests (Epskamp et al., 2018). This would provide an optimal design for cross-sectional between-subjects data studies that investigate whether connectivity is associated with depression level increase or MDD onset.

Alternatively to a cross-sectional between-subjects study, one may investigate depression elevation with time-series within-subjects data according to the typical MAD design. This is recommendable due to the possibility of MAD studies to study networks of single individuals and how their networks change over time (see Wichers et al., 2021). To be able to compute networks that can change in structure over time, it is important to use time-varying vector autoregressive models that account for parameter change within networks over time (Nemesure et al., 2024).

Another advantage of time-series within-subjects studies is their potential clinical utility. In comparison to cross-sectional between-subjects studies, MAD studies have shown that with individual time-series data, it is possible to investigate early warning signals (EWS) which indicate a potential switch to a depressed state (Wichers et al., 2016; Wichers et al., 2020). Contrarily, current cross-sectional between-subjects studies merely try to prove that connectivity is associated with MDD persistence, elevation, or onset (Lee et al., 2023; Schweren et al., 2018; van Borkulo et al., 2015). However, no possible measures are investigated which may be used as preventive monitoring tools with preventive utility.

In summation, future research in the field of the NTP primarily need high sample sizes and a stringent methodology in order to yield reliable data that may show whether there is an association between network connectivity and MDD persistence, elevation, or onset (Epskamp et al., 2018; Epskamp & Fried, 2018; Lee et al., 2023; Schweren et al., 2018). For clinical implications, MAD studies that investigate potential EWS offer promising opportunities (Wichers et al., 2016; Wichers et al., 2020).

Conclusion

The study at hand does not show any association between network connectivity and change in depression levels. However, this result has to be interpreted with caution due to low power, which increases the likelihood of a Type-II error. Based on this thesis, it can be concluded that there seems to be no connection between high network connectivity and a subsequent elevation in depression levels. For a more confident answer, future research with higher sample sizes and a more stringent methodology will be necessary to investigate our research question.

References

Bekhuis, E., Hartman, T. C. O., Boschloo, L., & Lucassen, P. L. (2019). A new approach to psychopathology: The example of depression. *British Journal of General Practice, 69*(680), 146-147. https://doi.org/10.3399/bjgp19X701717

Bringmann, L. F., Elmer, T., Epskamp, S., Krause, R. W., Schoch, D., Wichers, M., Wigman, J. T. W., & Snippe, E. (2019). What do centrality measures measure in psychological networks? *Journal of Abnormal Psychology, 128*(8), 892–903. https://doi.org/10.1037/abn0000446

Bringmann, L. F., & Eronen, M. I. (2018). Don't blame the model: Reconsidering the network approach to psychopathology. *Psychological Review, 125*(4), 606–615. https://doi.org/10.1037/rev0000108

Burger, J., Isvoranu, A. M., Lunansky, G., Haslbeck, J., Epskamp, S., Hoekstra, R. H., Fried, E. I., Borsboom, D., & Blanken, T. F. (2022). Reporting standards for psychological network analyses in cross-sectional data. *Psychological Methods, 28*(4), 806–824. https://doi.org/10.1037/met0000471

Coie, J. D., Watt, N. F., West, S. G., Hawkins, J. D., Asarnow, J. R., Markman, H. J., Ramey, S. L., Shure, M. B., & Long, B. (1993). The science of prevention: A conceptual framework and some directions for a national research program. *American Psychologist*, *48*(10), 1013-1022. https://doi.org/10.1037/0003-066X.48.10.1013

Cramer, A. O., van Borkulo, C. D., Giltay, E. J., Van Der Maas, H. L., Kendler, K. S., Scheffer, M., & Borsboom, D. (2016). Major depression as a complex dynamic system. *Plos One*, *11*(12). https://doi.org/10.1371/journal.pone.0167490

Cuijpers, P., Beekman, A. T., & Reynolds, C. F. (2012). Preventing depression: A global priority. *JAMA Psychiatry*, *307*(10), 1033-1034. https://doi.org/10.1001/jama.2012.271

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

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage.

Foygel, R., & Drton, M. (2010). Extended Bayesian information criteria for Gaussian graphical models. *Advances in Neural Information Processing Systems, 23,* 1-9.

Fried, E. I., & Nesse, R. M. (2015). Depression sum-scores don't add up: Why analyzing specific depression symptoms is essential. *BMC Medicine, 13*(72), 1-11. https://doi.org/10.1186/s12916-015-0325-4

Fried, E. I., Proppert, R., & Rieble, C. L. (2022). Building an early warning system for depression: Rationale, objectives, and methods of the WARN-D study. *Clinical Psychology in Europe, unpublished*.

JASP Team (2022). JASP (Version 0.18.1.0) [Computer software].

Kazdin, A. E. (2021). *Research Design in Clinical Psychology.* Cambridge University Press.

Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals*, *32*(9), 509-515. https://doi.org/10.3928/0048-5713-20020901-06

Lee, C. T., Kelley, S. W., Palacios, J., Richards, D., & Gillan, C. M. (2023). Estimating the prognostic value of cross-sectional network connectivity for treatment response in depression. *Psychological Medicine*, *54*(2), 317-326. https://doi.org/10.1017/S0033291723001368

Liu, Y., Liang, M., Zhou, Y., He, Y., Hao, Y., Song, M., Yu, C., Liu, H., Liu, Z., & Jiang, T. (2008). Disrupted small-world networks in schizophrenia. *Brain*, *131*(4), 945-961. https://doi.org/10.1093/brain/awn018

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics, 18*(1), 50-60.

Munakata, Y., & Pfaffly, J. (2004). Hebbian learning and development. *Developmental Science, 7*(2), 141-148. https://doi.org/10.1111/j.1467-7687.2004.00331.x

McNally, R. J. (2021). Network analysis of psychopathology: Controversies and challenges. *Annual Review of Clinical Psychology, 17*, 31-53. https://doi.org/10.1146/annurev-clinpsy-081219-092850

Nemesure, M. D., Collins, A. C., Price, G., Griffin, T. Z., Pillai, A., Nepal, S., Heinz, M. V., Lekkas, D., Campbell, A. T., & Jacobson, N. C. (2024). Depressive symptoms as a heterogeneous and constantly evolving dynamical system: Idiographic depressive symptom networks of rapid symptom changes among persons with major depressive disorder. *Journal of Psychopathology and Clinical Science, 133*(2), 155–166. https://doi.org/10.1037/abn0000884

R Core Team (2022). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. https://www.r-project.org/

Robinaugh, D. J., Hoekstra, R. H., 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

Santos, H., Fried, E. I., Asafu-Adjei, J., & Ruiz, R. J. (2017). Network structure of perinatal depressive symptoms in Latinas: Relationship to stress and reproductive biomarkers. *Research in Nursing & Health*, *40*(3), 218-228. https://doi.org/10.1002/nur.21784

Schweren, L., van Borkulo, C. D., Fried, E., & Goodyer, I. M. (2018). Assessment of symptom network density as a prognostic marker of treatment response in adolescent depression. *JAMA Psychiatry*, *75*(1), 98-100. https://doi.org/10.1001/jamapsychiatry.2017.3561

Siddique, M. P., & Hanif, R. (2021). Measurement of vulnerability markers for depression: A study on translation, and validation of DAS-A. *Journal of Islamic International Medical College (JIIMC)*, *16*(1), 32-36.

Steen, O. D., van Borkulo, C. D., & van Loo, H. M. (2021). Symptom networks in major depression do not diverge across sex, familial risk, and environmental risk. *Journal of Affective Disorders*, *294*, 227-234. https://doi.org/10.1016/j.jad.2021.07.002

van Borkulo, C. D., Boschloo, L., Borsboom, D., Penninx, B. W., Waldorp, L. J., & Schoevers, R. A. (2015). Association of symptom network structure with the course of depression. *JAMA Psychiatry*, *72*(12), 1219-1226. https://doi.org/10.1001/jamapsychiatry.2015.2079

van Borkulo, C. D., van Bork, R., Boschloo, L., Kossakowski, J. J., Tio, P., Schoevers, R. A., Borsboom, D., & Waldorp, L. J. (2023). Comparing network structures on three aspects: A permutation test. *Psychological Methods, 28*(6), 1273–1285. https://doi.org/10.1037/met0000476

Wichers, M., Arnout C. Smit, A. C., & Snippe, E. (2020). Early warning signals based on momentary affect dynamics can expose nearby transitions in depression: A confirmatory single-subject time-series study. *Journal for Person-Oriented Research*, *6*(1), 1-15. https://doi.org/10.17505/jpor.2020.22042

Wichers, M., Groot, P. C., Psychosystems, E. S. M., & EWS Group. (2016). Critical slowing down as a personalized early warning signal for depression. *Psychotherapy and Psychosomatics*, *85*(2), 114-116. https://doi.org/10.1159/000441458

Wichers, M., Riese, H., Hodges, T. M., Snippe, E., & Bos, F. M. (2021). A narrative review of network studies in depression: What different methodological approaches tell us about depression. *Frontiers in Psychiatry*, *12*, 1-16. https://doi.org/10.3389/fpsyt.2021.719490

Appendix A

Patient Health Questionnaire-15

In this Appendix, an overview of all PHQ-15 items is provided. In the following section, it is described how the PHQ-15 differs from the PHQ-9. In the last section, we detail how to construct PHQ-9 severity scores based on PHQ-15 data.

Items

1. Little interest or pleasure in doing things (PHQ_1)
2. Feeling down or depressed (PHQ_2)
3. Feeling hopeless (PHQ_3)
4. Trouble falling asleep or staying asleep (PHQ_4)
5. Sleeping too much (PHQ_5)
6. Feeling tired or having little energy (PHQ_6)
7. Poor appetite (PHQ_7)
8. Overeating (PHQ_8)
9. Feeling bad about yourself – or that you're a failure or have let yourself or your family down (PHQ_9)
10. Trouble concentrating on things, such as reading or watching television (PHQ_10)
11. Moving or speaking so slowly that other people could have noticed (PHQ_11)
12. Being so fidgety or restless that you have been moving around a lot more than usual (PHQ_12)
13. Thoughts that you would be better off dead or of hurting yourself in some way (PHQ_13)
14. Little interest in sex (PHQ_14)
15. If you checked any problems, how difficult have they made it for you to do your work, take care of things at home, or get along with other people? (PHQ-15)

Items 1 to 14 can be scored as either “Not at all (0)”, “Several Days (1)”, “More Than Half the Days (2)”, or “Nearly Every Day (3)”. Item 15 can be scored as “Not difficult at all (0)”, “Somewhat difficult (1)”, “Very difficult (2)”, or “Extremely difficult (3)”.

Comparison to the PHQ-9

Compared to the PHQ-9, the PHQ-15 includes six more questions. In detail, the PHQ-15 splits four questions of the PHQ-9 into two. These disaggregated items are the following. First, items 2 and 3 form item number 2 in the PHQ-9: "Feeling down, depressed, or hopeless…" (Kroenke & Spitzer, 2002, p. 6). Secondly, items 4 and 5 form item number 3 in the PHQ-9: "Trouble falling or staying asleep, or sleeping too much…" (Kroenke & Spitzer, 2002, p. 6). Items 7 and 8 of the PHQ-15 are merged in the PHQ-9 as: "Poor appetite or overeating…" (item number 5) (Kroenke & Spitzer, 2002, p. 6). The last disaggregated items are items 11 and 12, which form the 8th PHQ-9 item: “Moving or speaking so slowly that other people could have noticed? Or the opposite – being so fidgety or restless that you have been moving around a lot more than usual…” (Kroenke & Spitzer, 2002, p.6). Items 14 and 15 are new in the PHQ-15 and do not exist in the PHQ-9 (Fried et al., 2022; Kroenke & Spitzer, 2002).

Construction of PHQ-9 Severity Scores Based on PHQ-15 Data

The PHQ-9 score, which indicates the severity of depression, was computed from the PHQ-15 data in the following way. Regarding the disaggregated items in the PHQ-15 (items 2/3, items 4/5, items 7/8, and items 11/12), the respectively higher score of both items was used as the overall score for their aggregated PHQ-9 version. Data for items 14 and 15 of the PHQ-15 was omitted. Accordingly, after the disaggregated items were merged by taking the higher score out of both and the data of the additional items was omitted, the sum-score of all remaining items formed the reconstructed PHQ-9 score (Fried et al., 2022). For the group construction of our study, we used those reconstructed PHQ-9 depression severity scores.

Appendix B

Power Analysis

Epskamp et al. (2018) detail that for a partial correlations network, the edges to be expected for a given network can be calculated by the following formula: $P * (P - 1) / 2$. P indicates the number of nodes that are included in the network. For a network of 6 nodes, one can expect a total of 15 edges. From here, a power analysis can be calculated with the probable effect size (.2) and the number of participants needed per parameter ($N = 10$ per parameter) (Cohen, 1977). Since edges are described as parameters in network studies, for a network of 6 nodes, at least 150 participants are required per condition. Following the formula, a network of 7 nodes necessitates a network of 210 participants. Since the sample size per condition for this study equals 163, a network with a maximum of 6 nodes can be estimated.

Appendix C

Equalizing Group Sizes

The NCT works more concisely with equal group sizes (Steen et al., 2021; van Borkulo et al., 2023). We obtained equal group sizes by, first, randomly deleting participants from the groups *Improved* (n = 216) and *Same* (n = 367) until their sample sizes were equal to the sample size of group *Worse* (n = 163). Following the random removal of participants, we compared the mean and standard deviation of included PHQ-15 items for the *Improved* group before the removal of participants (with original n = 216) with group *Improved* after the removal of participants (n = 163). The comparison was statistically computed with the Mann-Whitney U test. This was being done to test the null hypothesis that groups do not systematically differ in scoring on the included PHQ-15 items when their sample size is decreased to n = 163. Only if no systematic differences are found can a group design with equal sample sizes be applied for statistical analyses (Kazdin, 2021). The same procedure of comparing original group data with group data after the removal of participants was repeated for group *Same*.

Table C1 shows the results for group *Improved,* whereas the results for group *Same* can be found in Table C2. The p-values of the Mann-Whitney U test supported the null hypothesis that the mean and standard deviation of the included PHQ-15 items do not change upon removal of participants for both groups, respectively. Therefore, we can conclude that the *Improved* and *Same* groups did not have systematically different sample characteristics with regard to the included PHQ-15 items following the random removal of participants. As a consequence, we decided to use a group design with equal sample sizes for the analyses of this study. This means that the *Improved* and *Same* group exhibit sample sizes of n = 163 each.

Table C1

Descriptive Statistics and Mann-Whitney Test Results of Group Improved

Items	Loss of Interest		Depressed Mood		Hopelessness		Fatigue		Worthlessness		Concentration Problems	
Sample Size Type	Equal	Unequal	Equal	Unequal	Equal	Unequal	Equal	Unequal	Equal	Unequal	Equal	Unequal
M	1.1	1.2	1.1	1.1	1.0	0.9	2.0	2.0	1.1	1.2	1.6	1.6
SD	0.8	0.8	0.8	0.8	0.9	0.9	0.9	0.8	0.9	0.9	1.0	1.0
Mann-Whitney (*W*)	17261.50		17740.00		17891.00		17193.50		17582.00		17382.00	
p-value	.723		.888		.770		.682		.983		.827	
n	163	216	163	216	163	216	163	216	163	216	163	216

Note. This Table displays the mean and standard deviation for group *Improved* after random deletion of participant (*n = 163)* and group *Improved* with the original sample size (*n* = 216). The mean (*M*) and standard deviation (*SD*) of the PHQ-items of group *Improved* (*n* = 163) and group *Improved* (*n* = 216) were compared via the Mann-Whitney U test (Mann & Whitney, 1947). The p-values of this test are also displayed in the Table.

Table C2

Descriptive Statistics and Mann-Whitney Test Results of Group Same

Items	Loss of Interest		Depressed Mood		Hopelessness		Fatigue		Worthlessness		Concentration Problems	
Sample Design	Equal	Unequal	Equal	Unequal	Equal	Unequal	Equal	Unequal	Equal	Unequal	Equal	Unequal
M	0.6	0.6	0.5	0.6	0.4	0.4	1.3	1.3	0.5	0.5	0.9	0.9
SD	0.6	0.7	0.7	0.7	0.6	0.6	0.8	0.8	0.7	0.7	0.9	0.9
Mann-Whitney (*W*)	29189.00		30692.50		29722.50		29027.00		29323.50		29074.00	
p-value	.620		.588		.892		.553		.681		.582	
n	163	367	163	367	163	367	163	367	163	367	163	367

Note. This Table displays the mean (*M*) and standard deviation (*SD*) for group *Same* after random deletion of participant (*n = 163)* and group *Same* with the original sample size (*n* = 367). The mean and standard deviation of the PHQ-items of group *Same* (*n* = 163) and group *Same* (*n* = 367) were compared via the Mann-Whitney U test (Mann & Whitney, 1947). The p-values of this test are also displayed in the Table.

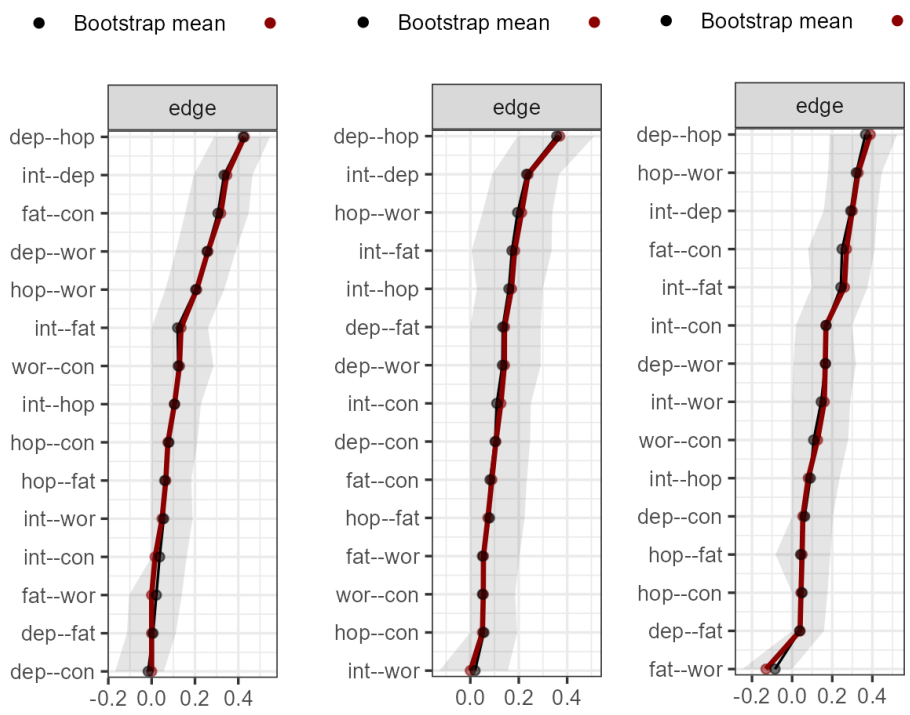
Appendix D

Edge-Weight Accuracy Test

We tested the accuracy of the estimated edge-weights by constructing a non-parametric bootstrapped confidence interval (CI) for every edge (see Figure D1). Bootstrapping means that a sampling distribution is estimated by repeatedly sampling from the original data set. Further, non-parametric means that observations from the original data set are randomly picked. After each selection, the randomly chosen observation stays in the data set. Accordingly, in the next random selection, there is a chance that the same observation is re-selected. This is repeated until the bootstrapped sample has reached the same size as the original data set from which the observations are drawn. For the construction of the confidence intervals in Figure D1, 1000 samples are drawn upon which the CI are estimated. The displayed CI are sufficiently small in width for each group, which indicates accurate edge-weights. As a consequence, the estimated networks of all groups can be interpreted.

Figure D1

Network Stability Confidence Intervals



Note. The graph on the left depicts the *Improved* group, the graph in the middle displays the *Worse* group, and the graph on the right entails the edge accuracy information for the *Same* group. The black lines indicate the original edge-weights based on the balanced data. The red dots are the mean bootstrapped edge-weights. The grey area represents the confidence interval (95% range). The smaller the CI, the more accurately the edge can be estimated (Santos et al., 2017).

Appendix E

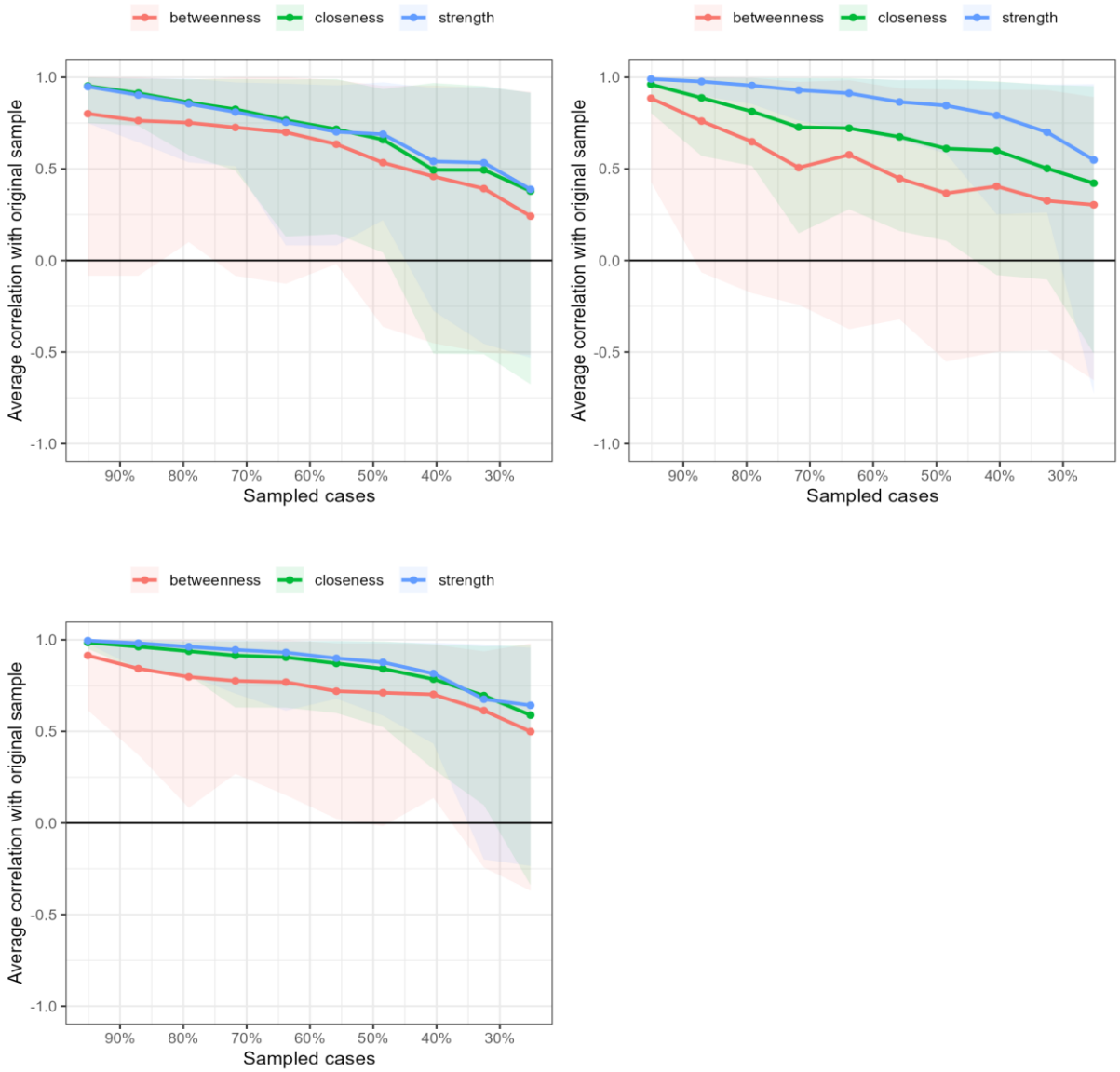
Centrality Stability Tests

In this section, the stability of the centrality indices is displayed. Stability indicates the degree to which the centrality indices are consistent in their result when data of individuals of the group is randomly removed (Epskamp et al., 2018). If the centrality indices are unstable when data is removed, it indicates that the results are not reliable and should therefore, not be interpreted (Epskamp et al., 2018).

As can be seen in Figure E1, the mean correlation of original data and bootstrapped data is less stable within groups *Improved* (upper right graph within Figure E1) and *Same* (upper left graph within Figure E1) compared to group *Worse* (lower left graph in Figure E1). Accordingly, the 95% confidence intervals of Improved and Same include the possibility of no correlation between original and bootstrapped data after 50%-65% of the data of participants was removed. This suggests a high instability of centrality indices. For the group *Worse,* the confidence intervals and the mean correlation indicate higher stability. That is specifically true for the centrality index *strength* of the *Worse* group. Furthermore, within group *Worse,* the less stable centrality indices *closeness* and *betweenness* showed a high correlation with *strength* (r = .96 with closeness and r = .81 with betweenness). This indicates that a change in indices *betweenness* or *closeness* is correlated with a change in the index *strength.* As a consequence, because *closeness* and *betweenness* are highly correlated with *strength* in the group *Worse*, and because *strength* is the most stable index, only *strength* is reported and interpreted in this thesis. Ultimately, due to the instability of all indices of groups *Improved* and *Same*, their centralities cannot be interpreted. Accordingly, their results are not reported.

Figure E1

Centrality Stabilities of Groups Improved, Worse, and Same



Note. The correlation of case-dropped bootstrapping data with original data is shown. The lines represent the means, and the lightly coloured areas indicate the confidence interval (2.5th to the 97.5th quantile). The upper left graph represents the *Same* group. The right graph represents the I*mproved* group, and the lower left graph represents the *Worse* group.

Appendix F

Statistical Analyses With Data Based on Original Group Sizes

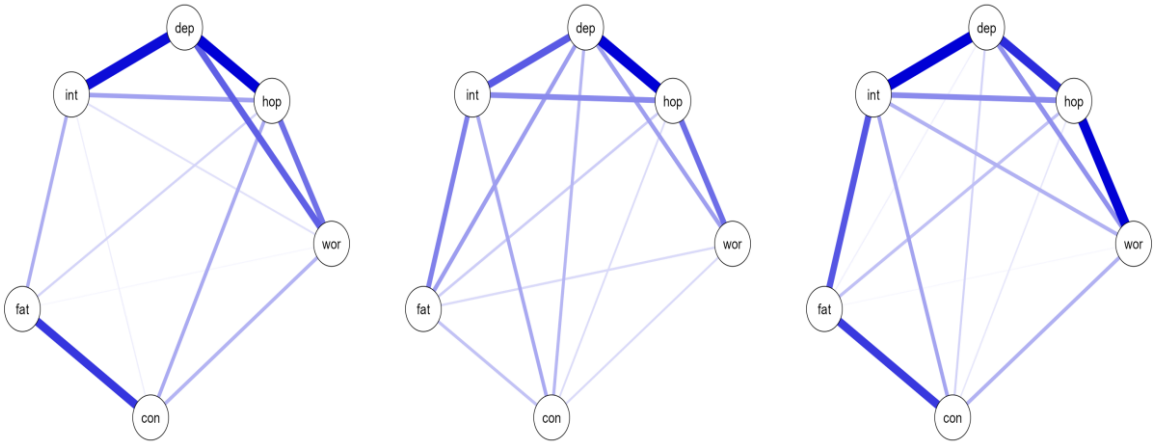
Network Description

Figure F1 displays the group networks based on the original group data. This means that the sample sizes were $n = 216$ for group *Improved*, $n = 367$ for group *Same*, and $n = 163$ for group *Worse*. The *Same* network was the most dense network (*degree of sparsity:* 0.00) followed by the *Worse* network (*degree of sparsity:* 0.07) and the *Same* network (*degree of sparsity:* 0.13).

The non-parametric bootstrapped 95% confidence intervals (CI) are displayed in Figure F2. As for the analyses in the main text based on equal group sizes, the bootstrapped samples were repeated 1000. The CI in Figure F2 were all sufficiently small in width, which means that the edges were accurate and, therefore, interpretable (Epskamp et al., 2018).

Figure F1

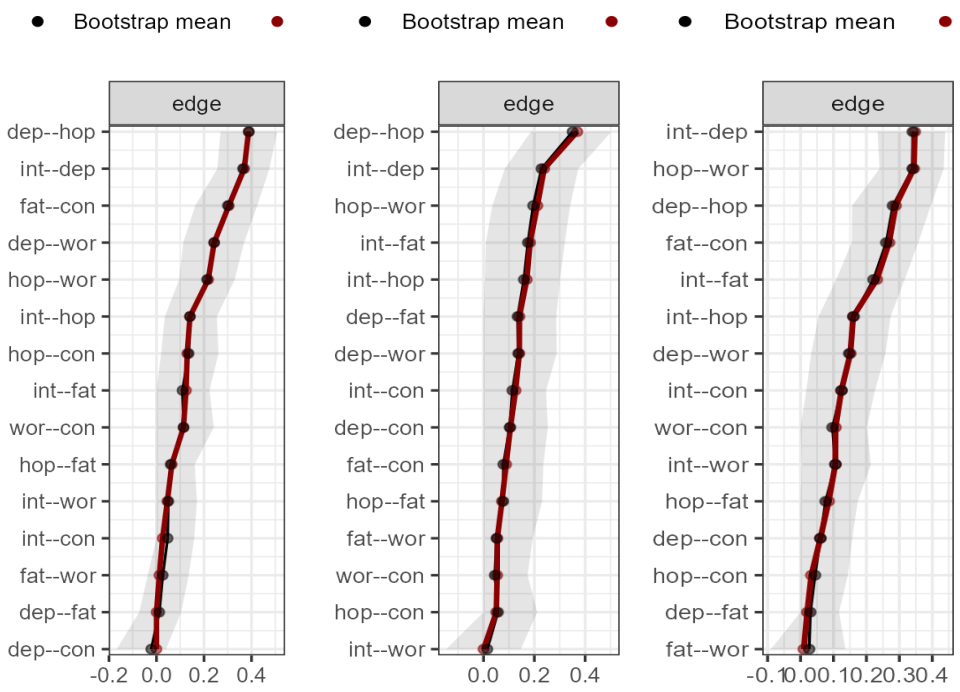
Visualized Networks of Groups Improved, Worse, and Same



Note. Network structures of the I*mproved* (left), *Worse* (middle), and *Same* (right) groups based on baseline data. Node names are shortened in order to fit in the figure: int = loss of interest; dep = depressed mood; hop = hopelessness; fat = fatigue; wor = worthlessness; con = concentration problems. Edges are visualized as lines that connect the lines. The colour blue indicates a positive association between nodes. The networks were visualized via the Gaussian Graphical Model (Foygel & Drton, 2010).

Figure F2

Network Stability Confidence Intervals



Note. The graph on the left depicts the *Improved* group, the graph in the middle displays the *Worse* group, and the graph on the right entails the edge accuracy information for the *Same* group. The black lines indicate the original edge-weights based on the unbalanced data. The red dots are the mean bootstrapped edge-weights. The grey area represents the confidence interval (95% range). The smaller the CI, the more accurately the edge was estimated (Santos et al., 2017).

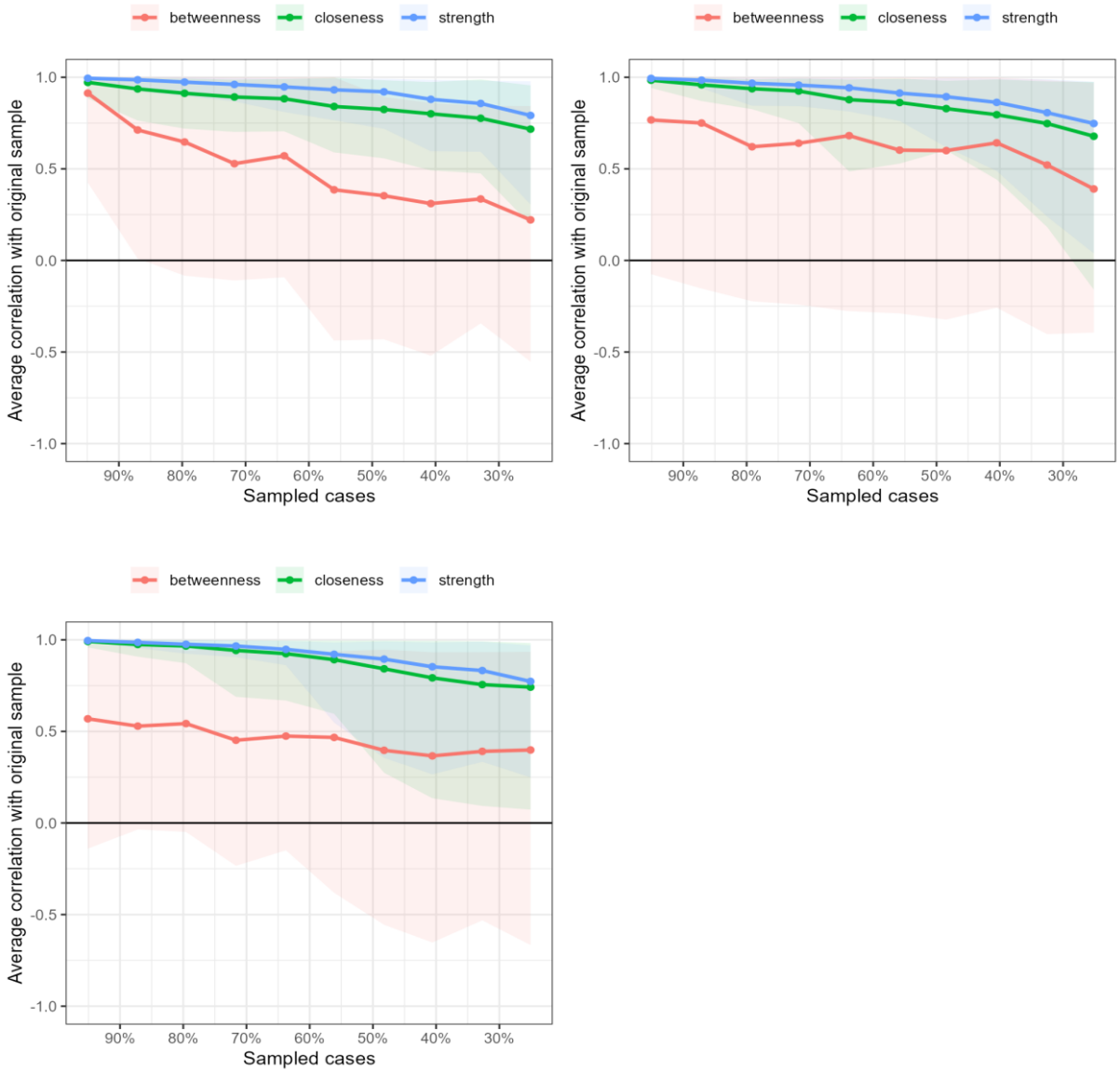
Centrality Indices

Concerning the centrality stability, the centrality index *betweenness* was not stable and therefore, not interpretable for every group (see Figure F3). *Strength* was the most stable parameter across all groups. And since *strength* and *closeness* showed a strong correlation in every group (*Improved:* r =.96; *Worse:* r = .96; *Same:* r = .87), we exclusively interpret the centrality *strength* for the analysis based on the original, unequal group sample sizes.

Regarding the results of the centralities, the node *depressed mood* exhibited the highest edge strength in groups *Improved* (1.36) and *Worse* (1.40), while having been the third highest node in the *Same* group (0.56). Moreover, the node *loss of interest* showed a higher value in the *Same* group (1.19) compared to *Improved* (-0.10) and Worse (0.21) (see Table F1).

Figure F3

Centrality Stabilities of Groups Improved, Worse, and Same



Note. Correlation of case-dropped bootstrapping data with original data is shown. The lines represent the means, the lightly coloured areas indicate the confidence interval (2.5th to the 97.5th quantile). Upper left graph is representing the *Improved* group. Right graph represents the *Worse* group and lower left graph is representing the *Same* group.

Table F1

Strength Centrality Values of Groups Improved, Worse, and Same

	Group *Improved*	Group *Worse*	Group *Same*
Centrality Index	*Strength*	*Strength*	*Strength*
Loss of Interest	-0.10	0.21	1.19
Depressed Mood	1.36	1.40	0.56
Hopelessness	1.07	0.88	0.83
Fatigue	-1.10	-0.55	-1.02
Worthlessness	-0.47	-0.90	-0.39
Concentration Problems	-0.76	-1.05	-1.17

Note. This Table shows the values for the centrality index s*trength* for every node of the respective group networks.

Network Comparison

The Network Comparison Test consists of two separate tests, which assess the network structure (Network Invariance Test) and the overall network connectivity (Global Strength Test). The Global Strength Test did not find significant results for the comparisons of networks *Improved/ Worse* (p = .158) and *Improved/ Same* (p = .157), which shows that the compared networks were equally strong regarding their overall connectivity. However, the network comparison of *Worse/ Same* (p = .050) did show a significant difference. However, following a Bonferroni correction for the number of performed significance tests (N = 6), the *Worse/ Same* network comparison did not remain significant. Moreover, the trend toward significance can be explained by the power imbalance between both groups (Steen et al., 2021) – the *Same* group (n = 367) had nearly twice as many participants as the *Worse* group (n

= 163). Accordingly, the *Same* network had substantially higher power than the *Worse* network. Therefore, the likelihood of a false positive was very high (Steen et al., 2021). It can be concluded that all networks were equally strong in terms of overall connectivity. Lastly, the Network Invariance Test did not show any significant differences between groups (Network Invariance Tests: *Improved/ Worse: p* = .267; *Improved/ Same: p* = .861; *Worse/ Same: p* = .436). This shows that the structure of the networks was not different across the groups.