Psychologie
Faculteit der Sociale Wetenschappen

Moderation of symptom-symptom associations in network models of psychopathology data: a primer and tutorial

Nikki van Eijk

MSc in Psychology (research), specialisation Clinical and Health Psychology

Institute of Psychology

Faculty of Social and Behavioural Sciences – Leiden University

Date: November 2020

Student number: s1655469

Internal supervisor at Leiden University: Eiko Fried

Contact: nikkilyka@hotmail.com

Abstract

The network approach has gained considerable attention over the past decade in psychological research. Ideas about the complex interplay between symptoms have been studied in a variety of different mental disorders. Different statistical methods have been introduced to study these networks of symptoms with wide ranging complexity. While recently more attention is focused on other variables that could play a role in these networks, there has not been much research into potential moderation effects within these networks. Because of the large variety in the presentation of mental disorders, moderators grant access to more complex models of psychopathology. This paper aims to present a novel statistical method that can be used to assess interaction effects on specific connections in the network. This method is then illustrated using a clinical sample with 81 participants. We do so by creating a network model based on experience sampling surveys and subsequently conducting moderation analyses on the connections in this model. Our model is based on a sample of healthy controls and patients with anxiety and mood disorders. Moderators include various questionnaires aimed at their psychopathology and personality traits. Extending the use of psychological networks to include moderation can help to gain insight into the complex processes by which mental disorders arise and perpetuate within patients. In addition to scientifically useful information, studying the interactions between networks of symptoms and other variables such as personality traits may also lead to valuable clinical information, guiding therapeutic interventions.

Introduction

Theoretical background

In the past decade, the network approach to psychopathology has gained considerable attention (Contreras, Nieto, Valiente, Espinosa, & Vazquez, 2019; Fried et al., 2017). This approach states that mental disorders arise from complex interactions among variables. This means that these variables do not need to share one underlying cause, but may be caused by each other or other variables (Borsboom & Cramer, 2013). Symptoms of psychopathology can evolve and perpetuate because of the dynamic structure of the network. For example, if a person experiences insomnia, that may cause feelings of fatigue at a later timepoint (Ferentinos et al., 2009). This approach attempts to explain why mental illness persists despite the absence of a certain trigger, by allowing the symptoms to directly cause other symptoms at later timepoints. The other view that is held by many is sometimes referred to as the common cause model. This model refers to the idea common in large areas of the medical world, where one illness causes many symptoms. These symptoms have no other connection to each other than the fact they are all caused by the same illness, which means that symptoms of a disorder would be uncorrelated after taking their common cause (the disorder) into account (Borsboom & Cramer, 2013; McNally, 2016). When taking this view, eliminating the illness would automatically lead to a disappearance of all symptoms, and interventions need to focus on the underlying illness, rather than on certain symptoms.

From experience in clinical practice, scientists realised that various symptoms of psychopathology were likely correlated on a larger scale than expected based on their shared cause, which was one of the reasons that led to the development of the network approach (Borsboom, 2017). After the network perspective was introduced, many studies have been published using this approach in many fields of psychology, including psychopathology (Contreras et al., 2019). This network perspective has been used in a variety of different areas

of psychopathology as well: depression (Bolger, Davis, & Rafaeli, 2003; Csikszenthmihalyi & Larson, 1987; Scott, Wozencroft, & Waller, 2019), anxiety (Beard et al., 2016; Langer et al., 2019), posttraumatic stress disorder (PTSD; Armour, Fried, Deserno, Tsai, & Pietrzak, 2017; McNally et al., 2015), binge eating disorder (BED; Wang, Jones, Dreier, Elliott, & Grilo, 2018), and many others.

Since the introduction of the network approach, statistical methods have been introduced to investigate the network theory. These network analysis methods likely simplify the complexity of true psychopathology networks by estimating linear relations between nodes. However, recently scientists have started asking more complex questions. One of the ways to answer these questions has been to include other variables such as cognitive functions and personality traits as well, to get a more complete overview, rather than just symptoms of certain mental illnesses (Fried & Cramer, 2017). Reasons behind this include the idea that symptom-symptom associations might be dependent on certain demographic variables.

Another trend that has emerged is an increase in studies taking a temporal approach to the network model, rather than cross-sectional studies based on one time point (e.g. Aalbers, McNally, Heeren, de Wit, & Fried, 2018; Boschloo, Schoevers, van Borkulo, Borsboom, & Oldehinkel, 2016; van Borkulo et al., 2015). In the past, cross-sectional research has been used to estimate networks. These studies are less time-consuming because they simply look at between-subjects effects. However, there have been many criticisms on the use of cross-sectional research for estimating psychological networks. For example, it has been shown that these networks are not well-suited to processes happening within one participant over time (Bos et al., 2017). Cross-sectional research in general is limited to correlational research, as different time points are needed to measure causality (Robinaugh, Hoekstra, Toner, & Borsboom, 2020). This makes the predictability of nodes in cross-sectional research less stable than in time-series data (Haslbeck & Waldorp, 2018). Replicability itself has also been

shown to be relatively low for cross-sectional research, often exacerbated by small sample sizes (Robinaugh et al., 2020). Thus, to investigate within-subjects processes, researchers have started using temporal data. These types of studies are based on time-series data, which means that they are longitudinal in nature and measure the same variables at different moments in time. This makes them different from the studies based on cross-sectional data because the difference is based on time rather than between-subjects variables. With this approach, predictions can be made about the potential causal order of variables and the way different variables could influence each other. An advantage of this is to be able to test hypotheses to get the biggest effect on the entire network structure, which can then be tested in real clinical settings. With technological advancements in data collection, it has become possible to collect time-series data quickly and easily. One of these methods is the Ecological Momentary Assessment (EMA), a form of Experience Sampling Methods (ESM), where data is collected repeatedly over time (Bolger et al., 2003; Csikszenthmihalyi & Larson, 1987; Scott et al., 2019). Using this method creates an opportunity for longitudinal data to be acquired within a fairly short amount of time. This allows for a closer look at the potential order in which variables influence each other.

Time-series data have been estimated with multilevel networks to gain new insights on the group level. For example, one study by Bringmann and colleagues (2013) investigated the effect of being in a therapy group on the structure of a network of mood-related factors. This meant that they combined time-series data with between-subjects factors to explore the effect of therapy on network structure. They also looked at variability in individual networks by investigating centrality measures at different values of neuroticism (Bringmann et al., 2013). This multilevel approach to networks allows for many new ways to study psychopathology, both at the group level and on the level of the individual.

However, many of these multilevel analyses which combine time-series with between-subjects variables, or moderators, have only investigated global measurements such as centrality. A major disadvantage of this approach is that you only find differences between people depending on their value on the moderator based on the entire network structure as a whole. An example of this would be a difference in the network structure between men and women. This comes with a second issue because analysing groups separately means a smaller sample size per analysis, which decreases the power of your estimates. There might be more intricate interactions between symptom-symptom associations and explanatory variables, which would be missed with this approach. We know that moderators influence many associations (Kazdin, 2017), but this knowledge has not been extended to the network approach in great detail, mainly because of a lack of statistical methods to implement this. One study has applied local structural equation models to assess moderation effects in comorbidity networks (Madole, Rhemtulla, Grotzinger, Tucker-Drob, & Harden, 2019). However, this approach was fairly complex, did not rely on free open-source software that can easily be used and obtained, and we are not aware of any simpler solutions that have been proposed. Hence, it will be the purpose of this paper to study the moderation of symptom-symptom associations in dynamic network analyses using a simpler methodology.

Goals of the current study

The current paper aims to present a statistical method suitable for estimating moderation effects within temporal models of psychopathology. Researchers are increasingly using time-series models in psychopathology research (Robinaugh et al., 2020), but there has been insufficient attention to the potential effect of moderators, especially on the scale of symptom-symptom associations. In this paper, the presented statistical method will be exemplified in an EMA dataset of individuals presenting with generalised anxiety disorder (GAD) and/or major depressive disorder (MDD; Fisher, Reeves, Lawyer, Medaglia, & Rubel,

2017). Specific details on the included variables and measures can be found in the method section of this paper.

The first goal of this paper is to put forward a simple statistical method that may improve how well psychopathology research reflects real life by including moderators into dynamic network models. Specifically studying temporal models allows for investigating moderation effects on within-subject processes. The relevance of this paper is based on gaining clinically and scientifically relevant information on the psychological processes involved in mental disorders, so future research will be able to give a more complete overview of the complex interactions between symptoms and other variables in psychopathology research. For example, the strong link between neuroticism and mental disorders has been found in many different studies (for a meta-analysis, see Kotov, Gamez, Schmidt, & Watson, 2010a), so it might prove fruitful to study if neuroticism has specific effects directly on symptoms or symptom interactions.

The second goal is to showcase this method in an empirical example with clinical data. Individuals with anxiety and/or depressive symptoms were included in this study to illustrate the potential use in psychopathology research. Healthy controls were included in the final analyses for a broader spectrum of responses and increased power. With the proposed method, a network analysis will be conducted for each subject. Then, the values of the edges between the symptom nodes will be extracted and correlated with a moderator. For example, the association between feeling anxious and experiencing fatigue may be moderated by a person’s age.

Methods

Data

The data were taken from a project by Fisher and Boswell (2016). Time-series data were collected from 47 individuals with GAD, MDD, or comorbid GAD and MDD, as well as from 35 healthy participants in a control group. DSM diagnoses were assessed using structured clinical interviews after patients passed a short screening survey. Other comorbid disorders were also assessed in the patient group, though manic and psychotic symptoms functioned as an exclusion criterion. Participants enrolled for personalised psychotherapy, and the EMA assessment took place in the 30 days before starting this study, to assess their symptoms before therapy. The participants were asked to answer 21 questions about their mood and behaviour four times a day, for approximately 30 days. These questions measured the DSM 5 symptoms for GAD and MDD, as well as related measures pertaining to positive affect, negative affect, rumination, behavioural avoidance, and reassurance seeking. The researchers used web-based surveys for the experience sampling. Texts were sent to the participants with a personalised hyperlink to the survey. Before the experience sampling survey, two rating scales were administered for measuring anxiety and depression symptomatology as baseline measures, as well as demographic information. Several other questionnaires were also administered. A full description of the data set is described in the original articles (A. J. Fisher & Boswell, 2016; A. J. Fisher et al., 2017).

Measures

From the vast battery of tests, we only used four measures in this study. Because of the two principal diagnoses in our group of participants, we decided to leave out questionnaires measuring other psychopathologies, such as personality disorders and post-traumatic stress disorder (PTSD). We decided to choose the Depression Anxiety Stress Scale because of the

added stress component that is missing from many other depression and anxiety scales, in order to measure a more general sense of distress (Lee, Lee, & Moon, 2019). To focus in more detail on anxiety symptoms and their interference with daily life, the Generalized Anxiety Disorder Questionnaire was used. Since the use of certain coping strategies correlates with psychopathology (Schäfer, Naumann, Holmes, Tuschen-Caffier, & Samson, 2017), we decided to use the Emotion Regulation Questionnaire. Lastly, we decided to use the NEO five-factor personality inventory because of the strong links between personality traits and psychopathology (Malouff, Thorsteinsson, & Schutte, 2005).

Depression Anxiety and Stress Scale. The Depression Anxiety and Stress Scale (DASS) is commonly used to measure negative emotions (Lovibond & Lovibond, 1995). The DASS consists of 42 questions, with 14 relating to depression, 14 relating to anxiety, and 14 relating to stress. The items consist of certain statements ("I felt sad and depressed") and the respondent is asked how much they agree with the statement on a Likert scale from 0-3. Scores within the subscale are then added together to form a range between 0-42 for each subscale. In order to minimise the number of moderators, the three subscales of the DASS were collapsed together to represent a general sense of distress and negative emotions.

Generalized Anxiety Disorder Questionnaire – IV. The Generalized Anxiety Disorder Questionnaire-IV (GAD-Q-IV) is a self-report questionnaire for the entire clinical syndrome of generalised anxiety (Newman et al., 2002). Originally created for the criteria of the DSM-III-R, it has since been revised to match the criteria posed by the DSM-IV-TR and the DSM 5. The scale has nine questions, of which five are yes/no questions. One question asks the participant to list the topics they worry about, one question asks participants to check all physical symptoms they experience, and two questions follow a 0-8 scale intended to measure the amount of distress and interference with daily life.

Emotion Regulation Questionnaire. The Emotion Regulation Questionnaire (ERQ) is designed to measure two different emotion regulation strategies: cognitive reappraisal and expressive suppression (Gross & John, 2003). The scale distinguishes between emotional experience and emotional expression. The subscale cognitive reappraisal consists of six statements ("I control my emotions by changing the way I think about the situation I'm in"); the expressive suppression scale consists of four statements ("I keep my emotions to myself"). Scores are based on 7-point Likert scales. The final scoring takes the average scores in each subscale, with higher scores indicating a higher use of that strategy. To focus on fewer moderators, we took a closer look at the ERQ and its predictive validity. Since the subscales represent such different methods of coping with (negative) emotions, we were not able to combine them into one as we did with the DASS. Thus, eventually, we decided to pick the reappraisal subscale since a recent study found that reappraisal was related to anxiety outcomes, whereas suppression was not (Goldin et al., 2014).

NEO Personality Inventory. The NEO Personality Inventory (NEO-PI) is a well-established inventory for assessing personality traits (Costa & McCrae, 2013). The NEO-PI assesses participants on five global personality domains. These can be used as five factors, or three of these domains can be further broken down into facets. For this study, only the neuroticism subscale was used due to its strong links with depression (Kotov, Gamez, Schmidt, & Watson, 2010b).

Analyses

All analyses will be performed in RStudio version 1.1.456 (R version 3.5.1). Missing data will be handled in the following manner: if less than 5% of data is missing, single predictive-mean-matching-based imputation will be used on the remaining cases using the MICE package (van Buuren & Groothuis-Oudshoorn, 2011). This is because single

imputation is considered to be appropriate when less than 5% is missing, and substantially less complicated than multiple imputation methods (Graham, 2009). Above 5% listwise deletion will be used to handle missing data.

Network estimation and visualisation. Networks consist of two elements: nodes and edges. Nodes represent symptoms or other variables included in the network, while edges represent the connections between these variables. Edges can be computed in various ways, though they often represent some type of correlation. In the networks we will be creating, the edges represent the partial correlations between two variables. The network structure will be estimated using the GraphicalVar package in the R software environment (Epskamp, 2018). A function called MLGraphicalVar will be used to estimate both individual networks as well as networks for the entire sample. Thicker edges represent stronger connections. Blue edges will be used for positive associations, whereas red edges will be used for negative associations. One network will be estimated for the entire sample including the healthy controls, mainly to increase the power of the model.

Model assumptions will be checked before interpreting the final model. The first assumption relates to the difference between day and night. The last observation from one day will not be correlated with the first observation from the second day, because the interval between these measurements is larger than between other measurements. Hence, it is unlikely to assume these associations to be similar to those during the day.

The next assumption is stationarity. This means that the mean and variance of the series must stay the same over time (Box, Jenkins, Reinsel, & Ljung, 2015). To check this, all variables are regressed with time as the independent variable. If a variable shows a time effect, it will be detrended.

For sparsity, only six nodes were selected to create the final estimation from the wide selection of variables measured during the EMA period. Due to the sample's mix of anxiety and depression patients, we sought to find a balanced mix between nodes. In the end, two variables primarily representing depression were picked: rumination and feelings of hopelessness; two variables primarily representing anxiety were picked: feeling worried and procrastination; and two variables with overarching symptoms were picked: feeling positive and fatigue.

Network moderation. First, the values of all edges were extracted from the networks of each individual participant. The MLGraphicalVar function within the GraphicalVar package creates all subject-specific networks so a loop can be used to extract the partial correlation matrices from each participant. From these individual networks, all specific edges between variables were extracted. Then, for the five largest symptom-symptom edges on the group level, the individual edges were correlated with the four specific moderators measured at baseline to test for moderation effects. Pearson correlations were used to assess the correlations between edges and moderators. To adjust for possible false positives, the false discovery rate (fdr) method was used to correct for multiple analyses. The fdr method was used within each of the four analyses, but no correction was used across the different moderation analyses, given the exploratory nature of this study.

Results

Descriptives

The sample used for our analyses consisted of 82 participants. Most of these participants were White (42.7%). Of the non-White participants, the largest group of participants identified as Asian/Asian American (24.4%) or Latino (13.4%). The rest of the

participants identified as Black (4.9%), other (6.1%), or a combination of races (8.5%). The ages of the participants ranged from 18 to 98, with a mean age of 32.4 years (SD=15.2 years). 15 participants were under 20 (18.3%), 32 participants fell into the 20-29 age group (40.2%), 13 participants were between 30-39 years of age (15.9%), 6 fell into the 40-49 age group (7.3%), 11 fell into the 50-59 age group (13.4%), and 4 participants were over 60 (4.9%). Of the 82 participants, 32 had some college education (39.0%) and 29 had a college degree (35.4%). Some participants had a master's degree (17.1%) or a doctorate degree (3.7%). Only 4.9% of participants had a high school diploma or less. Based on American education attainment data, this is fairly representative of the general population (United States Census Bureau, 2019). With regard to their mental health, most patients had not previously received outpatient care (64.6%). During the time of the study, 16 participants were using medication for their mental health problems (19.5%).

Network estimation

Figure 1 presents the contemporaneous network for the entire sample, on which we base our moderation analysis. The nodes represent the psychopathology symptoms and the edges represent the unique links between them, quantified with partial correlations. Blue edges represent positive relationships, red edges represent negative relationships. The thickness of the line indicates how strong the relationship is.

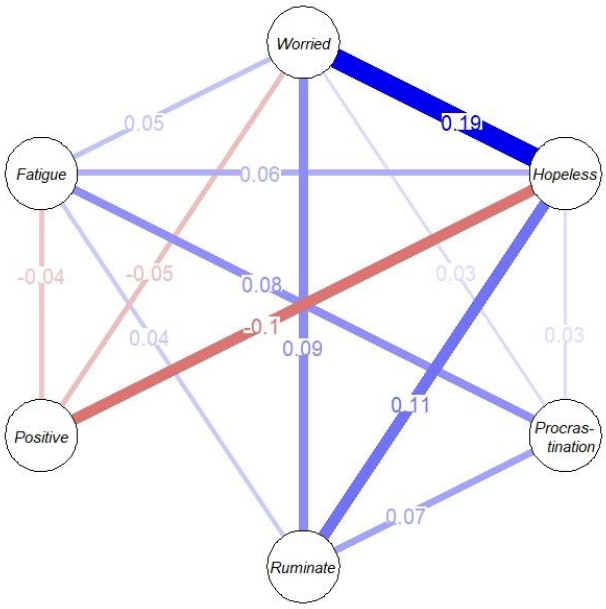

Figure 1: Group network of anxiety and depression symptoms. Edge weights are shown in their respective colour. Blue edges = positive partial correlation. Red edges = negative partial correlation.

Out of the 15 possible connections (6*5/2) between nodes, only the link between feeling positive and ruminating was not present. This showed clear interconnection between the symptoms and related variables. The five strongest edges were extracted from the model to make the moderation analysis more interpretable and minimise the probability of false positives. On top of that, stronger edges tend to mean fewer edge weights were fixed to zero on the individual level, which increases the chances of actually finding an effect. These were *worried – hopelessness*, *hopelessness – rumination*, *hopelessness – positivity*, *worried – rumination*, and *procrastination – fatigue*. Several of these edges linked the anxiety and depression symptoms, showing a clear overlap between the two constructs. Hopelessness generally presented as one of the variables with the strongest connections to all other variables.

Moderation analysis

The full set of moderation analyses are presented in Figure 2, which depicts the relationship between the measures used as moderators and the five strongest edge weights. 95% confidence intervals are shown in the figure, as well as in the text below.

Depression Anxiety Stress scale. The correlation between the DASS and the edge linking worried and hopelessness was small, $r = -.17,\ 95\%\ CI\ [-.37; .05]$. This indicated that people with a higher DASS score had on average a weaker link between worried and hopeless. Since this link was positive, this means that people who are inclined to worry more when they feel hopeless and vice versa, are more likely to have a lower DASS score. The same pattern was visible for worried and ruminating, $r = -.17,\ 95\%\ CI\ [-.38; .05]$. On the other hand, the correlation between the DASS and the link between hopeless and positive was positive, $r = .24,\ 95\%\ CI\ [.03; .44]$. The negative link between feeling hopeless and feeling positive indicates that people who feel hopeless are not likely to feel positive at the same time. This link was weaker for people with a higher DASS score because the coding of this link is reversed, meaning that people with more negative emotions are less likely to have a strong link between feeling hopeless and feeling positive. The correlation between the DASS and the edge linking hopelessness to rumination was negligible, $r = .01,\ 95\%\ CI\ [-.21; .23]$, as was the correlation with the link between procrastination and fatigue, $r = .00$, $95\%\ CI\ [-.22; .22]$.

Generalized Anxiety Disorder Questionnaire – IV. For the GAD-Q-IV, the total score was used as a moderator. The pattern was fairly similar to the one described for the DASS. The correlation with the edge between worried and hopeless was small to medium, $r = -.22,\ 95\%\ CI\ [-.41; .00]$, indicating that a lower anxiety score was related to a stronger link between feeling worried and feeling hopeless. Once again, the relationship with worried

and rumination was similar, $r = -.18,\ 95\%\ CI\ [-.39; .04]$. The GAD-Q-IV related positively to the relationship between hopelessness and positivity, $r = .24, 95\%\ CI\ [.03; .43]$. The correlations between the GAD-Q-IV and the link between hopeless and ruminate, as well as the link between procrastination and fatigue, were nearly zero, $r = .00$, $95\%\ CI\ [-.22; .21]$ and $r = -.03,\ 95\%\ CI\ [-.25; .19]$ respectively.

Emotion Regulation Questionnaire. The only correlation that stood out was the relationship between the ERQ and the link between worried and ruminate, $r = .16$, $95\%\ CI\ [-.06; .37]$. This indicated that people who make more use of cognitive reappraisal were more likely to have a stronger link between feeling worried and ruminating. The correlation with the link between worried and hopeless was very small, $r = -.08$, $95\%\ CI\ [-.29; .14]$, as was the link between procrastination and fatigue, $r = .07$, $95\%\ CI\ [-.15; .28]$. The link between the ERQ reappraisal subscale and the edge linking hopeless and rumination was about zero, $r = .03,\ 95\%\ CI\ [-.19; .25]$, as was the correlation with hopeless and positive, $r = .00,\ 95\%\ CI\ [-.22; .22]$.

NEO Neuroticism scale. The neuroticism subscale on the NEO Personality Inventory was the only one which showed a larger – though still small – effect on the link between procrastination and fatigue compared to the other moderators, $r = -.13,\ 95\%\ CI\ [-.33; .10]$. This indicated that people who are more neurotic experience a smaller link between procrastination and fatigue. The same effect occurred for the link between worried and ruminate, $r = -.17,\ 95\%\ CI\ [-.37; .05]$, and between worried and hopeless, $r = -.19$, $95\%\ CI\ [-.40; .03]$. The correlation between the NEO and the edge between hopeless and positive was a small positive relationship, $r = .16,\ 95\%\ CI\ [-.06; .37]$, indicating that those with higher scores on neuroticism were less likely to have a strong link between feeling

hopeless and feeling positive. Lastly, the link between hopeless and ruminate correlated very weakly with the neuroticism scale, $r = .04,\ 95\%\ CI\ [-.18; .26]$.

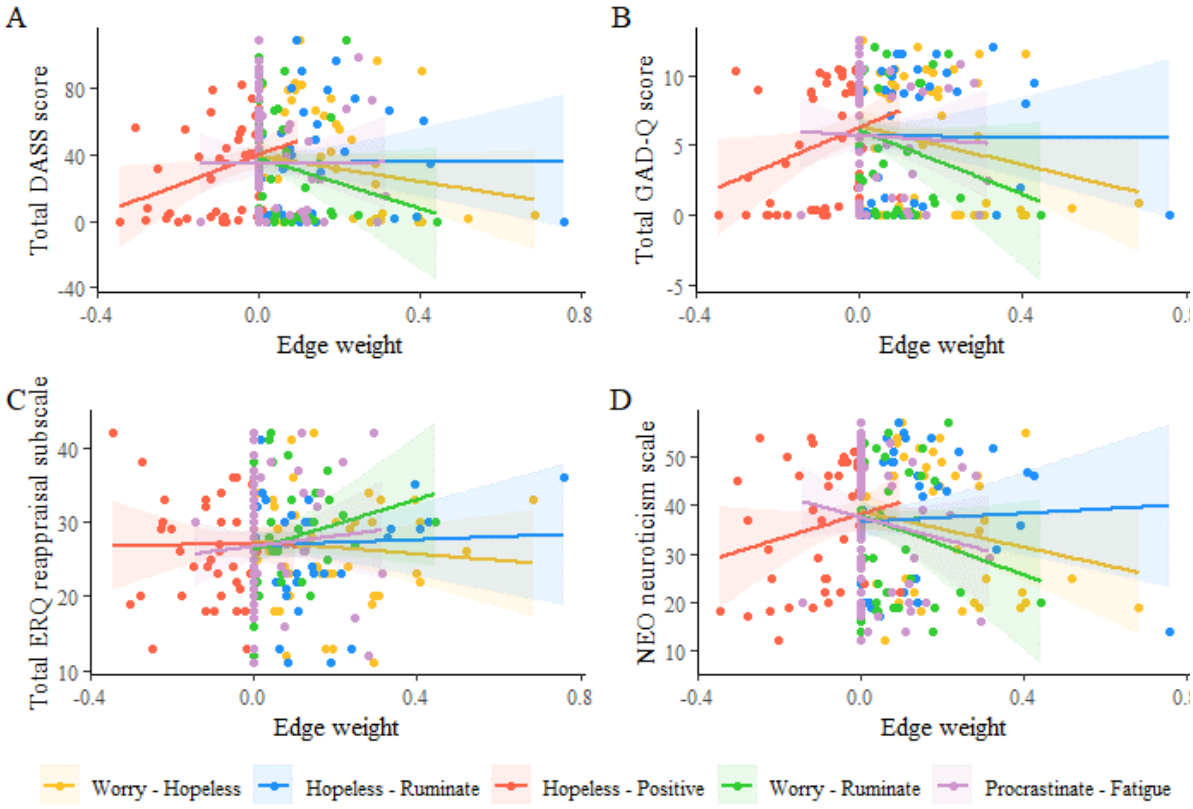

Figure 2: The individual edge weights for the five largest edges plotted against the total DASS score (A), the total GAD-Q score (B), the total ERQ reappraisal subscale (C), and the NEO neuroticism subscale (D). Confidence intervals are plotted at 95%.

Moderation per edge

Another way of describing the results is by looking at the edge weights themselves instead of the moderators. An overview of all the correlations can be found in Table 1. As seen in the results above, most of the confidence intervals included zero, indicating nonsignificant results. The edge weight *hopeless – positive* had the strongest correlations, namely .24 with the DASS sum scale, and .24 with the GAD-Q-IV. While these correlations were not significant when using the fdr correction for multiple comparisons, these were the

only correlations with raw confidence intervals fully above zero. Since this edge weight was the only negative one, interpretation is reversed as well. Positive correlations indicated a weaker link between the nodes with increasing values on the moderator. This was the case for three out of four correlations linking *hopeless – positive* to a moderator. Many of the other correlations were negative, indicating that lower values on the moderator were associated with larger edge weights. This was the case for all the correlations with the edge between worried and hopeless, three out of four correlations with the edge between worried and ruminate, and for two of the four correlations with the edge between procrastination and fatigue. As you can see in Figure 2 (panel C), the only correlation that truly stood out from this pattern is the correlation linking *worried – ruminate* to the ERQ. This indicated that those who were more inclined to worry and ruminate used the reappraisal strategy more often. The edge *hopeless – ruminate* did not seem to be moderated by any of the questionnaires, all r<.04. Since higher scores on all moderators except the ERQ indicate a higher level of psychosocial dysfunction, the overall conclusion seemed to be that those with more dysfunction have weaker links in their network.

Table 1.
Pearson Correlations between edge weights and moderators.

	DASS	GAD-Q-IV	ERQ-R	NEO-N
Worried – Hopeless	-.17 [-.37;.05]	-.22 [-.41;.00]	-.08 [-.29;.14]	-.19 [-.40;.03]
Hopeless – Ruminate	.01 [-.21;.23]	.00 [-.22;.21]	.03 [-.19;.25]	.04 [-.18;.26]
Hopeless – Positive	.24 [.03;.44]	.24 [.02;.43]	.00 [-.22;.22]	.16 [-.06;.37]
Worried – Ruminate	-.17 [-.38;.05]	-.18 [-.39;.04]	.16 [-.06;.37]	-.17 [-.37;.05]
Procrastination – Fatigue	.00 [-.22;.22]	-.03 [-.25;.19]	.07 [-.15;.28]	-.13 [-.33;.10]

Notes: 95% confidence intervals are shown.
Abbreviations: DASS: Depression Anxiety Stress Scale, GAD-Q: Generalized Anxiety Disorder Questionnaire, ERQ-R: Emotion Regulation Questionnaire – reappraisal subscale, NEO-N: NEO Personality Inventory – Neuroticism subscale

Discussion

The aim of this study was to showcase a new method that can be used to analyse possible moderators in temporal networks of psychopathology data. This is important because mental illnesses consist of complex psychological processes and most statistical models have

not reflected this complexity in the past. Extending network models to include moderators is another step forward in examining how these processes work, both at the individual level and the group level. While numerous ways exist to implement this method in much more involved ways, we aimed to showcase a straightforward method that is easy to follow. Especially because the potential predictive use of this method lies in the clinical field rather than the academic field, we thought it was relevant to keep it interpretable even to those without a solid mathematical background. Thus, we only took six symptoms of depression and anxiety to compute the final network, and from this network we only took the five strongest edges to perform the moderation analyses. We also limited ourselves to only study four possible moderators, to focus on the interpretability of the design.

A few findings are worth discussing in more detail. To start, when considering the results of our study, one thing clearly stands out. Most of the correlations with the moderators were not in the direction that were expected based on previous research. For example, one study showed that those who recovered from depression had weaker connectivity in their network (van Borkulo et al., 2015). If generalised, it would be expected that experiencing fewer symptoms would correspond with smaller edge weights, but the opposite was found in our results. There were no clear explanations for this, though we outline some hypotheses here.

The first possible reason for finding correlations in the opposite direction than expected was a floor effect on some of the moderators for the people in the healthy control group. While these people endorsed the symptoms measured and their edge weights may have been similar, the questionnaires show a clear discriminatory effect between healthy people and those with a mental disorder (J. Fisher, Thach, & Tuan, 2013; M. T. Moore, Anderson, Barnes, Haigh, & Fresco, 2014; Vasconcelos-Raposo, Fernandes, & Teixeira, 2013). This could have been part of the reason why negative correlations were observed between edge

weights and the moderators. Indeed, post-hoc data exploration showed that some of the signs of the correlation coefficients switch when only taking into account the patient data, although still without reaching significance. However, this would only explain the effects for the DASS and the GAD-Q-IV, but not for the ERQ and the NEO-PI, since the latter measure a continuum across both healthy and mentally ill people.

A second explanation could be that the network estimation fixed many edge weights to zero if they did not reach a certain threshold. This meant that the variance of the edge weights was reduced, but potential bias increased. However, the least restrictive model was used for this study, so any changes would have only further reduced the number of significant edges.

A last possible explanation may be that the analysis method chosen (Pearson correlations) was not well-suited to the data. With many edges fixed to zero, data became very skewed. Nonparametric measures may have been better suited to analyse these interaction effects. However, since the main goal of this study was to introduce a possible method for integrating moderators into network models, this would have been more complex than we considered desirable. Studies with clear hypotheses in mind about the integration of their moderators might benefit from looking into nonparametric testing for skewed data, or using transformations to meet the assumptions for most statistical tests. On top of that, the variables chosen in a study influence the distribution of the scores, and many potential variables may not need additional transformations or robust testing to meet the necessary assumptions.

While this extension of network modelling allows for more complexity in estimating psychological models, there is still much to extend on. For example, one paper found that emotion regulation strategies mediated the role between neuroticism and depression (Yoon, Maltby, Joormann., & Yoon, 2013). We investigated emotion regulation and neuroticism separately in their relation to depression and anxiety symptoms, which might not as accurately

reflect the true relationships between these variables. Thus, future research could extend on this by examining higher-order interactions and looking at more complex combinations of edges and other predictors. However, we think that this is a good start in trying to improve the statistical models that we use to estimate the relationships between symptoms in a network model of mental disorders.

Another point of interest to mention here is that this method of analysing network models assumes stationarity. That means that we expect the connections between the nodes in the model to stay the same over time. However, it would be interesting to create networks where this assumption is not made and connections can vary over time. One study mentioned previously looked at the difference between patients who remitted and those who persisted (van Borkulo et al., 2015), but being able to estimate this difference in a longitudinal manner might lead to new insights. It would allow clinicians to track changes in the network as a person goes through treatment for example. This way, you can see which edges weaken first or where strong connections may lead to relapse. This can then help the clinical field to learn more about the underlying processes of their patients.

Limitations

What is important to keep in mind is that this study was exploratory in nature, and many other decisions could have been made. The original dataset by Fisher and Boswell (2016) featured 22 variables that could have been used to estimate the network, of which we used only six. To avoid cherry-picking, we picked the variables before any network was estimated. To do so, we looked at the distributions of how heavily an item was endorsed, as well as the underlying disorder the item was meant to represent. Another decision we made was to use the whole group for the estimation of the network. While many studies have focused on the differences between the control group and the clinical group, we decided to opt

for having more power due to a larger sample size, especially because we estimated the effects would be small. Moreover, mental health should not be seen as an absence of mental illness as though they are two sides on the same continuum (Keyes, 2002), and so it could be expected that people classified as healthy still endorse the measured symptoms, albeit to a lesser extent. After doing so, we performed another moderation analysis to see if this had been a justified decision. We compared all fifteen original edges across the two groups to see if there were any significant differences. None of the edges showed a significant difference at the 5% level, and only 2 edges showed a significant difference at the 10% level, namely the edge between worry and ruminate and the edge between hopeless and fatigue. Scatter plots for all edges can be found in Figure 1 of the Supplementary materials. Despite the combined sample, the sample size was still not sufficiently large to detect smaller effects. Larger samples will be necessary to detect effects across participants, especially if the variance in the measured moderators is large. However, the purpose of our study was not to find significant effects in the field of anxiety and depression research, but merely to show how our method could potentially be used in a clinically relevant sample.

The last decision we had to make was which moderators to use. Many justifiable decisions could have been made here, and we certainly do not claim ours is the best. Ideally, moderators should not be correlated if you put them in a model separately, but in psychopathology this tends to be very hard to achieve. We combined the three subscales of the DASS for two main reasons: first, the correlations between the subscales are significantly high that the overlap is considerable (Vasconcelos-Raposo et al., 2013); and second, we did not want the models to be overly complex and measure many different moderators, due to the increased likelihood of false positives. However, arguments have been made for the discriminatory power of the three subscales, so other decisions could have been made as well. Our decision to use the reappraisal subscale of the ERQ was based on literature stating that

reappraisal was significantly related to depression. However, another study has found contradicting evidence, stating that suppression was strongly related to depression and stress, whereas reappraisal was not (S. A. Moore, Zoellner, & Mollenholt, 2008). Thus, there is once again no one correct decision to be made. We hope that by using these variables as an example, we have been able to illustrate how a similar study could potentially be carried out.

Concluding, it is important to realise that we did not explore all of these different options and picked this one to present, but that only one possible outcome was explored. That leaves you free to criticise us for our choices and conduct a new study in which you make different choices along the way and see how the results compare.

With regards to the dataset itself, there were some limitations as well. The main one was already mentioned previously: the lack of power for what we wanted to study. Either a larger sample or more time points would be necessary to investigate small effects. On top of that, the sample does show some WEIRD characteristics. While the education level of the participants was fairly representative of the population of the United States, future studies should seek to include people from all types of backgrounds. Researchers from different countries and institutes should work together to validate the results across different cultures, as well as to increase the overall sample size for increased power.

Recommendations for future research

In conclusion, the current paper presented a method for combining time-series data with moderating variables. This method can be applied to many different models, even those outside of psychopathology. Any model that could benefit from the analysis of moderators could benefit from using this approach. We hope that our analysis of a clinically relevant sample has increased your understanding of how this can be applied in future research. While this paper only examined the contemporaneous network, this method can easily be extended

to the temporal network provided by many network estimation packages as well. More extensions on studying moderation could also be useful in moving forward, such as studying higher-order interactions within the model. An example already posed by Robinaugh and colleagues (2020), in which they suggest sleep might moderate the effect of trauma memories on emotional reactivity. Moreover, since many statistical tests are performed in these types of analyses, high power is needed to detect smaller effects. This calls for more consistent research so that data may be aggregated into larger samples across different research teams. Another use of this method could be to investigate specific factors in more idiographic research, rather than a nomothetic approach. This could potentially help patients in mental health care to improve faster by targeting specific aspects of their network. We believe that the extension of network analysis with moderation analyses is a valuable tool in the toolbox of psychological researchers, and very relevant for the current time in which increasing attention is being put on new methods of investigating psychopathology.

References

Aalbers, G., McNally, R. J., Heeren, A., de Wit, S., & Fried, E. I. (2018). Social Media and Depression Symptoms: A Network Perspective. *Journal of Experimental Psychology: General*, (2018), 1–9. https://doi.org/10.1037/xge0000528

Armour, C., Fried, E. I., Deserno, M. K., Tsai, J., & Pietrzak, R. H. (2017). A network analysis of DSM-5 posttraumatic stress disorder symptoms and correlates in U.S. military veterans. *Journal of Anxiety Disorders*, *45*(May 2013), 49–59. https://doi.org/10.1016/j.janxdis.2016.11.008

Beard, C., Millner, A. J., Forgeard, M. J. C., Fried, E. I., Hsu, K. J., Treadway, M. T., … Björgvinsson, T. (2016). Network analysis of depression and anxiety symptom relationships in a psychiatric sample. *Psychological Medicine*, *46*(16), 3359–3369. https://doi.org/10.1017/S0033291716002300

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary Methods: Capturing Life as it is Lived. *Annual Review of Psychology*, *54*(1), 579–616. https://doi.org/10.1146/annurev.psych.54.101601.145030

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, *16*(1), 5–13. https://doi.org/10.1002/wps.20375 LK - http://sfx.metabib.ch/sfx_locater?sid=EMBASE&issn=20515545&id=doi:10.1002%2Fwps.20375&atitle=A+network+theory+of+mental+disorders&stitle=World+Psychiatry&title=World+Psychiatry&volume=16&issue=1&spage=5&epage=13&aulast=Borsboom&aufirst=Denny&auinit=D.&aufull=Borsboom+D.&coden=&isbn=&pages=5-13&date=2017&auinit1=D&auinitm=

Borsboom, D., & Cramer, A. O. J. (2013). Network Analysis: An Integrative Approach to the Structure of Psychopathology. *Annual Review of Clinical Psychology*, *9*(1), 91–121. https://doi.org/10.1146/annurev-clinpsy-050212-185608

Bos, F. M., Snippe, E., de Vos, S., Hartmann, J. A., Simons, C. J. P., van der Krieke, L., … Wichers, M. (2017). Can We Jump from Cross-Sectional to Dynamic Interpretations of Networks? Implications for the Network Perspective in Psychiatry. *Psychotherapy and Psychosomatics*, *86*(3), 175–177. https://doi.org/10.1159/000453583

Boschloo, L., Schoevers, R. A., van Borkulo, C. D., Borsboom, D., & Oldehinkel, A. J. (2016). The network structure of psychopathology in a community sample of preadolescents. *Journal of Abnormal Psychology*, *125*(4), 599–606. https://doi.org/10.1037/abn0000150

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control*. John Wiley & Sons.

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., … Tuerlinckx, F. (2013). A Network Approach to Psychopathology: New Insights into Clinical Longitudinal Data. *PLoS ONE*, *8*(4). https://doi.org/10.1371/journal.pone.0060188

Contreras, A., Nieto, I., Valiente, C., Espinosa, R., & Vazquez, C. (2019). The study of psychopathology from the network analysis perspective: A systematic review. *Psychotherapy and Psychosomatics*, *88*(2), 71–83. https://doi.org/10.1159/000497425

Costa, P. T., & McCrae, R. R. (2013). Stability and change in personality assessment: The revised NEO personality inventory in the year 2000. *Emerging Issues and Methods in Personality Assessment*, pp. 87–95. https://doi.org/10.4324/9780203774618-12

Csikszenthmihalyi, M., & Larson, R. (1987). Validity and Reliability of the Experience-Sampling Method. *The Journal of Nervous and Mental Disease*, *175*(9), 526–536.

Epskamp, S. (2018). *Graphical VAR for Experience Sampling Data* (p. 12). p. 12. CRAN.

Ferentinos, P., Kontaxakis, V., Havaki-Kontaxaki, B., Paparrigopoulos, T., Dikeos, D., Ktonas, P., & Soldatos, C. (2009). Sleep disturbances in relation to fatigue in major depression. *Journal of Psychosomatic Research*, *66*(1), 37–42. https://doi.org/10.1016/j.jpsychores.2008.07.009

Fisher, A. J., & Boswell, J. F. (2016). Enhancing the Personalization of Psychotherapy With Dynamic Assessment and Modeling. *Assessment*, *23*(4), 496–506. https://doi.org/10.1177/1073191116638735

Fisher, A. J., Reeves, J. W., Lawyer, G., Medaglia, J. D., & Rubel, J. A. (2017). Exploring the idiographic dynamics of mood and anxiety via network analysis. *Journal of Abnormal Psychology*, *126*(8), 1044–1056. https://doi.org/10.1037/abn0000311

Fisher, J., Thach, D. T., & Tuan, T. (2013). Validation of the depression anxiety stress scales (DASS) 21 as a screening instrument for depression and anxiety in a rural ... Validation of the depression anxiety stress scales (DASS) 21 as a screening instrument for depression and anxiety in a rur. *BMC Psychiatry*, *13*(24), 1–7. https://doi.org/10.1186/1471-244X-13-24

Fried, E. I., & Cramer, A. O. J. (2017). Moving Forward: Challenges and Directions for Psychopathological Network Theory and Methodology. *Perspectives on Psychological Science*, *12*(6), 999–1020. https://doi.org/10.1177/1745691617705892

Fried, E. I., van Borkulo, C. D., Cramer, A. O. J., Boschloo, L., Schoevers, R. A., & Borsboom, D. (2017). Mental disorders as networks of problems: a review of recent insights. *Social Psychiatry and Psychiatric Epidemiology*, *52*(1), 1–10. https://doi.org/10.1007/s00127-016-1319-z

Goldin, P. R., Lee, I., Ziv, M., Jazaieri, H., Heimberg, R. G., & Gross, J. J. (2014). Trajectories of Change in Emotion Regulation and Social Anxiety During Cognitive-Behavioral Therapy for Social Anxiety Disorder. *Behaviour Research and Therapy*, *56*, 7–15. https://doi.org/10.1038/jid.2014.371

Graham, J. W. (2009). Missing Data Analysis: Making It Work in the Real World. *Annual Review of Psychology*, *60*(1), 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530

Gross, J. J., & John, O. P. (2003). Individual Differences in Two Emotion Regulation Processes: Implications for Affect, Relationships, and Well-Being. *Journal of Personality and Social Psychology*, *85*(2), 348–362. https://doi.org/10.1037/0022-3514.85.2.348

Haslbeck, J. M. B., & Waldorp, L. J. (2018). How well do network models predict observations? On the importance of predictability in network models. *Behavior Research Methods*, *50*(2), 853–861. https://doi.org/10.3758/s13428-017-0910-x

Kazdin, A. E. (2017). *Research Design in Clinical Psychology* (5th ed.). New York: Pearson.

Keyes, C. L. M. (2002). The mental health continuum: From languishing to flourishing in life. *Journal of Health and Social Behavior*, *43*(2), 207–222. https://doi.org/10.2307/3090197

Kotov, R., Gamez, W., Schmidt, F., & Watson, D. (2010a). Linking "Big" personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychological Bulletin*, *136*(5), 768–821. https://doi.org/10.1037/a0020327

Kotov, R., Gamez, W., Schmidt, F., & Watson, D. (2010b). Linking "Big" personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychological Bulletin*, *136*(5), 768–821. https://doi.org/10.1037/a0020327

Langer, J. K., Tonge, N. A., Piccirillo, M., Rodebaugh, T. L., Thompson, R. J., & Gotlib, I. H. (2019). Symptoms of social anxiety disorder and major depressive disorder: A network perspective. *Journal of Affective Disorders*, *243*(March 2018), 531–538. https://doi.org/10.1016/j.jad.2018.09.078

Lee, J., Lee, E. H., & Moon, S. H. (2019). Systematic review of the measurement properties of the Depression Anxiety Stress Scales–21 by applying updated COSMIN methodology. *Quality of Life Research*, *28*(9), 2325–2339. https://doi.org/10.1007/s11136-019-02177-x

Lovibond, P. F., & Lovibond, S. H. (1995). The Structure of Negative Emotional States: Comparison of the Depression Anxiety Stress scales (DASS) with the Beck Depression and Anxiety Inventories. *Behavioral Research and Therapy*, *33*(3), 335–343. https://doi.org/10.1007/BF02511245

Madole, J. W., Rhemtulla, M., Grotzinger, A. D., Tucker-Drob, E. M., & Harden, K. P. (2019). Testing Cold and Hot Cognitive Control as Moderators of a Network of Comorbid Psychopathology Symptoms in Adolescence. *Clinical Psychological Science*, *7*(4), 701–718. https://doi.org/10.1177/2167702619842466

Malouff, J. M., Thorsteinsson, E. B., & Schutte, N. S. (2005). The relationship between the five-factor model of personality and symptoms of clinical disorders: A meta-analysis. *Journal of Psychopathology and Behavioral Assessment*, *27*(2), 101–114. https://doi.org/10.1007/s10862-005-5384-y

McNally, R. J. (2016). Can network analysis transform psychopathology? *Behaviour Research and Therapy*, *86*, 95–104. https://doi.org/10.1016/j.brat.2016.06.006

McNally, R. J., Robinaugh, D. J., Wu, G. W. Y., Wang, L., Deserno, M. K., & Borsboom, D. (2015). Mental disorders as causal systems: A network approach to posttraumatic stress disorder. *Clinical Psychological Science*, *3*(6), 836–849. https://doi.org/10.1177/2167702614553230

Moore, M. T., Anderson, N. L., Barnes, J. M., Haigh, E. A. P., & Fresco, D. M. (2014). Using the GAD-Q-IV to identify generalized anxiety disorder in psychiatric treatment seeking and primary care medical samples. *Journal of Anxiety Disorders*, *28*(1), 25–30. https://doi.org/10.1016/j.janxdis.2013.10.009

Moore, S. A., Zoellner, L. A., & Mollenholt, N. (2008). Are Expressive Suppression and Cognitive Reappraisal Associated with Stress-Related Symptoms. *Behavioral Research and Therapy*, *46*(9), 993–1000. https://doi.org/10.1038/jid.2014.371

Newman, M. G., Zuellig, A. R., Kachin, K. E., Constantino, M. J., Przeworski, A., Erickson, T., & Cashman-McGrath, L. (2002). Preliminary reliability and validity of the generalized anxiety disorder questionnaire-IV: A revised self-report diagnostic measure of generalized anxiety disorder. *Behavior Therapy*, *33*(2), 215–233. https://doi.org/10.1016/S0005-7894(02)80026-0

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

Schäfer, J. Ö., Naumann, E., Holmes, E. A., Tuschen-Caffier, B., & Samson, A. C. (2017). Emotion Regulation Strategies in Depressive and Anxiety Symptoms in Youth: A Meta-Analytic Review. *Journal of Youth and Adolescence*, *46*(2), 261–276. https://doi.org/10.1007/s10964-016-0585-0

Scott, J. L., Wozencroft, A. J., & Waller, S. N. (2019). Ecological Momentary Assessment Research in Therapeutic Recreation. *Therapeutic Recreation Journal*, *53*(1), 76–85. https://doi.org/10.18666/trj-2019-v53-i1-8797

United States Census Bureau. (2019). Educational Attainment in the United States: 2018. Retrieved September 5, 2020, from https://www.census.gov/data/tables/2018/demo/education-attainment/cps-detailed-tables.html

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

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*. Retrieved from http://www.jstatsoft.org/v45/i03/

Vasconcelos-Raposo, J., Fernandes, H. M., & Teixeira, C. M. (2013). Factor structure and reliability of the depression, anxiety and stress scales in a large Portuguese community sample. *Spanish Journal of Psychology*, *16*(May), 1–10. https://doi.org/10.1017/sjp.2013.15

Wang, S. B., Jones, P. J., Dreier, M., Elliott, H., & Grilo, C. M. (2018). Core psychopathology of treatment-seeking patients with binge-eating disorder: A network analysis investigation. *Psychological Medicine*, (2017). https://doi.org/10.1017/S0033291718002702

Yoon KL, Maltby J, Joormann J., & Yoon K.L., M. J. , J. J. (2013). A pathway from neuroticism to depression. *Anxiety, Stress and Coping*, *26*(5), 558–572.

Supplementary Materials

Appendix One

Scatter plots of post-hoc analyses

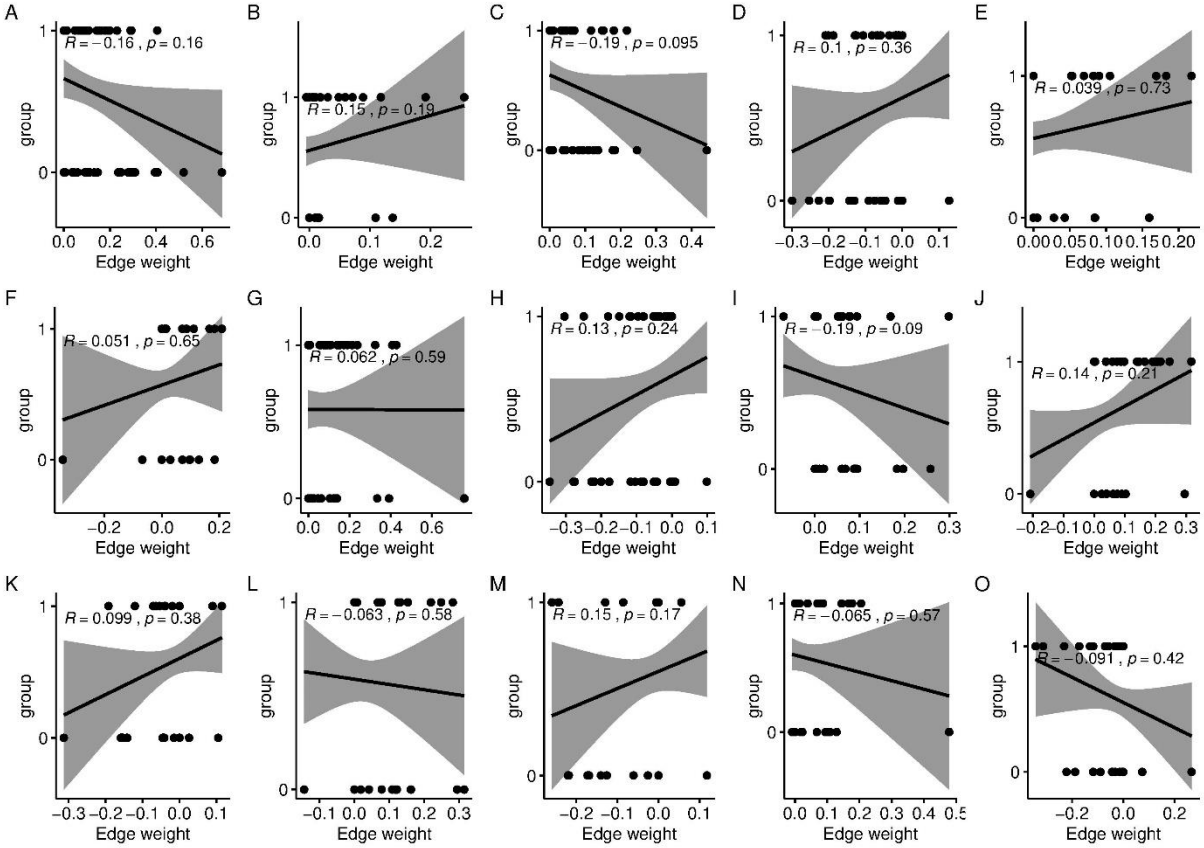

Figure 1: Scatter plots with the edge weights on the X-axis and the group membership on the Y-axis. Group 1: Clinical, Group 0: Control. Regression lines based on Spearman's rho. 95% confidence intervals are shown. 1A: Worry and Hopeless, 1B: Worry and Procrastinate, 1C: Worry and Ruminate, 1D: Worry and Positive, 1E: Worry and Fatigue, 1F: Hopeless and Procrastinate, 1G: Hopeless and Ruminate, 1H: Hopeless and Positive, 1I: Hopeless and Fatigue, 1J: Procrastinate and Ruminate, 1K: Procrastinate and Positive, 1L: Procrastinate and Fatigue, 1M: Ruminate and Positive, 1N: Ruminate and Fatigue, 1O: Positive and Fatigue.

Appendix Two

```
#################################
#                               #
#                               #
#     Master thesis analyses    #
#     Nikki van Eijk            #
#     November 2020             #
#                               #
#                               #
#################################

# ---------- 1. Load packages -----------------------------------------

library(dplyr)
library(tibble)
library(tidyr)
library(graphicalVAR)
library(ggpubr)
library(qgraph)
library(mlVAR)
library(Hmisc)
library(psych)
library(summarytools)

# ----------- 2. Recode the data to make ready for merging ---------------

## case one, 206 ##
#create new df so you don't change original data #
dat206PR <- dat206

#reorder the columns
dat206PR <- dat206[,c(1,2,29,28,5,23,6,22,21,20,7,
                      19,18,8,27,25,24,15,3,30,14,
                      11,13,9,10,4,12,17,16,26,31,32,33)]

#add new variable that was somehow not measured for this participant but is
in the column names
dat206PR$reassure <- NA
#put new variable in right position
dat206PR <- dat206PR[,c(1:22,33,23:32)]

#rename the columns
colnames(dat206PR) <- c(colnames(dat001[1:28]))
colnames(dat206PR)[29:33] <- c(1:5) #not needing these vars but colname
cannot be NA
#extract only variables everyone else has
dat206F <- select(dat206PR,1:28)


## case two, 244 ##
#create new df so you don't change original data
dat244PR <- dat244

#reorder the columns
dat244PR <- dat244[,c(1:21,25,22,23,28,27,26,24)]
```

```
#rename the columns
colnames(dat244PR) <- c(colnames(dat001[1:28]))

# ----------- 3. Merge the data -----------------------------------------

#### put all people with clinical diagnosis in one group ####

DataClin <- bind_rows(dat001,dat003,dat004,dat006,dat007,
                      dat008,dat009,dat010,dat012,dat013,
                      dat014,dat019,dat021,dat023,dat025,
                      dat033,dat037,dat040,dat048,dat068,
                      dat072,dat074,dat075,dat100,dat111,
                      dat113,dat115,dat117,dat127,dat137,
                      dat139,dat145,dat160,dat163,dat168,
                      dat169,dat173,dat202,dat203,dat204,
                      dat206F,dat215,dat217,dat219,dat220,
                      dat223,dat244PR, .id = "id")

#add group to make it distinguishable later
DataClin$group <- 1
DataClin2 <- select(DataClin,id,group)
DataClin2 <- distinct(DataClin2)

#### put all control people in one dataset ####

DataCont <- bind_rows(dat063,dat065,dat067,dat069,dat071,
                      dat078,dat082,dat087,dat092,dat094,
                      dat096,dat102,dat107,dat109,dat116,
                      dat120,dat123,dat126,dat128,dat129,
                      dat134,dat140,dat143,dat164,dat170,
                      dat174,dat177,dat178,dat181,dat189,
                      dat190,dat192,dat194,dat195,dat196,
                      .id = "id")

#### make one big dataset, everyone up to 244PR is clinical ####
DataFull <- bind_rows(dat001,dat003,dat004,dat006,dat007,
                      dat008,dat009,dat010,dat012,dat013,
                      dat014,dat019,dat021,dat023,dat025,
                      dat033,dat037,dat040,dat048,dat068,
                      dat072,dat074,dat075,dat100,dat111,
                      dat113,dat115,dat117,dat127,dat137,
                      dat139,dat145,dat160,dat163,dat168,
                      dat169,dat173,dat202,dat203,dat204,
                      dat206F,dat215,dat217,dat219,dat220,
                      dat223,dat244PR,
                      dat063,dat065,dat067,dat069,dat071,
                      dat078,dat082,dat087,dat092,dat094,
                      dat096,dat102,dat107,dat109,dat116,
                      dat120,dat123,dat126,dat128,dat129,
                      dat134,dat140,dat143,dat164,dat170,
                      dat174,dat177,dat178,dat181,dat189,
                      dat190,dat192,dat194,dat195,dat196,
                      .id = "id")

#add in variable to assign group membership
DataFull <- left_join(DataFull, DataClin2,copy=FALSE,by="id")
DataFull[["group"]][is.na(DataFull2[["group"]])] <- 0
```

```
# ---------- 4. Select variables for network analyses --------------------

summary(DataFull)

boxplot(DataFull[4:25],DataFull[27:29], las=2)

# ---------- 5. Extract all relevant information -------------------------

vars <- c("worried", "hopeless", "procrast", "ruminate", "positive",
"fatigue")

DataFull2 <-
select(DataFull,id,start,worried,hopeless,procrast,ruminate,positive,fatigu
e)

#convert the time
DataFull2$Time <- as.POSIXct(DataFull2$start,tz="","%m/%d/%Y %H:%M")

#extract days from the data
DataFull2$Day <- as.Date(DataFull2$Time, tz="","%m/%d/%y")

dayvar <- "Day"
idvar <- "id"

#check stationarity assumption
lm_wrr <- lm(worried ~ Day, data=DataFull2)
summary(lm_wrr)
lm_hpl <- lm(hopeless ~ Day, data=DataFull2)
summary(lm_hpl)
lm_prc <- lm(procrast ~ Day, data=DataFull2)
summary(lm_prc)
lm_rmn <- lm(ruminate ~ Day, data=DataFull2)
summary(lm_rmn)
lm_pst <- lm(positive ~ Day, data=DataFull2)
summary(lm_pst)
lm_ftg <- lm(fatigue ~ Day, data=DataFull2)
summary(lm_ftg)

# ----------- 6. Run the mlGraphicalVAR procedure -------------------------

MainRes <- mlGraphicalVAR(DataFull2,
                          gamma = 0,
                          vars=vars,
                          dayvar=dayvar,
                          idvar=idvar,
                          lags=1)

saveRDS(MainRes, "resultsMLGraphvar.rds")

#### alternatively, load the data file #####
MainRes <- readRDS(file="resultsMLGraphvar.rds")
####



# ---------- 7. Plot the results -----------------------------------------
```

```
dev.new()
qgraph(MainRes$fixedResults$PCC,
       layout="circle",
       posCol=c("blue2"),
       esize=20,
       vsize=10,
       vsize2=3,
       label.prop=0.9,
       labels=c("Worried","Hopeless","Procras-
   tination","Ruminate","Positive","Fatigue"),
       label.font=3,
       label.scale.equal=TRUE,
       edge.labels=TRUE,
       edge.label.cex = 1.3,
       edge.label.position = 0.55)


# ---------- 8. Create new df to add individual edges --------------------

# in this matrix we will put the values of individual edges per person
# as well as the moderators that we are interested in so we can eventually
# correlate these

#create as many columns as you have unique edges in the model
#for contemporaneous  self-loops not included
M <- matrix(NA, ncol=16, nrow=82)
M <- as.data.frame(M)

colnames(M) <- c("id",
                 "wrrhpl", "wrrprc",
                 "wrrrmn", "wrrpst",
                 "wrrftg", "hplprc",
                 "hplrmn", "hplpst",
                 "hplftg", "prcrmn",
                 "prcpst", "prcftg",
                 "rmnpst", "rmnftg",
                 "pstftg")

M[,1] <- MainRes$ids   #add the IDs of the participants

# ------------ 9. Extract the data from the individual networks ----------

# first, the adjacency matrices for each individual are put into a list
# then, from this list the individual edges can be put into the df

n=82

x <- list()
for (i in 1:n)
{
  x[[i]] <- as.matrix(MainRes$subjectPCC[i])
}

##### now, add this into the new df #####

#number 1
```

```
for (i in 1:n)
  if (is.null(x[[i]][[1]])) {

  } else {

    M[i,2] <-  x[[i]][[1]][1,2]   #worried and hopeless
  }

#number 2
for (i in 1:n)
  if (is.null(x[[i]][[1]])) {

  } else {

    M[i,3] <-  x[[i]][[1]][1,3]   #worried and procrastinating
  }

#number 3
for (i in 1:n)
  if (is.null(x[[i]][[1]])) {

  } else {

    M[i,4] <-  x[[i]][[1]][1,4]   #worried and ruminate
  }

#number 4
for (i in 1:n)
  if (is.null(x[[i]][[1]])) {

  } else {

    M[i,5] <-  x[[i]][[1]][1,5]   #worried and positive
  }

#number 5
for (i in 1:n)
  if (is.null(x[[i]][[1]])) {

  } else {

    M[i,6] <-  x[[i]][[1]][1,6]   #worried and fatigue
  }

#number 6
for (i in 1:n)
  if (is.null(x[[i]][[1]])) {

  } else {

    M[i,7] <-  x[[i]][[1]][2,3]   #hopeless and procrastinating
  }

#number 7
for (i in 1:n)
  if (is.null(x[[i]][[1]])) {
```

```
  } else {

    M[i,8] <-  x[[i]][[1]][2,4]   #hopeless and ruminate
  }

#number 8
for (i in 1:n)
  if (is.null(x[[i]][[1]])) {

  } else {

    M[i,9] <-  x[[i]][[1]][2,5]   #hopeless and positive
  }

#number 9
for (i in 1:n)
  if (is.null(x[[i]][[1]])) {

  } else {

    M[i,10] <-  x[[i]][[1]][2,6]   #hopeless and fatigue
  }

#number 10
for (i in 1:n)
  if (is.null(x[[i]][[1]])) {

  } else {

    M[i,11] <-  x[[i]][[1]][3,4]   #procrastinate and ruminate
  }

#number 11
for (i in 1:n)
  if (is.null(x[[i]][[1]])) {

  } else {

    M[i,12] <-  x[[i]][[1]][3,5]   #procrastinate and positive
  }

#number 12
for (i in 1:n)
  if (is.null(x[[i]][[1]])) {

  } else {

    M[i,13] <-  x[[i]][[1]][3,6]   #procrastinate and fatigue
  }

#number 13
for (i in 1:n)
  if (is.null(x[[i]][[1]])) {

  } else {

    M[i,14] <-  x[[i]][[1]][4,5]   #ruminate and positive
```

```
  }

#number 14
for (i in 1:n)
  if (is.null(x[[i]][[1]])) {

  } else {

    M[i,15] <-  x[[i]][[1]][4,6]   #ruminate and fatigue
  }

#number 15
for (i in 1:n)
  if (is.null(x[[i]][[1]])) {

  } else {

    M[i,16] <-  x[[i]][[1]][5,6]   #positive and fatigue
  }

# ------------ 10. Add/create the moderators ----------------------------

# first moderator: group membership (0: control, 1: clinical)
Grps <- select(DataFull,id,group)
Grps <- distinct(Grps)
Grps[["group"]][is.na(Grps[["group"]])] <- 0

M <- left_join(M,Grps,by="id")

##### Moderators from baseline data #####


#extract all cases from the clinical group
Baseline <- baselinedata[c(1,3,4,6,7,8,9,10,12,
                           13,14,19,21,23,25,33,
                           37,40,48,68,72,74,75,
                           100,111,113,115,117,127,
                           137,139,145,160,163,168,
                           169,173,202,203,204,206,
                           215,217,219,220,223,244),]

#extract all cases from the control group
BaselineC <- baselinedata[c(63,65,67,69,71,78,82,
                            87,92,94,96,102,107,109,
                            116,120,123,126,128,129,
                            134,140,143,164,170,174,
                            177,178,181,189,190,192,
                            194,195,196),]

#make one big dataframe with all participants
BaselineF <- bind_rows(Baseline,BaselineC,.id = "group")
BaselineF$id <- seq.int(nrow(BaselineF))

#select only the interesting moderators
Baseline2 <-
select(BaselineF,group,participantid,dassd,dassa,dasss,gadqtotal,erqrtotal,
erqstotal,neon,id)
```

```
Baseline2 <- Baseline2[,c(1,2,10,3:9)]


#add moderators to the full results dataset
Baseline2$id <- as.character(Baseline2$id)
M <- left_join(M,Baseline2,by="id")


### optional: write to SPSS file ###
library(haven)
write_sav(M, "Finaloutput.sav")


# ----------- 11.Perform subsequent analyses with moderators -------------

#select the 5 largest edges in the model
MainRes$fixedPCC

FO <-
select(M,id,wrrhpl,hplrmn,hplpst,wrrrmn,prcftg,group.x,dassd,dassa,dasss,ga
dqtotal,erqrtotal,erqstotal,neon)

#create sum score for the DASS
FO$dassSum <- FO$dassa + FO$dassd + FO$dasss

#correlation between edges and moderators
ct1 <- corr.test(FO[2:6],FO[15],adjust = "fdr",method="pearson")
ct2 <- corr.test(FO[2:6],FO[11],adjust = "fdr",method="pearson")
ct3 <- corr.test(FO[2:6],FO[12],adjust = "fdr",method="pearson")
ct4 <- corr.test(FO[2:6],FO[14],adjust = "fdr",method="pearson")

print(ct1,short=FALSE)
print(ct2,short=FALSE)
print(ct3,short=FALSE)
print(ct4,short=FALSE)


### 11.1 Visualisation ###
library(reshape2)
plot_Data <- melt(FO[1:6], id="id")
plot_Data <- left_join(plot_Data,FO[,c(1,7:15)],by="id")

#create 4 scatterplots (one for each moderator)
library(ggplot2)

mod1 <- ggplot(plot_Data, aes(x=value,y=dassSum,color=variable)) +
geom_point() +
  geom_smooth(method = "lm", alpha=0.1,aes(color=variable,fill=variable)) +
  theme_classic(base_family ="serif")

mod2 <- ggplot(plot_Data, aes(x=value,y=gadqtotal,color=variable)) +
geom_point() +
  geom_smooth(method = "lm", alpha=0.1,aes(color=variable,fill=variable)) +
  theme_classic(base_family ="serif")
```

```
mod3 <- ggplot(plot_Data, aes(x=value,y=erqrtotal,color=variable)) +
geom_point() +
  geom_smooth(method = "lm", alpha=0.1,aes(color=variable,fill=variable)) +
  theme_classic(base_family ="serif")

mod4 <- ggplot(plot_Data, aes(x=value,y=neon,color=variable)) +
geom_point() +
  geom_smooth(method = "lm", alpha=0.1,aes(color=variable,fill=variable)) +
  theme_classic(base_family ="serif")

#create new color palette
my_pal <-
palette(c("goldenrod1","dodgerblue","tomato1","limegreen","plum3"))


mod11 <- mod1 + labs(x="Edge weight", y="Total DASS score",tag="A") +
  scale_color_manual(values = palette(),
                     name="",
                     labels=c("Worry - Hopeless","Hopeless - Ruminate",
                              "Hopeless - Positive","Worry - Ruminate",
                              "Procrastinate - Fatigue")) +
  scale_fill_manual(values=palette(),
                    name="",
                    labels=c("Worry - Hopeless","Hopeless - Ruminate",
                             "Hopeless - Positive","Worry - Ruminate",
                             "Procrastinate - Fatigue"))

mod21 <- mod2 + labs(x="Edge weight", y="Total GAD-Q score",tag="B") +
  scale_color_manual(values = palette(),
                     name="",
                     labels=c("Worry - Hopeless","Hopeless - Ruminate",
                              "Hopeless - Positive","Worry - Ruminate",
                              "Procrastinate - Fatigue")) +
  scale_fill_manual(values=palette(),
                    name="",
                    labels=c("Worry - Hopeless","Hopeless - Ruminate",
                             "Hopeless - Positive","Worry - Ruminate",
                             "Procrastinate - Fatigue"))

mod31 <- mod3 + labs(x="Edge weight", y="Total ERQ reappraisal
subscale",tag="C") +
  scale_color_manual(values = palette(),
                     name="",
                     labels=c("Worry - Hopeless","Hopeless - Ruminate",
                              "Hopeless - Positive","Worry - Ruminate",
                              "Procrastinate - Fatigue")) +
  scale_fill_manual(values=palette(),
                    name="",
                    labels=c("Worry - Hopeless","Hopeless - Ruminate",
                             "Hopeless - Positive","Worry - Ruminate",
                             "Procrastinate - Fatigue"))

mod41 <- mod4 + labs(x="Edge weight", y="NEO neuroticism scale", tag="D") +
  scale_color_manual(values = palette(),
                     name="",
                     labels=c("Worry - Hopeless","Hopeless - Ruminate",
                              "Hopeless - Positive","Worry - Ruminate",
```

```
                                "Procrastinate - Fatigue")) +
  scale_fill_manual(values=palette(),
                    name="",
                    labels=c("Worry - Hopeless","Hopeless - Ruminate",
                             "Hopeless - Positive","Worry - Ruminate",
                             "Procrastinate - Fatigue"))

fs <- ggarrange(mod11,mod21,mod31,mod41,ncol=2,nrow=2,
                common.legend = TRUE, legend = "bottom")

#################

# ----------- 12. Visualise outcomes based on group ----------------------

sc1 <-  ggscatter(M, x = "wrrhpl", y = "group",
                  add = "reg.line", conf.int = TRUE, na.rm = TRUE,
                  cor.coef = TRUE, cor.method = "spearman")
sc2 <- ggscatter(M, x = "wrrprc", y = "group",
                 add = "reg.line", conf.int = TRUE, na.rm = TRUE,
                 cor.coef = TRUE, cor.method = "spearman")
sc3 <- ggscatter(M, x = "wrrrmn", y = "group",
                 add = "reg.line", conf.int = TRUE, na.rm = TRUE,
                 cor.coef = TRUE, cor.method = "spearman")
sc4 <- ggscatter(M, x = "wrrpst", y = "group",
                 add = "reg.line", conf.int = TRUE, na.rm = TRUE,
                 cor.coef = TRUE, cor.method = "spearman")
sc5 <- ggscatter(M, x = "wrrftg", y = "group",
                 add = "reg.line", conf.int = TRUE, na.rm = TRUE,
                 cor.coef = TRUE, cor.method = "spearman")
sc6 <- ggscatter(M, x = "hplprc", y = "group",
                 add = "reg.line", conf.int = TRUE, na.rm = TRUE,
                 cor.coef = TRUE, cor.method = "spearman")
sc7 <- ggscatter(M, x = "hplrmn", y = "group",
                 add = "reg.line", conf.int = TRUE, na.rm = TRUE,
                 cor.coef = TRUE, cor.method = "spearman")
sc8 <- ggscatter(M, x = "hplpst", y = "group",
                 add = "reg.line", conf.int = TRUE, na.rm = TRUE,
                 cor.coef = TRUE, cor.method = "spearman")
sc9 <- ggscatter(M, x = "hplftg", y = "group",
                 add = "reg.line", conf.int = TRUE, na.rm = TRUE,
                 cor.coef = TRUE, cor.method = "spearman")
sc10 <- ggscatter(M, x = "prcrmn", y = "group",
                  add = "reg.line", conf.int = TRUE, na.rm = TRUE,
                  cor.coef = TRUE, cor.method = "spearman")
sc11 <- ggscatter(M, x = "prcpst", y = "group",
                  add = "reg.line", conf.int = TRUE, na.rm = TRUE,
                  cor.coef = TRUE, cor.method = "spearman")
sc12 <- ggscatter(M, x = "prcftg", y = "group",
                  add = "reg.line", conf.int = TRUE, na.rm = TRUE,
                  cor.coef = TRUE, cor.method = "spearman")
sc13 <- ggscatter(M, x = "rmnpst", y = "group",
                  add = "reg.line", conf.int = TRUE, na.rm = TRUE,
                  cor.coef = TRUE, cor.method = "spearman")
sc14 <- ggscatter(M, x = "rmnftg", y = "group",
                  add = "reg.line", conf.int = TRUE, na.rm = TRUE,
                  cor.coef = TRUE, cor.method = "spearman")
sc15 <- ggscatter(M, x = "pstftg", y = "group",
```

```
                  add = "reg.line", conf.int = TRUE, na.rm = TRUE,
                  cor.coef = TRUE, cor.method = "spearman")

# add extras to the scatteplots
sc11 <- sc1 + scale_y_continuous(breaks=seq(0, 1)) +
  scale_x_continuous(name = "Edge weight") + labs(tag="A")
sc21 <- sc2 + scale_y_continuous(breaks=seq(0, 1)) +
  scale_x_continuous(name = "Edge weight") + labs(tag="B")
sc31 <- sc3 + scale_y_continuous(breaks=seq(0, 1)) +
  scale_x_continuous(name = "Edge weight") + labs(tag="C")
sc41 <- sc4 + scale_y_continuous(breaks=seq(0, 1)) +
  scale_x_continuous(name = "Edge weight") + labs(tag="D")
sc51 <- sc5 + scale_y_continuous(breaks=seq(0, 1)) +
  scale_x_continuous(name = "Edge weight") + labs(tag="E")
sc61 <- sc6 + scale_y_continuous(breaks=seq(0, 1)) +
  scale_x_continuous(name = "Edge weight") + labs(tag="F")
sc71 <- sc7 + scale_y_continuous(breaks=seq(0, 1)) +
  scale_x_continuous(name = "Edge weight") + labs(tag="G")
sc81 <- sc8 + scale_y_continuous(breaks=seq(0, 1)) +
  scale_x_continuous(name = "Edge weight") + labs(tag="H")
sc91 <- sc9 + scale_y_continuous(breaks=seq(0, 1)) +
  scale_x_continuous(name = "Edge weight") + labs(tag="I")
sc101 <- sc10 + scale_y_continuous(breaks=seq(0, 1)) +
  scale_x_continuous(name = "Edge weight") + labs(tag="J")
sc111 <- sc11 + scale_y_continuous(breaks=seq(0, 1)) +
  scale_x_continuous(name = "Edge weight") + labs(tag="K")
sc121 <- sc12 + scale_y_continuous(breaks=seq(0, 1)) +
  scale_x_continuous(name = "Edge weight") + labs(tag="L")
sc131 <- sc13 + scale_y_continuous(breaks=seq(0, 1)) +
  scale_x_continuous(name = "Edge weight") + labs(tag="M")
sc141 <- sc14 + scale_y_continuous(breaks=seq(0, 1)) +
  scale_x_continuous(name = "Edge weight") + labs(tag="N")
sc151 <- sc15 + scale_y_continuous(breaks=seq(0, 1)) +
  scale_x_continuous(name = "Edge weight") + labs(tag="O")

# create grid with all figures
figure <- ggarrange(sc11, sc21, sc31, sc41, sc51,
                    sc61, sc71, sc81, sc91, sc101,
                    sc111, sc121, sc131, sc141, sc151,
                    ncol = 5, nrow = 3, widths=1.5)
figure

# ----------- 13. Descriptives - get age groups from data ----------------

dates <- BaselineF$dob
dates <- data.frame(dates)
dates$dates <- as.character(dates$dates)

dates$dob <- as.Date(  sub("/(..$)", "/19\\1",dates$dates)   , "%m/%d/%Y")
dates$dates <- as.Date(dates$dates, format="%m/%d/%y",
origin=as.Date("1900-01-01"))
dates$age <- NA
dates$age <- floor(age_calc(dates$dob,enddate = as.Date("2014-12-31"),
units="years"))


labs <- c(paste(seq(0, 95, by = 10), seq(0 + 10 - 1, 100 - 1, by = 10),
```

```
                sep = "-"), paste(100, "+", sep = ""))

dates$AgeGroup <- cut(dates$age, breaks = c(seq(0, 100, by = 10), Inf),
labels = labs, right = FALSE)
summary(dates$AgeGroup)

summarytools::freq(dates$AgeGroup, order = "freq")
```