A systematic review on the match of research questions, methods, and conclusions, regarding the within-person level versus the between-person level in the psychological network literature on MDD.

Bachelor's Project in Methodology and Statistics
Project number: 9
Utrecht University

Student: Babet de Wit
Student number: 6504841
Date: 28 june 2021

Supervisor: Noémi Schuurman

1. Introduction

Major depressive disorder (MDD) is a mental illness that is prevalent among more than 264 million people worldwide. According to the World Health Organization, it is a leading cause of disability worldwide, also for low-income and middle-income countries which have less money and possibilities for treatment of mental disorders (Wang et al., 2007; WHO, 2020). MDD contributes significantly to the overall global burden of disease and appears at every age group. In recent years, suicide due to MDD has occurred about 800.000 times a year worldwide (WHO, 2020). Given this worldwide prevailing mental illness which is causing a lot of societal damage, it is important that proper treatments and preventions for MDD are developed. To do this, it is essential that current theories are investigated, and supplemented or revised when necessary. New theories will emerge from this to better understand the disorder of MDD. Research on those existing and the development of new theories must be reliable and valid. This is important as preventing and treating MDD is about human lives.

In psychological research, a relatively new but increasingly popular theory is the network theory of mental disorders. The network theory has been applied to various mental disorders in scientific research, including MDD. According to this theory, mental disorders arise from the causal interactions in a network of symptoms (Borsboom, 2017). For example, taken a person who has a lot of stress due to their work-pressure. This work pressure then causes poor sleep. The bad night rest leads to concentration problems, which have an effect on the persons work performance (Choi et al., 2017). This accumulation of effects, created by the experience of stress, gives the person a sense of failure and a depressive mood develops. The person can't fall a sleep at night because he is mulling over his bad performance at work and his depressed mood (Elovainio et al., 2018). This mulling than leads to the experience of feelings of worthlessness. The person experiences this feeling worthlessness for a long time, and after a few weeks the person gets suicidal thoughts. This network of all these related symptoms together can be conceptualized as MDD, according to the network theory. This theory relates to within-person differences. The theory focusses on causal processes of mutually reinforcing symptoms that take place within individuals over time. The network structure in which these processes take place is complex and consists of direct causal relations (Borsboom, 2017). For this reason, the network theory looks at how mental disorders arise within a person, and how the interaction of symptoms differs over time. Hence, the most direct way to investigate this is with repeated measurements of one or more individuals.[1] Note however, that the composition of these network structures for mental disorders may be different for each individual. If someone wants to analyze the within-person structure for a particular individual, one needs to perform idiographic research. To perform this type of research, a lot of repeated measures for that individual are needed. The idiographic analysis is performed on an individual level by comparing one person with himself or herself over time (Blaauw et al., 2014).

There is a considerable number of studies conducted on psychological psychopathology networks. Robinaugh et al. (2020) found 204 empirical studies that addressed a mental disorder and incorporated the perspective of the network theory. But there is a debate going on in the literature

---

[1] Note that, within-person relations may also be investigated with RCTs (Randomized Control Trials) or correlational cross-sectional datasets. But, for RCTs you only get the average within-person relationship across that whole group, you don't see how the effect differs from person to person. And regarding correlational cross-sectional datasets, if you want to find a within effect you will have to take additional measures to separate the within-person variance and the between-person variance. An example of how to do this, is by the use of a control variable. The between-person variance could be filtered out by making a control variable that is only strongly associated with between-person differences in your variable. In this way, only within-person differences remain. But it is very difficult to find a good control variable which doesn't filter out other things, and you will not be sure if the variances are actually separated properly from each other unless you research this by another measurement method (Schuurman & Ryan, 2021).

about the validity of these psychological network studies (Bos & Wanders, 2016; Bringmann & Eronen, 2018; Forbes et al., 2017; Tzur-Bitan, Meiran & Shahar, 2010). In particular, there is criticism that cross-sectional research is performed which is not able to demonstrate causality and cannot be expected to generalize to the level of the network theory: the level of the individual.

An example of an article in which this problem is addressed is that of Bos & Wanders (2016). They analyzed the study of Van Borkulo et al. (2015). Concerning the network theory of MDD, Van Borkulo et al. (2015) found that a network of MDD patients with persistent MDD were more densely connected at baseline than the network of the remitted MDD patients. They argued that more strongly connected symptoms implicate higher risk to MDD because there is stronger feedback among the symptoms. Regarding this research, the critique of Bos & Wanders (2016) was that just because symptoms co-occur across cases at the group-level, this does not imply they influence each other over time within individuals. To illustrate, take for example the MDD symptom's loss of interest and fatigue: If there is a cross-sectional association, people who score high on loss of interest compared to other people, often score high on fatigue compared to other people. This is something else than a within-person association, which would mean that if a person scores high on loss of interest at t=1, this person will tend to score high on fatigue at t=2.

In this thesis I will investigate whether there is a more widespread problem with the internal validity of the empirical psychological network literature, with respect to this discussion. Specifically, my aim is to examine whether research questions, methods, and conclusions, in the psychological network literature on MDD, are consistent with each other regarding the within-person level versus the between-person level. If the levels of the research question and methods don't match, it may be difficult (or not possible) to answer that research question. If the conclusions don't match the level analysis, these conclusions may be not valid, as is argued by Bos & Wanders (2016) in their critique of Van Borkulo et al. (2015).

In this systematic literature review, I will build upon a recent systematic review on the psychological network literature performed by Robinaugh et al. (2020). In this review I will include an analysis of 20 randomly selected empirical articles on MDD which were included in the study of Robinaugh et al. (2020). The analysis will be performed by the manual coding of the articles. Based on this coding process, I will manually create a codebook. To ensure reliability and validity of my codebook, my thesis supervisor Noémi Schuurman will also code 10 of the 20 articles. Before I make my codebook, I will compare our coding's to check whether we analyzed the articles more or less in the same way. In the future, this codebook could be used to analyze more psychological network studies regarding the within-person level versus the between-person level.

In the following, I will first provide a theoretical background in section two to provide more context on what psychological networks are, as well as what is meant with the 'within-person' and 'between-person' level. In section three the methodological approach for the systematic literature review is described, including the developed codebook, followed by the results in section four. At last, in section five I will provide a conclusion and discussion.

2. Background

In the following chapter, I will provide information about the within-person level, the between-person level, cross-sectional data and analysis, psychological MDD networks, and commonly used methods to estimate psychological networks. To illustrate the differences between the within-person level, the between-person level and the cross-sectional level, I will use an empirical example. To this

end I will make use of time-series data from the study of Bringmann et al. (2013). The time-series data is obtained by an ESM (Experience Sampling Methodology) study in 2009. The data is collected from 129 adult participants with residual symptomatology after at least one episode of MDD. The participants were followed over a 12-day period. Participants were notified by a beeper ten times (every 90 minutes) between 07:30 a.m. and 10:30 p.m. each day. When they received a signal, they had to complete the ESM self-assessment form. In this form they had to assess their mood and social context in daily life. All self-assessments were rated on 7-point Likert scales (with 1 = not at all and 7 = very) (Geschwind et al., 2011). For the study of Bringmann et al. (2013), there are a number of items for different types of moods displayed. Given the focus on MDD for this thesis, I am interested in finding out to what extent ''feeling sad'' as an item for mood differs per day. To keep things simple for the following examples, I only looked at the fourth beep on five consecutive days for three participants (10720, 10726 and 10742).

2.1 Within-person level and intra-individual processes

Many important research questions in psychology are about causal processes that happen within persons over time. The within-person level concerns such *intra*-individual processes. Intra-individual processes are the dynamics of a psychological variable within a subject over time (Schuurman & Ryan, 2021). The development of a child learning how to talk and the regulation of serotonin levels in relation to eating are examples of intra-individual processes. The network theory of mental disorders focusses on such intra-individual processes. An example of an intra-individual process for MDD is that within person X, feeling guilty influences depressed mood and concentration. Depressed mood influences suicidality, and concentration influences psychomotor retardation. If for person X, his feelings of guilt increase at t(in months)=5, it is likely that depressed mood and concentration will increase at t=6. It is also likely that suicidality and psychomotor retardation will increase at t=7.

Examples of within-person research questions are: ''Is the baseline network structure of MDD symptoms associated with the <u>longitudinal course</u> of MDD?'' (Van Borkulo et al., 2015), ''What is the <u>dynamic relationship</u> between experienced traumatic events, daily stressors, and symptoms of psychological distress in post-war settings?'' (Bringmann et al., 2015), and ''Do characteristics of a <u>dynamic symptom network</u> provide insight <u>regarding long-term symptom</u> persistence?'' (Groen et al., 2019).

These within-person processes may differ from person to person. Every individual may even be seen as a unique system of interacting dynamic processes (Molenaar, 2004). For example, in the case of one person suffering from MDD, a depressed mood might lead to fatigue. This fatigue in turn leads to decreased appetite, which in turn leads to psychomotor retardation. This while in another person suffering from MDD, concentration problems and hypersomnia lead to loss of interest, which in turn lead to a depressed mood, which eventually leads to suicidality. Both of these individuals suffer from MDD but have completely different within-person processes of interacting MDD symptoms.

If a psychological variable within a subject takes on different values at different times, there is within-person variance. The most direct way to study this type of variance is to observe the process as it is taking place within an individual over time. This can be done by using intensive sampling methods over time, that is, taking many repeated measures per person (Hamaker, 2012). Within-person variance is the variance of the values for a certain psychological variable for a subject across time (Schuurman & Ryan, 2021).

To illustrate this within-person variance, I will use the example on the data of Bringmann et al. (2013). The data for the three participants (10720, 10726 and 10742) on five consecutive days is shown in Table 1. The means for the subjects, taken over the five repeated measurements, are 4.2 (10720), 3.8 (10726) and 1.4 (10742). The within-person variances, calculated over the five repeated measures (4,4,3,5 & 5 for 10720; 2,6,3,4 & 4 for 10726, and 2,2,1,1 & 1 for 10742) for the subjects are: 0.7 (10720), 2.2 (10762), and 0.3 (10742).

Table 1

Empirical example: differences in within-person level, between-person level & cross-sectional level

Scores of feeling Sad	Day 1	Day 2	Day 3	Day 4	Day 5	Person's Mean across days	Within-person variance
10720	4	4	3	5	5	4.2	0.7
10726	2	6	3	4	4	3.8	2.2
10742	2	2	1	1	1	1.4	0.3
Cross-sectional variance	1.33	4	1.33	4.33	4.33		
Between-person variance						2.29	

Note. Data are from the study by Bringmann et al. (2013).

2.2 Between-person level and interindividual differences

Most psychological research focuses on *inter*-individual differences: differences between persons. The between-person level concerns inter-individual differences. Inter-individual differences indicate that the value of a psychological variable varies from subject to subject (Schuurman & Ryan, 2021). For example, two children learning how to talk may know a different number of words. Or serotonin levels produced from food may differ between different individuals. Inter-individual differences for MDD are for example the different values for depressed mood of different subjects.

Examples of between-person research questions are: ''Do evening types have an increased susceptibility to MDD?'' (Merikanto et al., 2013), ''Are there gender differences in factors associated with smartphone addiction?'' (Chen et al., 2017) and ''Do many patients with MDD suffer from bipolar disorder?'' (Angst, 2006).

Inter-individual differences are often studied utilizing cross-sectional data sets, which are datasets where many individuals are measured at a certain point of time. But, in the context of the within-between discussion, between-subject variance of a variable is not the variance of the observed values for one time point (the cross-sectional variance). The between-person variance looks at a more specific type of inter-individual differences. Between person-variance is the variance of the *expected values – the mean -* of a certain psychological variable taken across the different subjects (Schuurman & Ryan, 2021). These expected values of the subjects represent their general tendency of the value for a psychological variable. It is a characteristic that is considered persistent over time.

To illustrate this between-person variance I will again use the example on the data in Table 1 of Bringmann et al. (2015). The expected values (means) for the three subjects are 4.2 (10720), 3.8

(10726) and 1.4 (10742). This may indicate that subject 10720 on average is more sad than subject 10726, and that subject 10726 is on average more sad than subject 10720. The between-person variance is the variance of these expected values (4.2, 3.8 and 1.4), which is 2.29.

2.3 Cross-sectional data/analysis

When performing cross-sectional analysis, one also looks at inter-individual differences. But the cross-sectional variance is different than the between-person variance discussed in the previous subsection. Cross-sectional variance is the variance of the scores of a certain variable across different subjects at a certain point of time (Hamaker, Mulder & van IJzendoorn, 2020; Schuurman & Ryan, 2016).

To illustrate this type of variance and to show the differences with within-person variance and between-person variance I will use the example on the data of Bringmann et al. (2015) in Table 1. The cross-sectional variance – taken over the scores of the three subjects at one day - is 1.33 for day 1, 4 for day 2, 1.33 for day 3, 4.33 for day 4, and 4.33 for day 5. The key point that I want to make with the example illustrated in Table 1 is that it is very important to make a right distinction in the type of variance that is being looked at. This is really important because the three types of variances give different types of results. When these variances are mixed up, incorrect results can come out and wrong conclusions can be drawn.

Specifically, it can be shown, that the cross-sectional variance approximates the sum of the average within-person variance and the between-person variance (Schuurman & Ryan, 2021). How large the within-person and between-person variance will be, and hence which contributes the most to the cross-sectional variance, varies from study to study. In order to draw conclusions about within-level results, or about the between-variance in the means, with cross-sectional data, you need to be able to distinguish those types of variation in your cross-sectional dataset. Given that psychological network theory is clearly about the within-person level, I expect that it will then be important for many network studies to isolate the within-variance in the cross-sectional dataset.

As with the variance of the variables, covariances and correlations may also be different at within-, between-, and cross-sectional levels. To illustrate this, instead of looking at the variance of one variable, in the following two examples I will look at the within-, between-, and cross-sectional correlation of two variables.

The first example is about the association on muscle ache (1 = no ache, 10= worst ache imaginable) and speed of running (km per hour) for 20 participants, running t=100 times. The association is positive for the within-person level, this means that the faster people run, the more muscle ache they have the next day (0.55). The within-person variance for muscle ache is 0.4 and the within-person variance for speed of running is 1. Opposed to this, the association on the between-person level is negative: participants that on average run harder than other participants, have on average relatively less muscle ache the next day than other participants (-0.4; because they are more experienced runners). This example is illustrated in Figure 1. In this example most of the variation comes from the variation in the means of the participants: the between-subject variance. For this reason, this cross-sectional association here resembles the between-subject association the most and is equal to -0.44 at t=20.

The second example is also about the association on muscle ache (1 = no ache, 10= worst ache imaginable) and speed of running (km per hour), but now the 20 participants are top sporters. The participants are still running t=100 times. The association is positive for the within-person level

(0.73). This means that the faster the participants run, the more muscle ache they get the next day. The within-person variance for muscle ache is 0.7 and the within-person variance for speed of running is 1.5. There is a really weak association on the between-person level here, all the top sporters are good at running and they all run at about the same speed (0.07). The between-person variance for muscle ache is 0.4 and the between-person variance for speed of running is 1. This example is illustrated in Figure 2. Here you can see that most of the variation in the association comes from the within-subject variance. This could be due to the fact that the participants have some days training where they get more muscle ache the next day than from than other trainings. This could be because, for example, on one day they were in a hurry and did not do a cooling down, while on another day they did a very serious warming up and cooling down. In this example, the cross-sectional association (0.53) resembles almost only the within-subject association.

Figure 1

Within-person association, between-person association and cross-sectional association for 20 participants
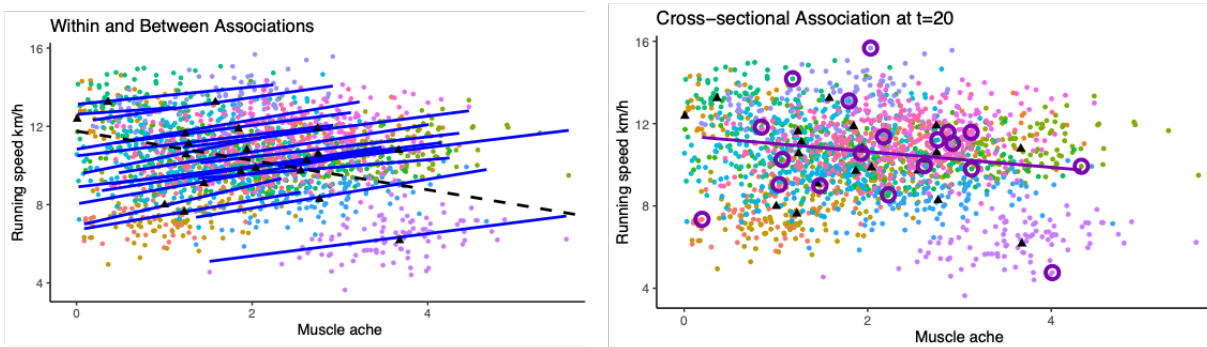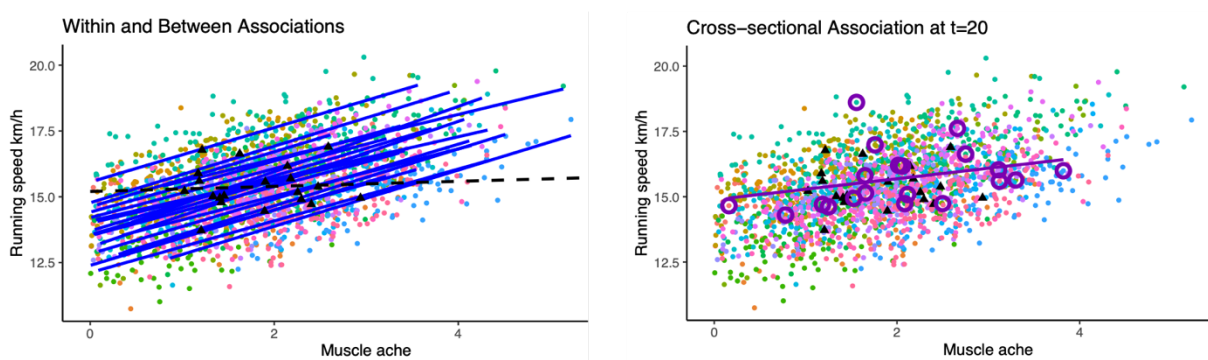

Figure 2

Within-person association, between-person association and cross-sectional association for 20 top sporters

What can be made up of these examples is that the cross-sectional association is a blend of the within-person association and the between-person association. What can be seen is that how a variable is distributed within a person, can be different per person. It can differ per study whether the cross-sectional association resembles the within-person association or the between-person association more. Both associations are meaningful and maybe interesting, but they are relevant for different types of research questions (Schuurman & Ryan, 2021). For the internal validity of a study,

it is therefore very important to distinguish in a proper way between these different levels.
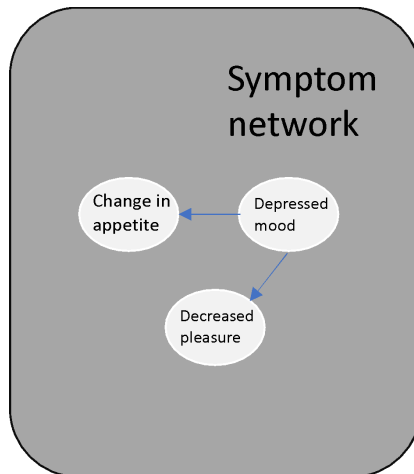
2.4 Psychological networks of MDD

MDD as a mental disorder is generally diagnosed by ''symptoms''. These symptoms are illustrated in diagnostic manuals such as DSM-5 and ICD-10. Researchers don't have a consensus of what MDD precisely is. In the case of a disease such as the flu, it is clear that the symptoms are all the result of that virus – the virus is a common cause of the symptoms. Given this clarity of the cause of the symptoms, it is easy to state that a vaccine is a possibility to stop the virus in some way. It is not precisely clear what causes MDD and what it consists of (Borsboom, 2017). For this reason, there is not a concrete way to prevent or treat MDD. It is therefore important that theories of MDD are investigated, and that proper MDD theories are developed.

A new and increasingly popular theory on MDD comes from the psychological network theory. According to the psychological network theory, psychiatric symptoms cause each other instead of being effects of a common cause (Borsboom, 2017). According to this theory, symptoms are ''constitutive'' of mental disorders, not reflective of them. This means that symptoms are not caused by the mental disorder, but that mental disorders emerge from the causal interactions of symptoms. In the light of this approach, MDD occurs when the necessary number of symptoms are triggered for a sufficient time period (McNally, 2016). An example of symptoms which could influence each other in the case of MDD is as follows: at first a person experiences sadness and insomnia, these two symptoms keep on strengthening each other, eventually leading to feeling guilty and feeling worthless. This last symptom triggers sadness even more, causing a symptom of suicidal ideation. This network theory is about the within-person level: It concerns how symptoms interact in a dynamic system within an individual.

The network structure in which these causal interactions of symptoms take place is called the ''symptom network''. Figure 1 represents an example of such a symptom network for Tim, who suffers from MDD. In this network, symptoms are *nodes*. In Figure 1, depressed mood, change in appetite and decreased pleasure are nodes which represent three symptoms of MDD. The *nodes* are connected through *edges*, which reflect the relationships between the symptoms. Edges can be categorized *unweighted* or *weighted*, and *undirected* or *directed*. If the edges are unweighted, you only know that the symptoms are associated. If edges are weighted, you also know the strength of their association. All edges in Figure 1 are unweighted. An undirected edge is a single line connecting two symptoms. If this is the case you don't know what the direction of the relation between two symptoms is. A directed edge has an arrow on the line representing the direction of the relation (McNally, 2016). In Figure 1, depressed mood is connected through direct edges to both change in appetite and decreased pleasure. This means that the depressed mood influences the change in appetite and decreased pleasure.

Figure 3

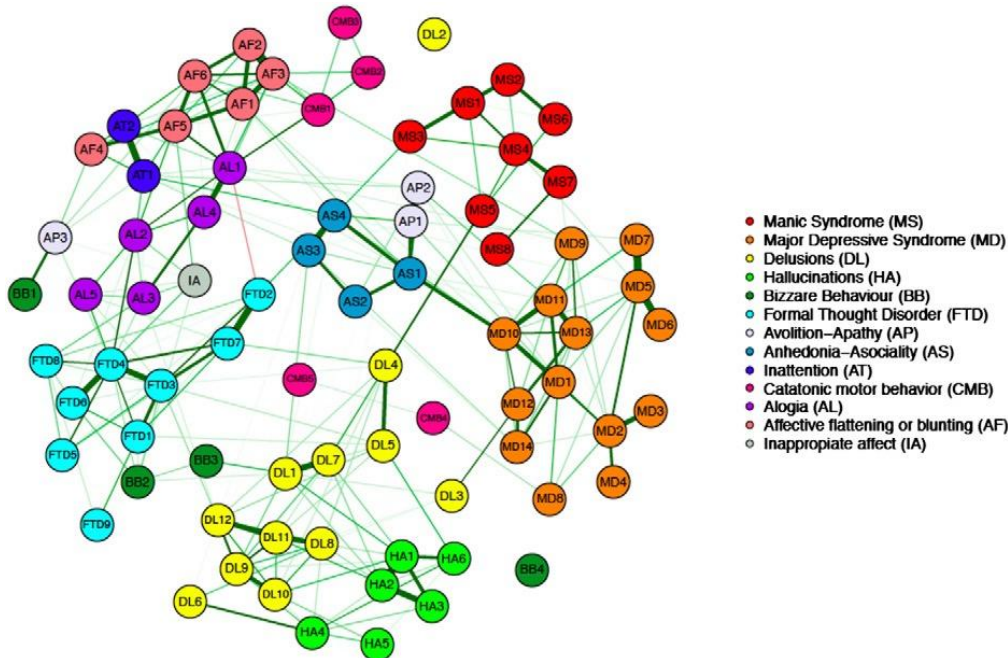*Example of a MDD symptom network*

2.5 Commonly used methods to estimate psychological networks

There are several research methods which can be used for estimating psychological networks. These methods are applied in one of the following three dimensions: cross-sectional network-analyses, dynamic network-analyses, and multi-level dynamic network-analyses.

For cross-sectional network analysis, the data used consists of a sample of multiple cases at a certain point of time. This type of data is called cross-sectional data (Hamaker, Mulder & Van IIzendoorn, 2020). In the case of MDD, the cases of interest are typically individual persons. Typical cross-sectional network methods are networks based on Pairwise Markov Random Fields, such as Gaussian Graphical Models for normally distributed variables or Ising models for dichotomous variables. The networks based on these techniques are also referred to as association networks and partial correlation networks. Association networks calculate whether symptoms correlate with each other and are not directed. But correlations are an insufficient basis for causal inference. Partial correlation networks calculate the partial correlation between symptom X and symptom Y after correcting for the influence of all other symptoms in a network. Partial correlation networks are also not directed. You don't know which symptom influences the other symptom, or it could also be the case both symptoms influence each other (McNally, 2016). Figure 4.1 represents a cross-sectional partial correlation network structure from the study of Van Rooijen et al. (2017, p. 79). The network model is based on the data of 408 male participants. It represents 79 symptoms; the colors represent a priori symptom domains, and the numbers are specific symptoms. The green associations are positive, and the red associations are negative. The thicker the lines, the stronger the association is. There are no directions in the correlations of the symptoms.
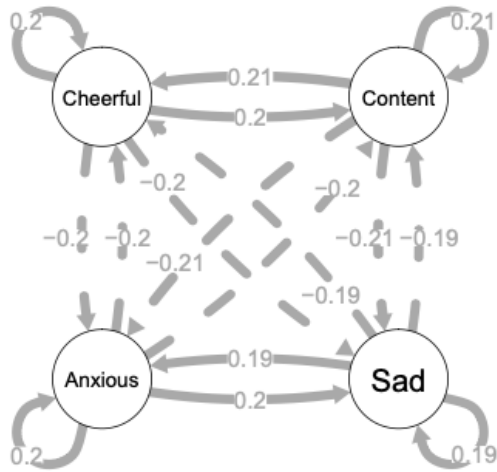
Figure 4.1

Cross-sectional network structure

Note. Reprinted from ''A symptom network structure of the psychosis spectrum'', by Van Rooijen et al., 2017, *Schizophrenia research, 189,* p. 79.

The problem with cross-sectional network analyses is that it cannot easily confirm causality among symptoms – which happens within persons over time - because they only measure at one single timepoint. For this reason, network researchers have designed methods that are better able to characterize mental disorders as causal systems (McNally, 2016). These methods are applied in dynamic network analyses and multi-level extensions of these analyses.

Dynamic network analyses try to find out what the relations between variables over time are. For this type of analysis, intensive longitudinal data is used. Intensive longitudinal data consists of many repeated measures for multiple individuals (Schuurman, 2016). A popular method for creating dynamic networks is vector autoregressive (VAR) modelling, which is a time-series model. In this type of method, each subject is analyzed individually. The researcher looks at how variables influence each other over time for one person. These analyses provide results at the within-level (Schuurman, 2016). Figure 4.2 represents a network structure estimated from a VAR model from the article of Haslbeck & Ryan (2020, p. 16). The network structure is based on time series data of the variables cheerful, content, anxious and sad. The network displays the matrix of lagged regression parameters, estimated from a VAR model (Haslbeck & Ryan, 2020, p. 16).
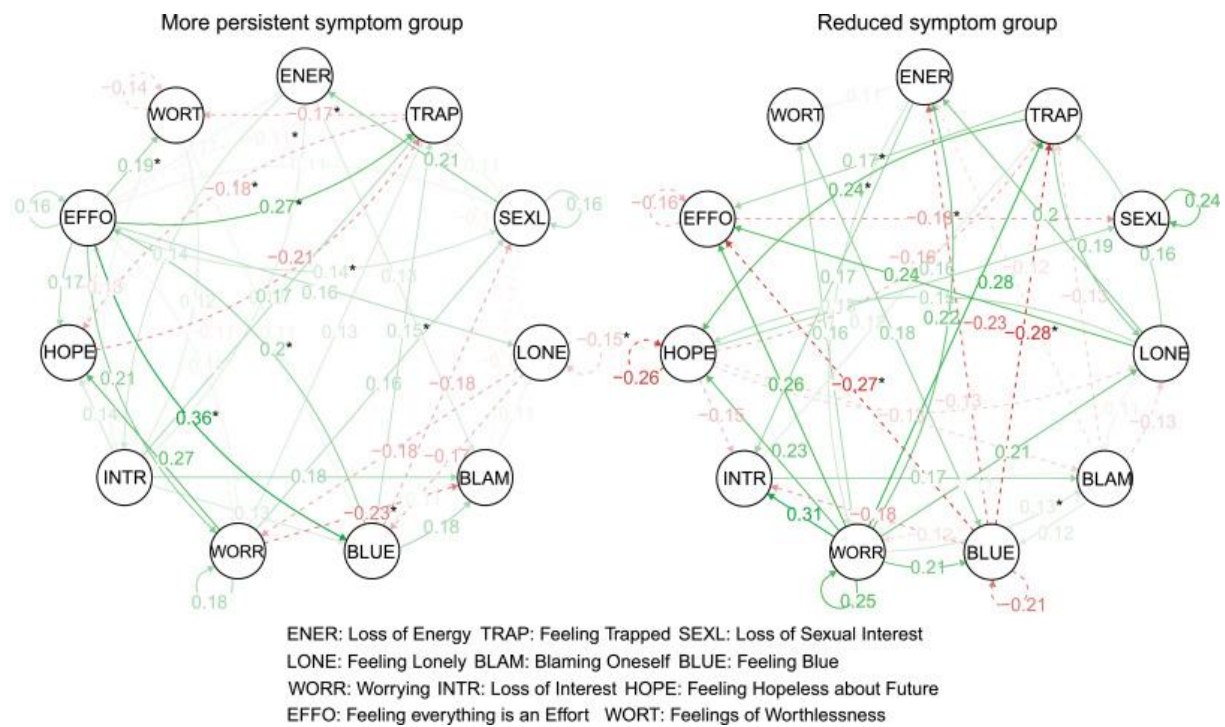
Figure 4.2

Dynamic network structure

 Note. Reprinted from ''Recovering within-person dynamics from psychological time series'', by J. Haslbeck & O. Ryan, 2020, p. 16.

Even though these dynamic analyses are better able to analyze causal systems, they don't take into consideration the between-level. This is due to the fact that they analyze results for one individual at time.

   Lastly, there is a method that combines research on the between-level and the within-level and is also able to separate the two levels properly: the multi-level VAR method. This method is used for multi-level dynamic network analyses. These multi-level models consist of two levels of modelling. At the first level there is a time-series model that describes within-person processes. At the second level, at the same time, between-person differences are modelled in the dynamic features (Hamaker & Wichers, 2017). Some models even create a within-person dynamic network and a between-person network. This is possible because with multi-level models you explicitly separate within from between and thus can create a network for both. This compared to correlational cross-sectional studies cannot easily separate the two from each other (Schuurman, 2016). Figure 4.3 represents a multi-level VAR network structure from Groen et al. (2019, p.645). The two distinct symptom networks represent two different symptom groups: the more persistent symptom group and the reduced symptom group. Strengths of the edges are based on the estimates of the multi-level VAR models and are implied by the thickness and darkness of the edges. Each edge is representing the association between symptoms or the autoregressive effect over time (one lag is one day). Green lines are positive associations and red lines are negative associations. Significantly different edges between groups are marked with an (*) in each group's network.

Figure 4.3

Multi-level VAR network structure

More persistent symptom group | Reduced symptom group

ENER: Loss of Energy TRAP: Feeling Trapped SEXL: Loss of Sexual Interest
LONE: Feeling Lonely BLAM: Blaming Oneself BLUE: Feeling Blue
WORR: Worrying INTR: Loss of Interest HOPE: Feeling Hopeless about Future
EFFO: Feeling everything is an Effort   WORT: Feelings of Worthlessness

Note. Reprinted from ''Capturing the risk of persisting depressive symptoms: A dynamic network investigation of patients' daily symptom experiences'', by Groen et al., 2019, *Psychiatry Research*, *271,* p. 645.

If the data or techniques don't match the level of the research question regarding the between-person level or the within-person level, sensible conclusions cannot be drawn. In the next section I will examine how much this inconsistency is a problem in psychological network literature by means of a systematic literature review.

3. Methods

One of the objectives of this study was to develop a codebook for analyzing 20 articles with a focus on MDD networks from the study of Robinaugh et al. (2020). To ensure reliability and validity of my codebook, my thesis supervisor Noémi Schuurman also coded 10 articles. Another objective of this study is that this codebook can be used later on to analyze studies that empirically examined psychological networks. I developed the codebook in the qualitative data software program NVivo 12 Pro Pro. In this method section I will first describe how I collected the papers for this systematic review. I will also provide a list of all the papers. After that I will clarify the codes and the codebook. At last, I will explain how I differentiated between the within-person level and between-person level.

3.1 Sample

I based my systematic review on a recently performed extensive systematic literature review by Robinaugh et al (2020). They reviewed literature (2008 – 2018) on the network approach to psychopathology. They analyzed 204 empirical articles that looked at psychological processes in the light of the network theory. From these 204 articles analyzed by Robinaugh et al. (2020), 69 articles have a focus on MDD, and are thus relevant for the aim of my review. Because I first had to create a codebook manually, I had limited time to analyze all the 69 articles. For this reason, I randomly selected 20 of the 69 articles in SPSS Statistics 26. I found out that one article was about a sociological network and not about MDD. For this reason, I excluded this article and performed another random sample on the remaining 49 articles in SPSS Statistics 26 for one new article. A list of references of the 20 articles is included in appendix A.

For this study, my main interest is in obtaining substantive information on whether the interests and goals of the researchers, with respect to the within-between levels, align with the methods and conclusions of the selected articles. For this reason, the unit of analysis in this review are sentences regarding: the motivation of the research question; the research question; interest in the network approach; data type; analysis technique; conclusions; clinical implications, and cross-sectional disclaimer.

3.1.1 Motivation of the research question

First of all, the motivation of the research question clarifies why the researchers perform their study. It represents more context behind the research question. It could for example be the case that a research question is formulated unclearly but the motivation and what they do in the study is consistent. If this is the case, it is clear that the research is inconsistent due to a poor formulation of the research question. Because of the motivation of the research question, you know this is the problem. An example of a piece of text which is coded as a motivation of the research question from one of the twenty articles is: ''Unfortunately, mainstream clinical assessment methods cannot be used in order to assess the causal relatedness often perceived to exist between co-occurring psychological problems.'' (Frewen et al., 2013).

3.1.2 The research question

The research question itself is very important because that is what the researchers are trying to answer with their study. This is what is relied upon throughout the study. For this reason, it is certainly a problem if it is on another level than the methods or conclusions, as this can lead to inconsistency of what they want to achieve with the study and what they actually do. An example of a research question of one of the articles is: ''The aims of the current study were to I) construct a symptom network and investigate interactions between a wide array of psychotic symptoms; II) identify the most important symptoms within this network and III) perform an explorative shortest pathway analysis between depressive and delusional symptoms.'' (Van Rooijen et al., 2017).

3.1.3 Interest in the network approach

Interest in the network approach is interesting because, just like the motivation for the research question, it outlines the context of the study. But also, it examines whether the networks they draw up are in line with what they want to describe in their study. In Afzali et al. (2017) the following piece

of text is coded as interest in the network approach: ''An alternative approach to conceptualizing comorbidity is the network approach… From the network perspective, each mental disorder represents a complex constellation of symptoms, clustered by pairwise relations (Borsboom et al., 2011a).''

3.1.4 Data type

The data type is a part of the methods and says something about on how many persons and over how many repeated measures variables were studied. I look at the data type because it differs for each type of data whether the between-person level, the within-person level and/or longitudinal network change can be studied. For example, ''…repeated administrations of the questionnaire to a group of depressed individuals who participated in a treatment study of an average of 14 weekly assessments.'' (Bringmann et al., 2015) and '' The sample consisted of 174 individuals (72% female) with primary DSM-IV-TR diagnosis of SAD'' (Heeren, Jones & McNally, 2018) are coded as datatypes.

3.1.5 Analysis technique

Another part of the methods is the analysis technique; this says something about how networks are constructed and analyzed based on the data. The analysis technique is important because it tells us whether the within-person effect can be separated in a valid way from the between-person effect. Examples of pieces of text coded as analysis technique are ''Using multi-level VAR models, I investigated the temporal relation between this perceived social pressure and depressive symptoms to determine directionality.'' (Dejonckheere et al., 2017) and ''Networks were estimated using Gaussian graphical models and lasso regularization for all the children, and separately for those who were adopted at different ages.'' (Elovainio et al., 2018).

3.1.6 Conclusions

The conclusions are important because they answer the research question, based on the methods. This is what the study has discovered according to the researchers. If it is inconsistent, there is a chance that it will be used incorrectly in other studies. A piece of text which is coded as a conclusion from Jaya et al. (2017) is: ''Additionally, as expected, the network analysis identified a unique connection between loneliness and symptoms of paranoia, specifically the impression that other people are giving odd looks and that other people are not what they seem to be. Other than expected, however, we did not find a connection between loneliness and hallucinations. Rather, hallucinations seemed to be associated with loneliness through a complex web of other positive symptoms.''

3.1.7 Clinical implications

The clinical implications are part of the conclusions. Clinical implications are possible medical recommendations based on the findings of the study. If these are not consistent with the level at which the research was performed, they are incorrect and may have harmful consequences for those who adopt them. A clinical implication from Jaya et al. (2017) is: ''One implication of the result is that interventions targeting loneliness are likely to be an effective adjunct to early interventions for

psychosis as they could improve both depression and psychotic symptoms.''

3.1.8 The cross-sectional disclaimer

The cross-sectional disclaimer is often mentioned in the conclusion and says something about whether cross-sectional methods were used incorrectly for the aim and conclusions of the study. When this is done, a study is inconsistent, but the researchers themselves acknowledge this. An example of a cross-sectional disclaimer in one of the articles is ''…our reliance on cross-sectional analyses precluded us from confirming causal relationships among symptoms.'' (Robinaugh et al., 2014).

I want to know something about what the researchers specifically did regarding MDD networks in their study. Therefore, sentences relating to anything else but networks; publications other than the selected articles; results; strengths; limitations (besides the cross-sectional disclaimer); directions for future research of the articles; and (potential) explanations for the results, were not considered in this thesis. Sentences relating to anything else but networks are omitted because they can also say something about the between-level or within-level, but not in the context of the discussion this paper is about. Results are not looked at because the main findings that answer the research question are presented in the conclusions. The results section provides a detailed overview of what the study found by the use of a lot of numbers. For my research question it is not necessary to look at these numbers for stating whether a paper is inconsistent regarding the between-person level and the within-person level. Strengths, limitations, directions for future research, and explanations also do not answer the research question and are beyond the scope of the purpose of this thesis.

3.2 Codes and Codebook

3.2.1 First codes: Type of sentences

At first, I coded sentences whether they were related to: the motivation of the research question; research question; interest in the network approach; data type; analysis technique; conclusions; clinical implications, and cross-sectional disclaimer. This coding was performed in NVivo12. Table 2 shows how all the selected sentences were defined while coding. This is the first part of the codebook.

Table 2
First part of the codebook

Section	Definition
Motivation of the research question	In the motivation of the research question the author(s) explain why this study is conducted. Most of the time these are both scientific and social reasons. Often it is a somewhat larger piece of text which can be found in the introduction. It is often referred to by sentences including ''In a recent study…'', ''Therefore, understanding… '', or ''To explain…''.

Research question	The research question is what the study aims to answer. It can be found in the introduction and is often repeated in the first sentence of the conclusion section. The research question is usually denoted with sentences like ''The main goal...'', ''The aim of this study…'' ''The current study…'', or ''This paper will…'''.
Interest in the network approach	If the study says anything about network approach, it is most of the time done introduction. A piece of text on this theory is often referred to as: ''An alternative approach…'', ''The network approach…'', ''The causal systems perspective…'' or ''Newer models of psychopathology…''.
Data type	What kind of data was collected, considering the number of people, and whether or not it was measured over time, and how often. The data type can be found at the beginning of the methods section. Often denoted with sentences like ''Data were derived from…'', ''This study used data from…'', ''We utilized data from...'', ''Our sample'' or '' ''''We selected…''
Analysis technique	The method that is used for analyzing the data. This can be found in the methods section, usually after the data type was discussed. Denoted with sentences including ''…models'', ''… analysis'', or ''…networks''. Techniques you often encounter in psychological network studies are ''Gaussian Graphical Models'', ''Ising models'', ''vector autoregressive (VAR) models'' and ''multi-level VAR models''.
Conclusions	The conclusion section provides a summary of the main results and answers to the research question of a study. It does not provide any new information compared to the result-section. The conclusion either has its own chapter or is the part of the discussion. It is often denoted with sentences like ''According to our findings …'' ''In conclusion…'' or ''Taken together…''.
Clinical implications	Clinical implications are possible recommendations of concrete clinical steps or actions, that are made based on the finding of a study. Clinical implications are often referred to by sentences including ''Clinical practice…'',''Interventions…'' & ''Targeting…''.
Cross-sectional disclaimer	Whether there is a disclaimer about the use of a cross-sectional methods. When an article has this disclaimer, it almost always appears in the conclusion and/or discussion under the limitations. It is often referred to as ''Due to the cross-sectional nature of the data…'', ''Our cross-sectional design…'', or ''Data were cross-sectional''.

3.2.2 Secondary codes: Within, between, longitudinal network change, neutral/unclear

In the second stage, the selected sentences were reviewed to determine if they were about a between-level network or a within-level network. This analysis was also performed in NVivo 12 Pro Pro. For the motivation of the research question, the research question, the interest in the network

approach, conclusions, and clinical implications I coded if the sentences were on the within-level, the between-level or neutral/unclear. I only coded a piece as within or between if the sentence provided information on whether the connections in the networks themselves should be at the within-person level or the between-person level. Other pieces were coded as neutral/unclear. While Noémi and I were coding, we found out that there was another level: longitudinal network change. At this network level, authors refer to changing networks. I added this level as code. At this level, change is on the within-level, but the networks themselves (that changed) could be cross-sectional, at the between-level or at the within-level. In the following, I will give some examples from the articles to illustrate when a sentence was coded at a certain level. Definitions for the codes and example identifier words can be found in the second part of the codebook in Table 3.

In the following research question from Bringmann et al. (2015) the underlined part was coded as within: ''…this paper is the first to investigate these patterns of short-term (i.e. session to session) dynamics for a widely used psychological questionnaire for depression – the Beck Depression Inventory (BDI-II).'' I coded this as within because it is about the dynamic relationships between symptoms within persons over time.

The clinical implication ''…clinicians could prioritize the assessment of identified bridging PTSD symptoms to screen patients with higher risk of MDD comorbidity in early post-trauma evaluations.'' from Afzali et al. (2017) is coded as between, because this clinical implication involves the targeting of symptoms which is based on the comparison of psychological variables between individuals.

Examples of conclusions that were coded as neutral/unclear are ''Low-energy was another highly central depression symptom'' and '' The symptoms of 'sad mood' and 'too much worry' were most central to the network'' (Beard at al., 2016). These pieces of text were coded as neutral/unclear because there are not within-person level or between-person level identifiers in the conclusions. The statements do say something about a network , but it is not clear what kind of network they are discussing.

As a result of analyzing the article of Beard et al. (2016), I added the code longitudinal network change. I also found this level back in the article of Bos et al. (2018). An example of a research question with a coding at this level from that article is: ''The aim of this study was to explore the possibilities of using cross-sectional network analysis to increase our understanding of symptom connectivity before and after SSRI treatment.'' (Bos et al., 2018). This research question falls under this level because it is about changes in the network over time: before and after treatment. But is not about network connections themselves.

3.2.3 Secondary codes: Data type and analysis type

The data types I refer to in this study are cross-sectional data, longitudinal (panel) data and intensive longitudinal (time-series) data. Regarding the data type I coded if the data was cross-sectional data, longitudinal (panel) data or intensive longitudinal (time-series) data. Cross-sectional data consists of data of subjects which are measured at a certain point of time. An example of an article using as cross-sectional data is that of Jayawickreme et al. (2017). Besides: ''Participants were 337 Sri Lankan Tamil survivors of war who received psychosocial assistance from the Family Rehabilitation Center (FRC).'', they say nothing related to longitudinal data or repeated measures used for their study. Longitudinal (panel) data consists of data of subjects for a few repeated measurements over time. An example of longitudinal (panel) data from the article of Malgaroli, Maccallum & Bonnano (2018) is ''

Participants were administered structured clinical interviews at 3 months post-loss (T1: M= 2.67; S.D. = 1.01), 14 months post-loss (T2: M= 14.25; S.D. = .98), and 25 months post-loss (T3: M= 24.92; S.D. = .64).'' Intensive longitudinal (time-series) data consists of data of subjects for many repeated measurements over time, and an example of a data type coded as intensive longitudinal (time-series) data is '' A sample of individuals (n=112) with elevated depression scores (Patient Health Questionnaire [PHQ-9] ≥ 10) took part in an online daily diary study in which they rated their depressive symptoms and perceived social pressure not to feel depressed or anxious for 30 consecutive days.'' (Dejonckheere et al., 2017).

Beside the data type, I also looked at the analysis technique. As described in the background section, examples of commonly used analysis techniques in psychological network studies are Gaussian Graphical Models, Ising models, vector autoregressive (VAR) models and multi-level VAR models. When looking at the analysis technique, I coded whether it was cross-sectional networks or multi-level time series (VAR). There were no single subject vector autoregressive (VAR) models. Typical cross-sectional network techniques are Gaussian Graphical Models and Ising models. An example of a piece of text from which a part is coded as cross-sectional networks is ''Networks were estimated using Gaussian graphical models and lasso regularization for all the children, and separately for those who were adopted at different ages.'' (Elovainio et al., 2018). The coding of a multi-level VAR analysis technique from Johnson & Hoffart (2018) is as follows: ''The data was analyzed using the multi-level vector autoregressive (mlVAR) model…''. At the article of Frewen et al. (2013) it was another analysis technique that was used: ''We calculated Mean Causal Association and Mean Effect Association scores as outlined in Frewen et al. (2013). The Mean Causal Association score for any given item is the average Causal Association score the symptom receives across all other items rated when occupying ''Symptom X'' in the causal association question: ''How much do you think your problems with [Symptom X] CAUSE your problems with [Symptom Y]?''. I coded this as ''other techniques''.

Based on how I coded these secondary codes, and by comparing my codes with those of Noémi Schuurman for the first ten articles, I created the second part of the codebook, which is provided in Table 3. The codebook shows when certain types of sentences (from the first part of the codebook) have been given a code for a particular level, data type or analysis technique. This second part of the codebook consists of 10 different codes overall. The first four codes relate to the level of the motivation of the research question, the research question, the interest in the network approach, conclusions, and clinical implications. The first column in the Table for these codes shows which level the row refers to: the within-person level, the between-person level, the level of longitudinal network change, or neutral/unclear. The second column shows a definition of what kind of statements in psychological network studies are at this level. The other six codes refer to the methods, of which the first three refer to the data type and the last three to analysis techniques. The first column shows which data type or analysis technique the row refers to. The second column gives a definition of what the data type or analysis technique consists of. For all 10 codes, the third column contains examples of identifier words. These identifier words are typical words that appear in sentences coded for a particular level, data type or analysis technique. The examples of identifier words come from the 20 analyzed psychological network studies. The codebook is developed to analyze the within-between distinction in psychological network studies.

Table 3

Second part of the codebook

Code	Definition	Examples of Identifier words
Within-person level	Statement is about psychological intra-individual network processes. It is about the variability over time of one or more psychological variables within an individual. So, the statement can be about causal effects of certain symptoms on other symptoms in a network. But also about dynamic relationships between symptoms taking place over the time within a person.	''…short-term dynamics'' ''…session-to-session dynamics'' ''…symptom- or cluster-level mechanisms of behavioral risk.'' ''…dynamic interrelations of different symptoms.'' ''…causal interplay between psychiatric symptoms.'' ''…increase in appetite is unlikely to directly cause suicidal ideation.'' ''…could help prevent, development of…'' ''Interventions targeted at … should focus on …''
Between-person level	Statement is about inter-individual network differences. The statement is about a comparison on psychological variables between individuals. So, it looks at how a person is in higher risk of a mental disorder than another person. Or what individuals should get a certain treatment.	''Screen patients with higher risk of …'' ''Not all individuals ….'' ''…heterogeneity in … symptoms across individuals'' ''…targeting interventions to those …[individuals] at highest risk.''
Longitudinal network change	Statement is about changes in the network over time, but not about the networks/networks connections themselves	''…whether the network changed over the course of the treatment.'' ''…increased from admission to discharge.'' '' We constructed networks before the start of treatment and after 8 weeks of treatment.'' ''…compare network structures before and after SSRI treatment.'' ''We found a cluster consisting of cognitive symptoms both at baseline and week 8.''
Neutral or Unclear	There are no within-person level or between-person level identifiers in the statement.	''… map the structure of symptom associations between.'' ''…the relationship between any pair of directly-connected nodes cannot be

	These can be statements that have nothing to do with the network connections. Or there is no clear indication in the pieces of text about what kind of connections are involved, in this case often vague words like "connections" or "associations" are used without further information about what kind of connections or associations.	explained by <u>the presence of</u> other nodes in the network.'' ''Results indicate that, in this network, all BDI-II symptoms are directly or indirectly <u>connected</u>.'' ''…for each <u>symptom–symptom relation</u>, we controlled for the effect of all other symptoms in the network.''
Cross-sectional data	The data consists of measurements from a sample of individuals at one point in time. Often this is identified primarily by a notion of absence of repeated measurements. Sometimes also explicitly named as cross-sectional by the authors.	''Data comes from a sample of 909 Australian adults…'' ''…data from a community sample who reported a traumatic event and at least one period of two weeks when most of the day they felt sad, discouraged, or uninterested.'' ''Cross-sectional baseline data…'' ''Our sample…comprised 778 internationally adopted children.''
Longitudinal (panel) data	The data consists of a sample of individuals for a few repeated measurements over time.	''…at admission and discharge.'' ''…at 4 waves: at baseline (T0e), and at 3 follow-ups after on average 1.6 (T1e), 3.5 (T2e), and 8.4 years (T3e), respectively.'' ''… 210 subjects participated in the study 18 months after the loss (Wave 2), and 106 participated 48 months after the loss (Wave 3).''
Intensive longitudinal (time-series) data	The data consists of a sample of individuals for many repeated measures over time.	''Each patient participated in 3–20 weekly individual sessions, depending on the progress of the patient or due to drop out. On average, patients completed 14 sessions (S.D. =5).'' ''…repeated administrations of the questionnaire to a group of depressed individuals who participated in a treatment study of an average of 14 weekly assessments.'' ''A sample of individuals (N=112)…took part in an online <u>daily diary</u> study…''
Cross-sectional networks analysis technique	A network technique computing a network of inter-individual differences.	''…pairwise Markov random fields'' ''…the Ising model…'' ''Partial correlation networks…''

		''Graphical Gaussian Model''
		''Cross sectional network analysis''
Multi-level time series (VAR) analysis technique	A network technique computing a network of both intra-individual processes and inter-individual differences.	''…vector autoregressive (VAR) multi-level method'' '' …multi-level VAR method'' ''…multi-level model consisting of fixed (average) and random (individual) effects'' ''Vector autoregressive (VAR) models with a multi-level extension''
Other analysis technique		''Mean Causal Association'' ''Mean Effect Association''

4. Results

4.1 Meta data analysis: differentiation between the within- and between-person level

To see whether the research questions, methods, and conclusions match with regard to the within-person level and between-person level in the psychological network literature on MDD, I evaluated 20 articles. I performed this evaluation through coding in the program NVivo 12 Pro Pro. Table 4 provides a summary of the main results, showing whether the articles were consistent regarding the level of the research question, methods, and conclusions. The Table shows for the research question conclusions and clinical implications what the distribution regarding the levels is. I calculated the proportion of within, between, longitudinal network change and neutral/unclear for the research questions, the conclusions, and clinical implications. For calculating this, I counted how many within-level, between-level, longitudinal network change and neutral/unclear codes each part (research question, conclusions, clinical implications) had. After this I added up the total number of codes for all the levels, and then divided the share per level by this total and did this proportion times 100%. For example for the research question of article 4: (4 within-person level codes/8 total of codes) x 100% = 50% within-level research question. Table 5 provides a summary of the complementary results. These results concern the context of the research: the motivation for the research question and the interest in the network approach. But also, whether the article addresses a disclaimer for the use of cross-sectional data. This Table shows for the motivation of the research question and the interest in the network approach what the distribution is for the levels. Note that in both Table 4 and Table 5 it could be that for some papers the percentages do not add up to exactly 100%, this is because I rounded to whole numbers. In the following, at first, I will discuss the frequencies of the types of coding's. After this I will provide the overall distribution of the: research questions; methods; conclusions; motivations for the research questions; interests in the network approach and cross-sectional disclaimer. And lastly, I will discuss the coherence of the findings.

Table 4

Summary of the main results

Article	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Research question																				
Within		17%		50%		100%	20%	33%		100%			50%	67%		33%			60%	
Between																				
Longitudinal network change		17%	29%														50%			
Neutral/Unclear	100%	67%	71%	50%	100%		80%	67%	100%		100%	100%	50%	33%	100%	67%	50%	100%	40%	100%
Methods																				
Data type:																				
Cross-sectional	✓			✓	✓		✓		✓	✓	✓	✓	✓	✓				✓	✓	
Longitudinal (panel)		✓	✓													✓	✓			✓
Intensive longitudinal (time series)						✓		✓							✓					
Analysis technique:																				
Cross-sectional networks	✓	✓	✓	✓	✓		✓		✓		✓	✓	✓	✓		✓	✓	✓	✓	✓
Multilevel time series (VAR)						✓		✓												
Other										✓					✓					
Conclusions																				
General:																				
Within	43%	29%	17%		30%	67%	40%	90%	39%	71%	55%	25%	13%	13%	17%	8%	19%		31%	29%
Between				33%					6%											
Longitudinal network change		25%	50%													50%	33%			
Neutral/Unclear	57%	46%	33%	67%	70%	33%	60%	10%	56%	29%	45%	75%	88%	88%	83%	42%	48%	100%	69%	71%
Clinical implications:																				
Within	50%					100%	33%	100%	67%			100%	100%	50%	25%			100%	100%	100%	
Between	17%						17%							50%	25%					
Longitudinal network change																				
Neutral/Unclear	33%						50%		33%						50%					
Consistent	No	No	No	No	No	Yes	No	Yes	No	X	No	No	No	No	X	No	No	No	No	No

Table 5

Summary of the complementary results

Article	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Motivation for the research question																				
Within		25%		50%		100%	100%	33%		55%	33%	50%	43%		38%	33%	57%		75%	
Between															38%					
Longitudinal network change																				
Neutral/Unclear	100%	75%	100%	50%	100%			67%	100%	45%	67%	50%	57%	100%	25%	66%	43%	100%	25%	100%
Interest in network approach																				
Within		80%	60%	25%	55%	50%	57%	50%	55%	60%	63%	66%		63%	100%	33%	69%		50%	80%
Between							14%										8%			
Longitudinal network change																				
Neutral/Unclear	100%	20%	40%	75%	44%	50%	29%	50%	44%	40%	36%	33%	100%	38%		66%	23%	100%	50%	20%
Cross-sectional disclaimer	✓	✓	✓	✓	✓		✓		✓		✓	✓	✓	✓		✓	✓	✓	✓	✓

4.2 Coding frequencies

In total, there have been 1.081 codes attached to the 20 articles. Of these, there are 541 codes related to the first coding: the type of text. There are 57 pieces of text coded as motivations of the research question, 45 pieces of text coded as research question, 56 pieces of text coded as interest in the network approach. Regarding the methods, 58 pieces of text were coded as data type, and 62 as analysis technique. 161 pieces of text were coded as conclusions, 44 as clinical implications and 29 as cross-sectional disclaimer. There were 540 pieces of text coded for the second coding: the level of the of the piece of text. Of this second coding, 200 pieces of text were coded as within, only 10 as between, 42 as longitudinal network change, and 247 pieces of text were coded as neutral/unclear. There were 12 pieces of text coded as cross-sectional data, five as longitudinal (panel) data and three as intensive longitudinal (time-series) data. Lastly regarding the analysis technique, there were 16 pieces of text coded as cross-sectional network analysis technique, two as multi-level time series (VAR) analysis technique and 2 as other analysis technique.

For the research questions, the conclusions, the motivations of the research question and the interests in the network approach, I calculated the proportion of a certain level (within, between, longitudinal network change, neutral/unclear) across all the twenty articles. I did this by adding up all the percentages for one level for all the articles and divided tis percentage by 2000% (20 articles x 100% per article). This number I did times 100% for the eventual proportion of that level. For example for proportion neutral/unclear in the research questions: (100% + 67% + 71% + 50% + 100% + 80% + 67% + 100% + 100% + 100% + 50% + 33% + 100% + 67% + 50% + 100% + 40% + 100%)/ 2000% = 0.6875 x 100% = 69% of the codes of the research questions was neutral or unclear.

4.2.1 Research question

Regarding the 57 pieces of text coded as research question, 26% was at the within-person level, 0% at the between-person level, 5% at the level of longitudinal network change and 69% was neutral or unclear. There were only two articles which had their research question only at the within-person level. Another eight articles had a research question which was partly at the within-person level and partly neutral or unclear. There were no articles for which the research question was at the between-person level, not even partly. Of the 20 articles, there were three articles which partly had their research question at the level of longitudinal network change, and partly within and/or neutral or unclear. We found that eight articles had a research question that was completely neutral or unclear.

4.2.2 Methods

Regarding the methods, I looked at the data type and the analysis technique. Of the 20 articles, 12 used cross-sectional data, five used longitudinal (panel) data and three used intensive longitudinal (time-series) data. The analysis techniques are fairly consistent with this given that 16 analysis techniques were cross-sectional networks, two were multi-level time series (VAR), and two were ''Other'' techniques. When I look at what methods the papers used, there were 16 studies inconsistent in the use of their methods, two studies consistent and of two studies I cannot tell.

4.2.3 Conclusions

I looked at the general conclusions and the clinical implications in the articles. With the general conclusions, I see roughly the same pattern as with the research questions. Of the 161 pieces of text coded as conclusions, 32% are at the within-person level, 2% at the between-person level, 8% at the level of longitudinal network change and 58% neutral/unclear. There is only one article which has their conclusions at only one level, which is neutral/unclear. Only 12 of the 20 articles had clinical implications. Of these 12 articles, there were seven articles which had clinical implications which were only at the within-person level. What is remarkable, is that four articles had clinical implications which were partly at the between-person level. There were no clinical implications found at the level of longitudinal network change. There were four articles which partly had clinical implications which were at the level of neutral/unclear. There where were no articles which had clinical implications which were for 100% at the between-person level, or 100% neutral/unclear.

4.2.4 Motivation for the research question

There were 57 pieces of text coded as motivation for the research question. These pieces of text were coded for 34% at the within-person level, 2% at the between-person level, 0% at the level of longitudinal network change, and 64% were neutral or unclear. For 2 articles, the motivation for the research question was entirely at the within-person level and for 7 articles it was entirely neutral or unclear which network level the motivation for the research was.

4.2.5 Interest in the network approach

Looking at the interest in the network approach, 51% of the 56 pieces of text were coded at the within-person level, 1% at the between-person level, 0% at the longitudinal network change level, and 48% of the pieces of text were coded as neutral/unclear.

4.2.6 Cross-sectional disclaimer

There were 16 of the 20 articles which included a cross-sectional disclaimer. Of the four articles that did not have a cross-sectional disclaimer, two did not use cross-sectional data and analysis techniques. And the other two used ''other'' analysis techniques. So, all the papers which purely performed a cross-sectional network analysis had a cross-sectional disclaimer.

4.3 Coherence

The results show that most of the research questions and conclusions are at the neutral/unclear, followed by the within-person level. Looking at the context of the studies, I see the same happening for the motivations of the research questions. Regarding the interests in the network approach, I coded the majority at the within-person level, followed by neutral/unclear. All 20 articles do not include any research questions and almost no conclusions that are at the between-person level. Looking at the methods, the most common data type is cross-sectional data and the most commonly used analysis technique is cross-sectional network analysis. Cross-sectional data and cross-sectional networks do not offer a completely reliable possibility to separate the within-person level from the

between-person level. This is remarkable considering that almost all studies include less or more a within-person level part in their research question and conclusions.

What can be made-up from the main results in Table 4 is that only two of the 20 articles are consistent in terms of their research question, methods, and conclusions regarding the between-person level and the within-person level. These two studies are not cross-sectional studies. Of these studies, the first one has a 100% within research question, and the other one has a 33% within and 67% neutral/unclear research question. The data type of both is intensive longitudinal (time-series) and the analysis technique is for both multi-level time series (VAR). The first one has conclusions at the within-person level (67%) and neutral/unclear (33%). And the clinical implications in this study are only at the within-person level. The second study has also conclusions at the within-person level (90%) and neutral/unclear (10%). And also for this study, the clinical implications are only at the within-person level. For two articles, I cannot say whether the two are consistent because I did not focus on their type of analysis techniques in this thesis. What I can see from the additional analysis is that the sixteen inconsistent papers do all have a cross-sectional disclaimer, indicating the authors are to some extent aware of the inconsistency.

5. Discussion

In closing, in this section I will discuss what the contribution of this review has been, the limitations and implications for future research, the overall conclusion, and lastly the practical implications. In this thesis I investigated whether there is a problem with the internal validity of psychological network studies. I examined this through a systematic literature review, whether research questions, methods, and conclusions in the psychological network literature on MDD, are consistent with each other regarding the within-person level and the between-person level. The first finding is that the research questions and conclusions were generally consistent in the studies that I analyzed. Most of the research questions had codes which were neutral/unclear, followed by most codes at the within-person level. The same distribution holds for the conclusions. I do not think it is a good sign that 69% of the pieces of text coded as research questions are neutral or unclear. This indicates that researchers are not researching one specific type of network, that they do not know which type of network they want to research, or that they deliberately want to leave it open to ensure that their research is not inconsistent. While eight research questions were completely neutral/unclear. Only one conclusion was totally neutral/unclear. This is not logically as the conclusions answer the research question. It can be seen that the other six neutral/unclear research questions are answered with within-person level and neutral/unclear conclusions. One neutral/unclear research question is answered by within-person level, between-person level and neutral/unclear conclusions.

Regarding the use of cross-sectional research for psychological network studies, there is a debate going on in literature (Bos & Wanders, 2016; Bringmann & Eronen, 2018; Forbes et al., 2017; Tzur-Bitan, Meiran & Shahar, 2010). Researchers state that cross-sectional research is not able to demonstrate causality and separate the within-person level from the between-person level properly. In line with this, the second finding is that there were 16 studies that used cross-sectional analysis techniques, for (partly) within-person level research questions and/or conclusions. There are only two studies for which I am sure that the methods were consistent with the level of the research questions and the conclusions. These two studies don't use cross-sectional analysis techniques but the multi-level VAR method. For two studies I cannot tell whether they were consistent since their analysis techniques were behind the scope of this study. These findings indicate a clear problem,

which is that network studies of MDD that I analyzed are mostly not examined at the level they aim to and draw conclusions about. The methods that don't match the research question, are not able to answer this research question properly. Conclusions that don't match the methods may be not valid and could be incorrectly used for other studies and for treatments and preventions of MDD. Despite this second finding of 16 inconsistent studies, the third finding is that all these studies had a cross-sectional disclaimer. This suggests that authors are to a certain extent aware of the inconsistency of their studies.

The first limitation of the study is that even though I tried to make the meta-analysis less subjective by having my supervisor also code 50% of the articles, it still is a subjective method of analysis. The codes and conclusions are probably influenced somewhere by our background and knowledge. Noémi Schuurman is an associate professor in the Department of Methods & Statistics at Utrecht University in the Faculty of Social and Behavioral Sciences. Her main research interests are idiographic modeling dynamic modeling, multi-level modeling, Bayesian modeling, networks, scientific integrity, and philosophy of science, particularly in the context of psychology. The models she is interested in including all the analysis techniques we have discussed except cross-sectional networks. This might have influenced the codes of this type of analyses techniques. I am a sociology student myself, writing my thesis at the faculty of methods and statistics. My lack of knowledge about psychology might have influenced codes and conclusions. To make our meta-analysis less subjective, we could have used statistical methods for inter-rater reliability, such as the Pearson r-correlation coefficient and Cohen's kappa. If we would have chosen to do that, we would have had to code separately and check afterwards whether we had placed approximately the same codes. We chose to not do that because I was exploratively creating a codebook. For this reason we coordinated how we were going to code things. If different coders are going to use my codebook in future research, they could perform statistical analysis to calculate their inter-rater reliability in the use of the codebook.

The second limitation is that I only looked at whether or not a study had a cross-sectional disclaimer. What matters regarding this cross-sectional disclaimer is whether the authors considered their study to be hypothetical and not ''hard facts'' because they are conducting it incorrectly. Or whether they see the use of cross-sectional methods as a minor limitation, but still consider their study consistent. This is an interesting and relevant distinction that I didn't include in the codebook. Future research could analyze the remaining 48 (since 1 study is excluded) articles of Robinaugh et al. (2020) and other psychological network studies using the codebook I developed in this study. Where the codebook currently does not make a distinction between the type of cross-sectional disclaimer, this distinction could be added.

Another limitation is that the codebook is based on empirical network studies looking at MDD. The idea of the codebook is that it can later be used to analyze other empirical studies that studied psychological networks. Given that the codebook is based only on studies with a focus on MDD, this could possibly have negative implications when used to analyze psychological network studies on other mental disorders. In the future, a modified mental-disorder-neutral codebook for empirical psychological network research might also be created.

Despite these limitations, the consensus in our results on certain trends across articles makes our conclusions credible. The results have shown that there is an inconsistency in doing empirical research on psychological MDD networks. Specifically, this study has shown that there is an inconsistency in the research questions, methods, and conclusions in the psychological network literature on MDD regarding the within-person level versus the between-person level. This is a

problem because if the research question and methods don't match, it may be difficult or not even possible, to answer the research question. If the conclusions don't match the methods these conclusions may not be valid and incorrectly used by other researchers or clinicians. The finding of this inconsistency confirms that more attention is needed for research on existing theories of MDD, as well as the development of new theories of MDD, combined with appropriate methodology for research. The need for more attention on MDD research is important because invalid conclusions can lead to incorrect adoption of these conclusions for other research, treatments, and interventions.

The inconsistency of psychological network studies analyzed in this study has important implications for both researchers and clinicians. Invalid conclusions may lead to the incorrect adoption of these conclusions for other research, treatments, and interventions. In relation to network research on mental disorders, a new standard methodology would be relevant. Instead of research that conducts a cross-sectional analysis with a cross-sectional disclaimer, the use of a multi-level VAR method would be preferable. This method is in fact capable of separating the within-person level from the between-person level. In this way, valid conclusions about the different levels can be drawn. As such, both within-person conclusions and between-person conclusions can contribute to the emerging demand for personalized treatments. With validated and standardized instruments, researchers know that they are measuring what they want to measure. Therefore, they are much less likely to overlook symptoms or individuals that need treatment. When there are consistent studies with valid conclusions about mental disorders, treatments and preventions can be better and more effectively identified, substantiated, and applied by clinicians.

References

Afzali, M. H., Sunderland, M., Teesson, M., Carragher, N., Mills, K., & Slade, T. (2017). A network approach to the comorbidity between posttraumatic stress disorder and major depressive disorder: The role of overlapping symptoms. *Journal of Affective Disorders*, *208*, 490-496.

Angst, J. (2006). Do many patients with depression suffer from bipolar disorder? *The Canadian Journal of Psychiatry*, *51*(1), 3-5.

Beard, C., Millner, A. J., Forgeard, M. J., 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.

Blaauw, F. J., van der Krieke, L., Bos, E., Emerencia, A., Jeronimus, B. F., Schenk, M., ... & De Jonge, P. (2014, November). HowNutsAreTheDutch: Personalized Feedback on a National Scale. In *AAAI Fall Symposia*.

Van Borkulo, C., 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.

Borsboom, D. (2017). A network theory of mental disorders. World psychiatry, 16(1), 5-13.

Bos, F. M., Fried, E. I., Hollon, S. D., Bringmann, L. F., Dimidjian, S., DeRubeis, R. J., & Bockting, C. L. (2018). Cross-sectional networks of depressive symptoms before and after antidepressant medication treatment. *Social psychiatry and psychiatric epidemiology*, *53*(6), 617-627.

Bos, F. M., Snippe, E., de Vos, S., Hartmann, J. A., Simons, C. J., 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.

Bos, E. H., & Wanders, R. B. (2016). Group-level symptom networks in depression. JAMA psychiatry, 73(4), 411-411.

Bringmann, L. F., & Eronen, M. I. (2018). Don't blame the model: Reconsidering the network approach to psychopathology. *Psychological Review*, *125*(4), 606.

Bringmann, L. F., Lemmens, L. H. J. M., Huibers, M. J. H., Borsboom, D., & Tuerlinckx, F. J. P. M. (2015). Revealing the dynamic network structure of the Beck Depression Inventory-II. *Psychological medicine*, *45*(4), 747.

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), e60188.

Chen, B., Liu, F., Ding, S., Ying, X., Wang, L., & Wen, Y. (2017). Gender differences in factors associated with smartphone addiction: a cross-sectional study among medical college students. *BMC psychiatry*, *17*(1), 1-9.

Choi, K. W., Batchelder, A. W., Ehlinger, P. P., Safren, S. A., & O'Cleirigh, C. (2017). Applying network analysis to psychological comorbidity and health behavior: Depression, PTSD, and sexual risk in sexual minority men with trauma histories. *Journal of consulting and clinical psychology*, *85*(12), 1158.

Dejonckheere, E., Bastian, B., Fried, E. I., Murphy, S. C., & Kuppens, P. (2017). Perceiving social pressure not to feel negative predicts depressive symptoms in daily life. *Depression and anxiety*, *34*(9), 836-844.

Elovainio, M., Hakulinen, C., Pulkki-Råback, L., Raaska, H., & Lapinleimu, H. (2018). The network structure of childhood psychopathology in international adoptees. *Journal of Child and Family Studies*, *27*(7), 2161-2170.

Forbes, M. K., Wright, A. G., Markon, K. E., & Krueger, R. F. (2017). Evidence that psychopathology symptom networks have limited replicability. *Journal of Abnormal Psychology*, *126*(7), 969.

Frewen, P. A., Schmittmann, V. D., Bringmann, L. F., & Borsboom, D. (2013). Perceived causal relations between anxiety, posttraumatic stress and depression: extension to moderation, mediation, and network analysis. *European journal of psychotraumatology*, *4*(1), 20656.

Geschwind, N., Peeters, F., Drukker, M., van Os, J., & Wichers, M. (2011). Mindfulness training increases momentary positive emotions and reward experience in adults vulnerable to depression: a randomized controlled trial. *Journal of consulting and clinical psychology*, *79*(5), 618.

Groen, R. N., Snippe, E., Bringmann, L. F., Simons, C. J., Hartmann, J. A., Bos, E. H., & Wichers, M. (2019). Capturing the risk of persisting depressive symptoms: A dynamic network investigation of patients' daily symptom experiences. *Psychiatry Research*, *271*, 640-648.

Hamaker, E. L. (2012). Why researchers should think" within-person": A paradigmatic rationale.

Hamaker, E. L., Mulder, J. D., & van IJzendoorn, M. H. (2020). Description, prediction and causation: Methodological challenges of studying child and adolescent development. *Developmental cognitive neuroscience*, 100867.

Hamaker, E. L., & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science*, *26*(1), 10-15.

Haslbeck, J., & Ryan, O. (2020). Recovering within-person dynamics from psychological time series.

Heeren, A., Jones, P. J., & McNally, R. J. (2018). Mapping network connectivity among symptoms of social anxiety and comorbid depression in people with social anxiety disorder. *Journal of Affective Disorders*, *228*, 75-82.

Hosenfeld, B., Bos, E. H., Wardenaar, K. J., Conradi, H. J., van der Maas, H. L., Visser, I., & de Jonge, P. (2015). Major depressive disorder as a nonlinear dynamic system: bimodality in the frequency distribution of depressive symptoms over time. *Bmc psychiatry*, *15*(1), 1-9.

Jaya, E. S., Hillmann, T. E., Reininger, K. M., Gollwitzer, A., & Lincoln, T. M. (2017). Loneliness and psychotic symptoms: The mediating role of depression. *Cognitive therapy and research*, *41*(1), 106-116.

Jayawickreme, N., Mootoo, C., Fountain, C., Rasmussen, A., Jayawickreme, E., & Bertuccio, R. F. (2017). Post-conflict struggles as networks of problems: A network analysis of trauma, daily stressors and psychological distress among Sri Lankan war survivors. *Social Science & Medicine*, *190*, 119-132.

Malgaroli, M., Maccallum, F., & Bonanno, G. A. (2018). Symptoms of persistent complex bereavement disorder, depression, and PTSD in a conjugally bereaved sample: a network analysis. *Psychological Medicine*, *48*(14), 2439-2448.

McNally, R. J. (2016). Can network analysis transform psychopathology?. *Behaviour Research and Therapy*, *86*, 95-104.

Merikanto, I., Lahti, T., Kronholm, E., Peltonen, M., Laatikainen, T., Vartiainen, E., ... & Partonen, T. (2013). Evening types are prone to depression. *Chronobiology international*, *30*(5), 719-725.

Molenaar, P. C. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, *2*(4), 201-218.

Robinaugh, D. J., LeBlanc, N. J., Vuletich, H. A., & McNally, R. J. (2014). Network analysis of persistent complex bereavement disorder in conjugally bereaved adults. *Journal of abnormal psychology*, *123*(3), 510.

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.

Van Rooijen, G., Isvoranu, A. M., Meijer, C. J., van Borkulo, C. D., Ruhé, H. G., & de Haan, L. (2017). A symptom network structure of the psychosis spectrum. *Schizophrenia research*, *189*, 75-83.

Schuurman, N.K. (2016). Introduction. In *Multi-level Autoregressive Modeling in Psychology: Snags and Solutions. Doctoral dissertation* (pp. 9-19). Utrecht University.

Schuurman, N.K., & Ryan, O. (2021). *On separating ''within-person variance'' from ''between-person variance''*. Manuscript in preparation.

Tzur-Bitan, D., Meiran, N., & Shahar, G. (2010). The importance of modeling comorbidity using an intra-individual, time-series approach. *Behavioral and Brain Sciences*, *33*(2-3), 172.

Wang, P. S., Aguilar-Gaxiola, S., Alonso, J., Angermeyer, M. C., Borges, G., Bromet, E. J., ... & Wells, J. E. (2007). Use of mental health services for anxiety, mood, and substance disorders in 17 countries in the WHO world mental health surveys. *The Lancet*, *370*(9590), 841-850.

World Health Organization (2020, January 30). *Depression.* Retrieved from https://www.who.int/news-room/fact-sheets/detail/Depression