Loopy Belief Propagation for Dynamic Modeling of Psychopathology Networks Using Cross-sectional Data

Jelle van Bezooijen

Clinical Psychology Unit, Institute of Psychology, Leiden University

Thesis for Research Master's Degree: Clinical and Health Psychology

Coordinated by Eiko Fried

December 4, 2020

Contents

Abstract

The network approach to psychopathology proposes that psychopathological conditions can be understood as networks of symptoms. These symptoms are thought to influence one another. However, modeling techniques for cross-sectional data produce static network models. To close the gap between theory and statistical modeling, I introduce *Loopy Belief Propagation* (LBP). This algorithm can be used to create dynamic network models using cross-sectional data by estimating how the network would behave over time. Three simulation studies were performed to investigate LBP. In study one, the convergence rate (i.e. output production rate) and the scalability (i.e. the relation between input size and run time) were investigated. Linear scaling was found, indicating that the run time should never blow up to infinity. However, substantial reductions in complexity (defined as the removal of nodes/edges) were required to reach acceptable convergence rates. In study two, the stability of LBP-based network models was investigated by checking for symmetry (defined as rough proportionality) between the static and dynamic parts of the network models. No evidence of symmetry was found, indicating that the LBP-based network models were unstable. In study three, an extension of LBP-based modeling was investigated: The similarity estimation method. This extension can be used to estimate a similarity coefficient for nested network models. The similarity estimation method was found to produce invalid results. In sum, the results were not promising, but due to certain caveats also inconclusive. More research is needed to see if LBP can estimate dynamic network models using cross-sectional data on psychopathology networks.

In network theory, constructs are conceptualized as networks of interrelated subconstructs. Accordingly, the network approach to psychopathology proposes that psychopathological conditions can be understood as networks of interrelated symptoms. If the relationships between symptoms are strong, the network can become stuck in a self-reinforcing state. A psychopathological condition is seen as an emergent property of a network in such a state (Borsboom, 2017). Note that throughout this paper, the relationship between symptoms refers to the conditional dependence relationship, which are usually estimated using partial correlations. For example, suppose insomnia and suicidal ideation are directly related to one another, and indirectly via rumination. The conditional dependence relationship would then be the direct relationship between these symptoms while controlling for the indirect relationship via rumination (Epskamp & Fried, 2018).

Networks of symptoms are commonly modeled statistically using partial correlation matrices and visually using graphs. In these graphs, symptoms are represented as *nodes* which are connected to one another via *edges*. Figure 1 gives an example of a graph. Note that in this paper, the word ‘network’ refers to theoretical networks, and the word ‘network model’ refers to statistical models. In psychopathology networks, symptoms are thought to influence one another (Borsboom, 2017). However, conventional modeling techniques for cross-sectional data, such as the Ising Model (Van Borkulo et al., 2014) or the Gaussian Graphical Model (GGM; Epskamp, 2020), produce static network models that do not capture this dynamic behavior. These static modeling techniques are often seen in the literature because most studies are based on cross-sectional research (Robinaugh, Hoekstra, Toner, & Borsboom, 2019). Techniques that allow dynamic modeling using cross-sectional data are required to bring network modeling closer to network theory. Such a technique could then be used to estimate how a network would evolve over time using cross-sectional data. To close the gap between theory and statistical modeling, I introduce an algorithm from graph theory, *Loopy Belief Propagation* (LBP), which can be used as an add-on to conventional modeling to create dynamic network models.

In this thesis, I will first introduce LBP and one of its extensions, the similarity estimation method (Koutra, Parikh, Ramdas, & Xiang, 2011a). Then, in study one, I will investigate two aspects of LBP that are important for the practical useability. In study two, I will investigate the stability of LBP-based network models for psychopathology networks. In study three, I will investigate the similarity estimation method. It should be noted that this thesis is exploratory. As such, certain parts of the analysis were performed post-hoc. This thesis is thus not meant to provide a final conclusion, but instead aims to provide an indication of whether it would be worth to continue studying the use of LBP in network-based psychopathology research.

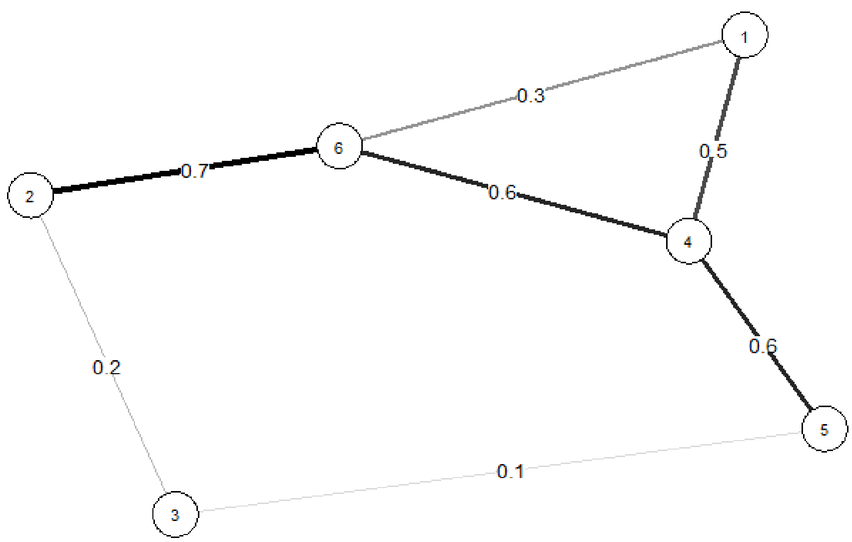

Figure 1. *Example of a graph with weighted edges, which can be seen as visual representation of a partial correlation matrix: Edges represent the conditional dependence relationships and range from -1 to 1.*

Loopy Belief Propagation

Loopy Belief Propagation (LBP) is a belief propagation (BP) algorithm (Yedidia et al., 2001) and was initially developed to estimate the marginal probabilities for nodes in a graph (Pearl, 1982; Pearl, 1986). What makes LBP useful for the current project is that the algorithm works by having nodes pass signals to one another. The network models estimated by LBP are not dynamic in the sense that they are estimated for time-series data. However, LBP can be used to estimate how network models based on cross-sectional data would evolve over time. In the following section, I give an interpretation of the technical details behind LBP following the network approach to psychopathology. For a more elaborate explanation of the mathematics behind LBP, the reader is referred to the paper by Yedidia and colleagues (2001).

Before the LBP algorithm can be used a network model needs to be estimated using a conventional method (e.g. GGM) to obtain the weighted adjacency matrix. Therefore, LBP is an add-on to conventional network modeling. After a network model has been estimated, all nodes need to be assigned at least one *prior belief*. A prior belief gives the probability for a specific node to be in a certain state. A node should be assigned as many prior beliefs as states (Yedidia et al., 2001). For the network approach to psychopathology, the prior beliefs would give the probability for a specific symptom to be in a certain state. For example, the prior belief may denote the probability for a certain symptom to be measured at the maximum level of severity. In such a case, the belief value could be interpreted as the severity of a symptom.

Just like all nodes need to be assigned prior beliefs, all edges need to be assigned an *initial message* based on the edge weight (Bayati, Borgs, Chayes, & Zecchina, 2007). Like nodes, edges have states, and so an initial message is required for every edge-state. Edges can have multiple states such that different initial messages can be used for the different states of a node. For the network approach to psychopathology, this initial message would be based on the partial correlation between symptoms. After the prior beliefs and initial messages have been assigned, each node will pass a signal to all of its direct neighbors (i.e. all nodes directly connected to said node). This first signal is based on the initial message and the prior beliefs. A new set of beliefs is then calculated based on all incoming signals (Yedidia et al., 2001). For the network approach to psychopathology, this process can be interpreted as symptoms influencing one another. For example, suppose there are two nodes, insomnia and rumination, with a weak relationship between these nodes. Suppose also that the intensity of insomnia is low and the intensity of rumination is high. Insomnia will then send a weak signal to rumination. Rumination will also send a weak signal to insomnia, albeit less weak then the signal received from insomnia because the intensity of rumination was high. A strong signal thus requires a high intensity symptom and a strong relationship between symptoms.

When the beliefs have been updated, the algorithm will calculate a new signal for each node to send to their neighbors. Based on this new signal, new beliefs will be calculated (see Figure 2). Eventually, the nodes should convergence on a single set of beliefs: The beliefs should stop changing. When this happens, LBP will produce a matrix of final beliefs (or a vector if there is one possible state per node; Yedidia et al., 2001). For the network approach to psychopathology, this can be interpreted as the psychopathological condition reaching a stable state. In this state, the severity of symptoms will remain stable unless some external factor (e.g. a major life event or treatment) influences the network. This stable state could be interpreted as analogous to the self-reinforcing state discussed by Borsboom (2017).

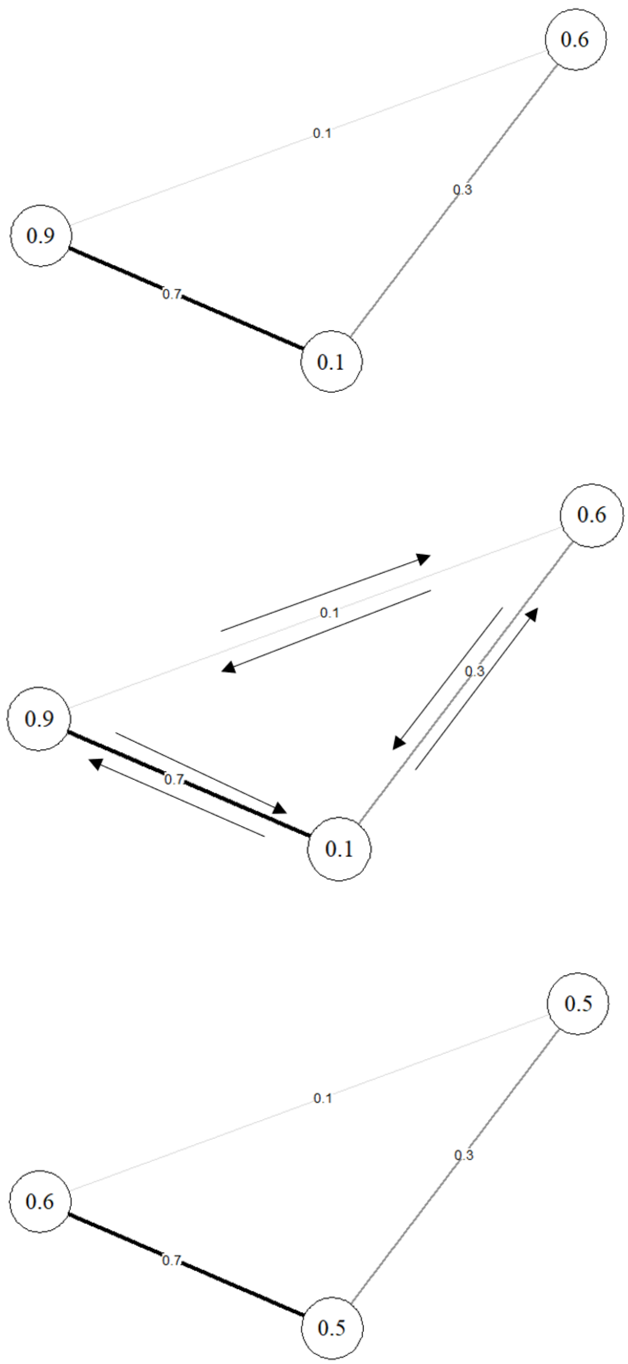

Figure 2. *An iteration of LBP visualized. The numbers in the nodes represent the belief value of a node. The numbers on the edges represent the edge weights. The arrows represent a signal being send. Note that the numbers are fictional and do not adhere to the mathematics behind LBP.*

Convergence may not occur. Instead, nodes may oscillate between two or more beliefs per state, and LBP is said to show nonconvergence. It is not entirely clear what causes nonconvergence. However, certain topological properties have been shown to influence the convergence rate. For example, if a network model is free of loops, then LBP is guaranteed to convergence (Gatterbauer et al., 2015). It is possible to force convergence using the convergence cutoff. When the degree by which a node oscillates is below some predetermined value, LBP will take the mean of the beliefs that the node is oscillating between. A higher convergence cutoff will therefore lead to less accurate final beliefs (McAuley, Caetano, & Barbosa, 2008). Despite the lack of output, nonconvergence may be meaningful. Nonconvergence could be interpreted as the psychopathological condition never reaching a stable state. The psychopathological condition may end up cycle, in which the symptoms go through a number of different states. Different versions and approximation of BP algorithms have been developed to overcome problems with convergence (e.g. Gatterbauer et al., 2015). However, LBP was used for this project because it is the most well-understood version.

Note that if node-states are theoretically defined (e.g. state defined as maximum severity level), and thus the prior beliefs are meaningful, then so are the final beliefs. In contrast, if the node-states are not defined, then neither the prior- nor final beliefs are meaningful. However, the difference between the prior and final beliefs is still meaningful because it is a function of the dynamic properties of the network. This meaningful difference can be used for various analyses, including the similarity estimation method (Koutra et al., 2011a).

The following assumptions are required for dynamic modeling of psychopathology networks using LBP. First, absolute values of edge weights can be used without loss of validity. Because LBP does not allow for negative messages, the absolute value of edge weights has to be used for the initial messages. Second, network models have the *Markov property*: the beliefs of a symptom at time t depend only on its belief at $t-1$, and the beliefs of all symptoms it is related to at $t-1$. For example, if we know the intensity of insomnia and of all other symptoms it is related to at time t, then we can predict the intensity of insomnia at $t+1$. This prediction must be within the margin of error imposed by the imperfect fit between the model and the data. Note that other BP algorithms may not require network models to have the Markov property (Yedidia et al., 2001). Third, relationships between symptoms are bidirectional in terms of the direction and size of the initial message. Fourth, the weighted adjacency matrix represents the true network. Thus, any bias in network estimation might be amplified by LBP. A proper method is therefore needed to estimate the strength of the conditional dependence relationships. Finally, symptoms influence each other via parallel duplication (Borgatti, 2005). Parallel duplication means that a symptom influences all other symptoms it is connected to, without the intensity of the symptom running out. Furthermore, the strength of a signal is not influenced by the amount of outgoing signals. For example, suppose we have a high intensity symptoms which is receiving only weak signals. The intensity of this symptom will then change only slightly due to the weak incoming signals. The fact that this symptom is sending out strong signals does not lead to a decrease in its intensity, no matter how many signals it sends. Thus, the intensity of a symptom is only influenced by the incoming signals and not the outgoing signals.

Similarity Estimation Method

Using the method developed by Koutra and colleagues (2011a), the similarity between two nested LBP-based network models can be estimated. This method could be valuable as future challenges may revolve around studying the replicability and heterogeneity of network models (Fried & Cramer, 2017).

Koutra and colleagues (2011a) employ the following definition of similarity: Two nodes are similar if their neighbors are similar. This definition can be interpreted as: A symptom in two different network models is similar if it shows a similar pattern of relationships with other symptoms. This definition is similar to the definition of topological overlap, which poses that if two nodes within a network model show a similar pattern of relationships to other nodes, then these nodes may represent the same construct (Fried & Cramer, 2017). The key difference is that topological overlap is about two nodes within the same network model, whereas similarity is about two nodes in two different network models. Nevertheless, the idea of topological overlap can be extended to the definition of similarity. Suppose we have two network models which we assume are generated by the same underlying process and for which we assume that all symptoms in the models represents the same constructs. Because a specific symptom in the two different network models is assumed to represent the same construct, one would expect that symptom to show a similar pattern of relationships to other symptoms. In other words, we would expect the symptom to show *between-network topological overlap*. The higher the percentage of symptoms that show between-network topological overlap, the higher the similarity between the two network models.

The similarity estimation method (Koutra et al., 2011a) works as follows. First, LBP is used to get a matrix of final beliefs for both network models. Then, a similarity measure (e.g. Pearson correlation) is used to produce a similarity coefficient using the two matrices of final beliefs. Note that this method of estimating a similarity coefficient satisfies the definition: A symptom is similar if it shows between-network topological overlap. If a symptom in two different network models shows similar patterns of relationships with other symptoms, then this symptom will obtain similar final beliefs. Thus, a high proportion of symptoms showing between-network topological overlap will lead to similar matrices of final beliefs, which will lead to a high similarity coefficient.

The similarity estimation method requires that the network models are nested because the final beliefs produced by LBP are compared pairwise for each node. Nodes that have the same label/position are thus assumed to represent the same symptom.

Studies

Study one.

The main objective is to investigate whether LBP can be used to create dynamic network models. Thus, LBP must be implemented into statistical software so that it can be used by researchers. Additionally, the algorithm must be of practical use. Therefore, in study one, I investigate two aspects of LBP that are important for practical use and are often suboptimal: run time and convergence rate (Gatterbauer et al., 2015). However, I investigate scalability instead of run time. Scalability is the degree by which the run time changes when the size of the input (i.e. the number of nodes/edges) is changed. Unlike run time, scalability is a property of the algorithm and generalizes to different machines. Ideally, the run time of the algorithm scales linearly with the size of the input. Linear scaling ensures that the run time cannot realistically blow up to infinity (Sanchez, Solarte, Bucheli, & Ordonez, 2018).

Study two.

In study two I investigate the stability of LBP-based network models by looking at the relationship between changes to the static part of the network model (the weighted adjacency matrix) and the resulting changes to the dynamic part of the network model (the final beliefs). The degree by which the dynamic part of the model changes should not be exactly equal to the degree by which the static part of the model was changed. Such a one-to-one relationship would mean that the dynamic part of the model does not add any extra information. However, the changes to the static and dynamic parts of the model should be roughly proportional: There should be *symmetry* between the static and dynamic parts of the model. If there is no symmetry, then processes that can cause minor changes to the static part of the model (e.g. sampling variance) can lead to major changes to the dynamic part of the model. In other words, the dynamic part of the model will be unstable. For example, in the absence of symmetry, a 1% change in the static part of the model produced by measurement error might lead to an 80% change in the dynamic part of the model. As such, completely different LBP-based network models could be estimated for the same population on consecutive days. Therefore, in the absence of symmetry, any analysis performed on LBP-based network models would be unreliable.

Study three.

In study three I investigate one of the extensions that arises when LBP is used to create dynamic network models: Estimating the similarity between two nested network models. This method was developed by Koutra and colleagues (2011a). However, there are some differences between the network models typically found in psychopathology literature and those used by Koutra and colleagues. The network models used by Koutra and colleagues showed no variation in edge weights. In terms of psychopathology network models, this would mean that all partial correlations have the exact same value, which is highly unlikely. Another important difference is that Koutra and colleagues used *Linearized BP*, which is an approximation of LBP and therefore does not produce the exact same beliefs (Koutra et al., 2011b). Thus, the results found by Koutra and colleagues (2011a) cannot be assumed to generalize to the current project.

Methods And Results

Data Collection

Empirical data were acquired for simulating network models. In addition, empirical data were acquired to test the similarity estimation method (Koutra et al., 2011a) using empirical network models. The empirical data needed to be cross-sectional data on the severity of psychopathological symptoms. In order to investigate the similarity estimation method, empirical data that could be used to fit nested network models needed to be acquired. The raw data were not required because LBP only uses the weighted adjacency matrix of a network model. Note that empirical data were only acquired such that realistic network models could be simulated. The results were not meant to be interpreted as clinically or theoretically meaningful.

In order to acquire empirical data, I first identified studies based on open datasets composed of symptom severity data of individuals suffering from a psychopathological condition. After suitable studies and open datasets were identified I contacted the lead author/researcher and/or institution to ask for permission to use the data. The resulting two datasets are described below.

Datasets

STAR*D dataset.

The Sequenced Treatment Alternatives to Relieve Depression (STAR*D; Fava et al., 2003; Rush et al., 2004) was a large-scale multisite clinical trial conducted in the USA and funded by the National Institutes of Health (NIH). The STAR*D dataset has been used before in psychometric literature (Fried, Epskamp, Nesse, Tuerlinckx, and Borsboom, 2016a; Fried et al., 2016b; Ron, Fried, & Epskamp, 2019). The STAR*D dataset contains a representative sample of depressed individuals (*n* = 4041). Participants were between the ages of 18 and 75 and were suffering from first-episode or recurrent nonpsychotic MDD. Possible participants were excluded if they had a history of- or suffered from psychosis, schizophrenia, schizoaffective disorder, bipolar disorder, primary obsessive compulsive disorder, anorexia, or bulimia. The data were collected via telephone interviews. A more detailed description of the dataset can be found elsewhere (Rush et al., 2004). For the current project, the correlation matrix for the baseline measures of the Quick Inventory of Depressive Symptoms, clinician-rated version (QIDS-C; Rush et al., 2003) was taken from the supplementary materials of the paper by Fried and colleagues (2016b). Note that Fried and colleagues (2016b) merged certain items to align with the DSM 5 criteria for MDD (American Psychiatric Association, 2013). Therefore, the dataset contained 14 instead of 16 QIDS-C items.

Posttraumatic stress disorder datasets.

For their study on the generalizability and replicability of posttraumatic stress disorder (PTSD) network models, Fried and colleagues (2018) used 4 independent samples. Sample 1 consisted of 526 Dutch individuals and refugees who had experienced complex traumatic events. In this sample, the severity of PTSD symptoms was assessed using the Harvard Trauma Questionnaire (HTQ; Mollica et al., 1992). A cut-off score of 2.5 on the HTQ led to a classification of probable PTSD for 66.7 percent of the participants in this sample. Sample 2 consisted of 365 Dutch individuals who had experienced traumatic events of various types. In this sample, the severity of PTSD symptoms was assessed using the Posttraumatic Stress Symptom Scale Self-

Report (PSS-SR; Foa, Cashman, Jaycox, & Perry, 1997). All the participants were diagnosed with PTSD based on the Structured Clinical Interview for DSM-IV (SCID-IV; Kübler, 2013). Sample 3 consisted of 926 Danish soldiers who had experienced deployment-related traumatic events. In this sample, the severity of PTSD symptoms was assessed using the PTSD Checklist, Civilian version (PCL-C; Weathers, Litz, Herman, Huska, & Keane, 1993). A cut-off score of 44 on the PCL-C led to a classification of probable PTSD for 59.3 percent of the participants in this sample. Sample 4 consisted of 956 refugees residing in Denmark. In this sample, the severity of PTSD symptoms was assessed using the HTQ (Mollica et al., 1992). All participants were diagnosed with PTSD. Fried and Colleagues (2018) pooled and rescaled items in order to be able to compare the measures. More information on the samples and the measures can be found elsewhere (Fried et al., 2018). The correlation matrices for the severity of PTSD symptoms for all samples were taken from the supplementary materials (Fried et al., 2018).

Parameters And Estimation

Network estimation.

Network models were fitted using Gaussian Markov random field estimation. Graphical Lasso regularization was used to trim any relationships between symptoms that were a function of random fluctuations. The extended Bayesian information criterion was used to tune regularization parameters (Epskamp & Fried, 2018).

LBP-based modeling.

The parameters for LBP were based on the examples found in the CRF R package (Wu, 2019). The absolute value of the weighted adjacency matrix was taken before estimating LBP-based network models. LBP-based network models were estimated using 2 states per node/edge. Initial messages were based on the edge weights from the weighted adjacency matrix by alternating between the edge weight and $1 - edge\ weight$ over edge-states. For each node, the prior beliefs were initialized as .5 and 1 for the two node states. Final beliefs were inferred using the sum-product rather than max-product equations. The maximum amount of iterations was set to 50,000. The cut-off value for forcing convergence was set to $1 * 10^{-4}$.

Similarity estimation method.

Euclidian distance was used as a similarity metric, because this metric produced the most intuitive results (Koutra et al., 2011a). The Euclidian distance (d) was converted into a similarity coefficient (s) using the equation used by Koutra and colleagues: $s = 1/(1 + d)$. The similarity coefficient was obtained by taking the average similarity coefficient over all node states. Under nonconvergence, the final beliefs tended to be biased towards state 1, causing the final beliefs to sometimes be rounded to 1. Therefore, state 1 was ignored in the calculation of the similarity coefficient under nonconvergence. Because the Euclidian distance is not scale invariant, final belief matrices with narrow ranges would have resulted in unintuitive similarity coefficients. As such, the final beliefs were standardized prior to computation of the similarity coefficient.

Software

Simulations and analyses were performed using R (version 3.5.1; R Core Team, 2018). Three R packages deserve special mention. The ‘bootnet’ package (Epskamp, Borsboom, & Fried, 2017) was used to estimate network models and simulate datasets. The ‘qgraph’ package (Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012) was used to visualize network models. The ‘CRF package (Wu, 2019) provided functions that could be used to model and perform computations on conditional random fields. With the authors permission, new functions were written based on the functions from the CRF package. These new functions were used for creating LBP-based network models and running the similarity estimation method. It should be noted that the new functions always produced the same output as the original functions when the new functions were programmed to mimic the original functions. Results were graphed using Microsoft Excel (version 365).

Study One: Scalability And Convergence Rates

Simulation, editing, and analyses.

For study 1, 3 sets of 5 datasets were simulated from the STAR*D dataset. Network models were then fitted, resulting in 3 sets of 5 network models. Every set contained 5 network models to minimize the random effects imposed by the simulation. Changes were made to the adjacency matrices in an attempt to affect the run time and convergence rate of LBP. These changes were also used by Koutra and colleagues (2011a). Specifically, the number of edges (set 1), the number of nodes (set 2), and the edge weights (set 3) were modified. Changes were made in 9 steps of 10 percent, resulting in 10 versions for each network model (0% to 90% change). LBP-based network models were then fitted for all versions. The run time was defined as the time spend inferring the final beliefs and did not include the time spend obtaining the weighted adjacency matrix through conventional modeling.

If the rate of nonconvergence was high, then the scalability analysis would not have been very informative because the run time would have been strongly related to the maximum amount of iterations. Therefore, 2 new sets of 5 simulated network models were created and reduced in complexity in an attempt to increase the convergence rate. These network models were reduced in complexity by randomly removing 6 out of the 14 nodes and 50 percent of the remaining edges. This base reduction in complexity was the largest reduction for which it was still possible to go through the 10 steps of decreasing complexity. The same changes were applied to the 8-node network models as were applied to the 14-node network models, except for the removal of nodes and associated edges. The effect of the removal of nodes and associated edges was not investigated because the 8-node network models would have gone down to 1 node. Set 2 of the 8-node network models was not analyzed because the complexity reduction did not lead to the intended result (higher convergence rates).

Results.

14-Node network models.

Convergence rates.

Very low convergence rates were observed for the 3 sets of 14-node network models. Convergence only occurred in set 1 when 90 percent of the edges were removed. Even then, the convergence rate was only at 66.67 percent. Therefore, the amount of edges had a weak effect on the convergence rate because a lot of edges needed to be removed before convergence occurred. Convergence did not occur in set 2, so the amount of nodes did not seem to have an effect on the convergence rate. In addition, convergence did not occur in set 3, so the value of edge weights did not seem to have an effect on the convergence rate. Thus, network models for psychopathology networks were likely too complex for LBP to reach acceptable convergence rates.

Scalability.

Figure 3 shows the average run times for the 3 sets 14-node network models. There was a weak negative relationship between the amount of edges and the average run time (set 1 in Figure 3). This relationship seemed to be linear, indicating good scalability. The big decrease in average run time that occurred between step 9 and 10 (see Figure 3) was caused by the increasing convergence rate (from .00% to 66.67%). If the network models that converged were excluded, the average run time was 3.56 seconds, following the established the linear trend. If the network models that did not converge were excluded, the average run time was .04 seconds. This decrease in run time upon convergence suggested that, under nonconvergence, most of the run time was spend reaching the maximum amount of iterations. There did not seem to be a relation between the average run time and the amount of nodes (set 2 in Figure 3) and the value of the edge weights (set 3 in Figure 3). In sum, the scalability of LBP-based modeling seemed good because there was either no effect on run time or a linear effect on run time.

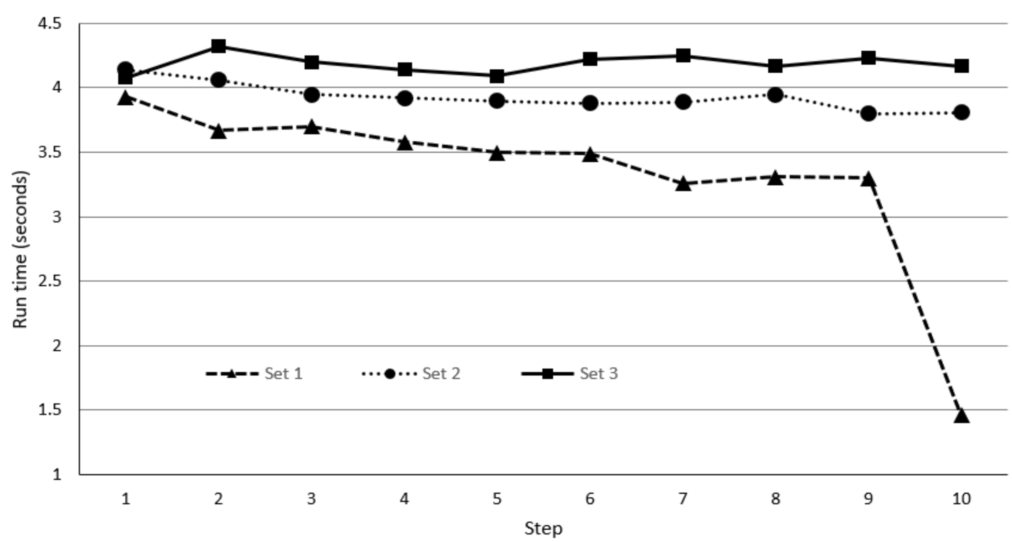

Figure 3. *Average run times for the 3 sets of 14-node network models. Every step on the x-axis represent a change of 10 percent, starting from 0 percent change at step 1.*

8-Node network models.

Convergence rates.

Table 1 gives the convergence rate per step for the set of 8-node network models, in which 10 percent of the edges were removed per step. The convergence rate improved substantially after the base complexity reduction. For the 14-node network models, convergence only occurred when 90 percent of the edges were removed. For the 8-node network models, convergence already occurred when 20 percent of the edges were removed. However, it should be noted that although complexity was an important factor, it was not the only factor. For one specific network model, convergence did not occur until 90 percent of the edges were removed, while all other network models converged when 20 to 60 percent of the edges were removed.

Scalability.

Figure 4 gives the average run times for the set of 8-node network models under nonconvergence. As found earlier, there was a linear relation between the amount of edges and run time, indicating good scalability. Apart from a dip in run time at step 8, run times seemed to decrease linearly. However, only 20 percent of network models showed nonconvergence at step 8 (see Table 1). Therefore, the dip in run time could have been an outlier caused by a small sample size.

Table 1. *Proportion converged per step for the 8-node network models*

Edges removed (%)	0	10	20	30	40	50	60	70	80	90
Proportion converged	0	0	0.4	0.4	0.4	0.6	0.8	0.8	0.8	1.0

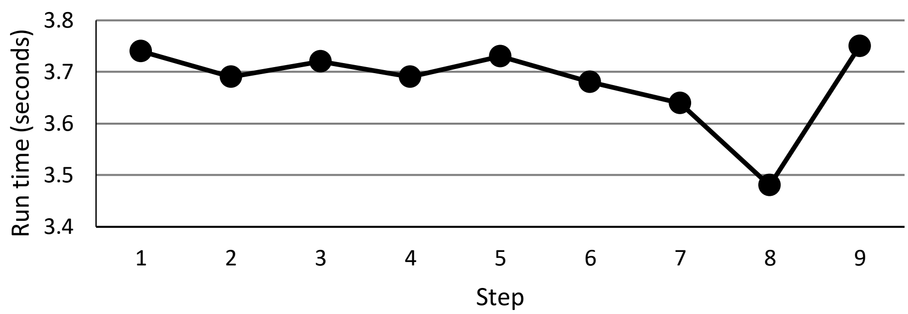

Figure 4. *Average run times for the 8-node network models under nonconvergence. Every step on the x-axis represent a change of 10 percent, starting from 0 percent change at step 1. Note that the x-axis stops at step 9 because nonconvergence was not seen at step 10.*

Figure 5 gives the average run times for the set of 8-node network models under convergence. As found earlier, the biggest factor influencing run time was whether there was convergence. Whereas run times ranged from 3.75 to 3.48 seconds for nonconvergence (see Figure 4), run times fluctuated around .015 seconds for convergence (see Figure 5). Unfortunately, it was impossible to draw conclusions about scalability under convergence because the small values of the run times allowed for a strong effect of random fluctuations.

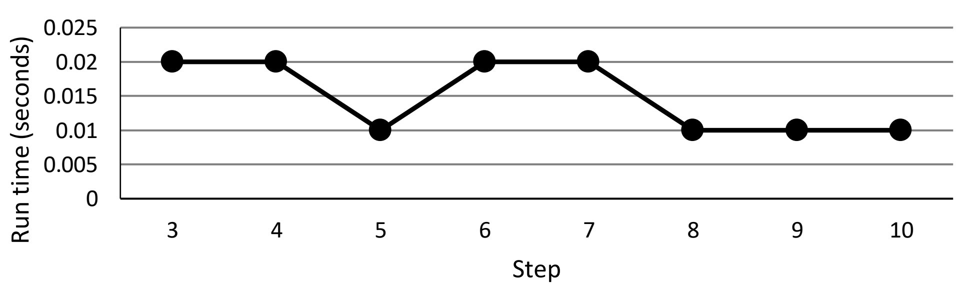

Figure 5. *Average run times for the 8-node network models under convergence. Every step on the x-axis represent a change of 10 percent, starting from 20 percent change at step 3. Note that the x-axis starts at step 3 because convergence was not seen at step 1 and 2.*

Study Two: Symmetry

Simulation, editing, and analyses.

The goal of this study was to investigate the stability of LBP-based network models. For this study, the 14-node and 8-node network models from study 1 were reused. However, set 2 of the 14-node network models, in which nodes were removed, could not be used for the symmetry analysis because the different versions of network models needed to be nested. Therefore, set 3 of the 14-node network models was renamed to set 2.

The similarity estimation method was used to compare the final beliefs for each of the 10 versions of a network model. Every version of a network model was compared to the previous version. For example, the version which had the 20 percent of the edges removed was compared to the version that had 10 percent of the edges removed. Every version was compared to the previous version and not the first version to ensure that if the first version showed nonconvergence, not all comparisons would include at least one nonconverged version. The sensitivity of unweighted adjacency matrices and Pearson's correlation over edge weights were computed for set 1. These metrics were then correlated to the similarity coefficient to see whether they indicated symmetry between the static and dynamic parts of the model. Correlations between metrics were computed separately for convergence and nonconvergence. For all metrics, 95 percent confidence intervals ($CI_{95\%}$) were computed based on the variance over different network models from a set at a specific step. The sensitivity and correlation over edge weights could not be computed for set 2 because difference between versions was that the value of the edge weights. Therefore, both the sensitivity and correlation over edge weights were always 1 and thus not informative. However, the similarity coefficient was still computed for set 2 to see if changes made to the edge weights affected the similarity coefficient. If changes to the edge weights did not affect the similarity coefficient, then this would provide evidence against symmetry between the static and dynamic parts of the network models.

Results.

14-Node network models.

Figure 6 gives the results of the symmetry analysis for set 1, in which 10 percent of the edges were removed per step. Convergence was not observed for any of the network models from this set, so only the similarity coefficient under nonconvergence could be calculated. The correlations between the nonconverged similarity coefficient and the correlation coefficient ranged between -.44 and .22, and the correlations between the nonconverged similarity coefficient and the sensitivity ranged between -.17 and .75 (see Table 1, Appendix A). Thus, there was no agreement between the metrics that quantify the changes made to the static part of the model (sensitivity and correlation over edge weights) and the metric that quantifies the resulting change in the dynamic part of the model (nonconverged similarity coefficient). In other words, there was no symmetry between the static and dynamic parts of the model under nonconvergence. This lack of symmetry suggested that, under nonconvergence, the model estimated by LBP was unstable.

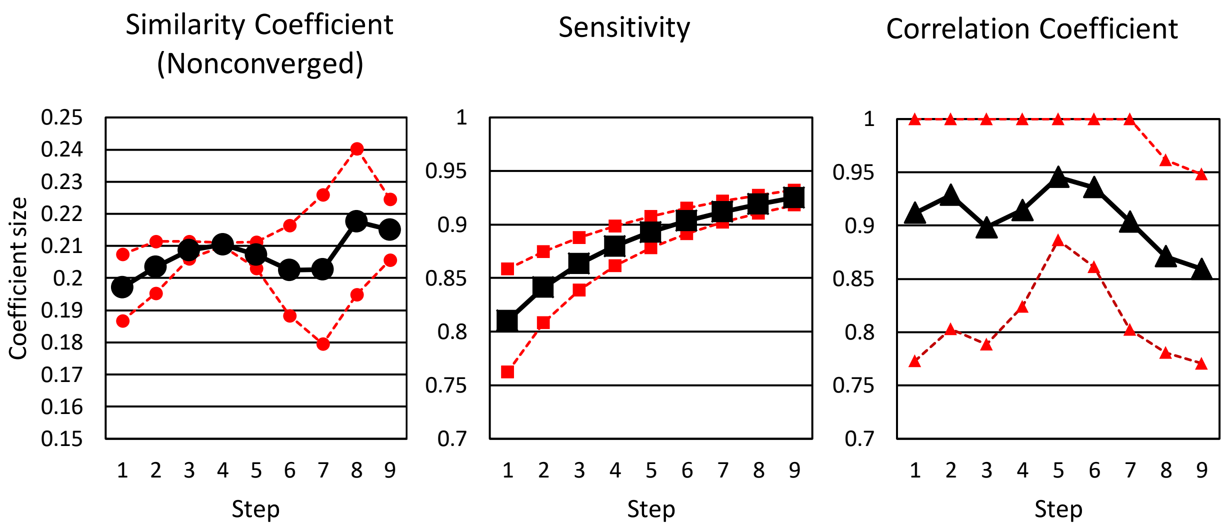

Figure 6. *Results of the symmetry analysis for set 1 of the 14-node network models. Every step on the x-axis represents a change of 10 percent, starting from 0 percent change at step 1. The dotted lines represent the upper and lower bounds of the $CI_{95\%}$. Note that if a bound hits 1 or 0 it actually exceeds it, but this was not shown in order to reduce the range of the y-axis.*

Figure 7 gives the results of the symmetry analysis for set 2, in which the edge weights were decreased by 10 percent per step. As in set 1, no convergence was observed and the nonconverged similarity coefficient fluctuated around 0.2. However, contrary to set 1, the $CI_{95\%}$ around the similarity coefficient had a comparatively constant range. This constant range suggested that the differences between the original network models were more important in determining the similarity coefficient than the changes made to the network models. The opposite pattern was seen in set 1, suggesting that changes to the structure of a network model (i.e. removing edges) had a stronger effect on the similarity coefficient than changes made to the edge weights. This lack of effect on the similarity coefficient when the edge weights were changed provided further evidence against symmetry under nonconvergence.

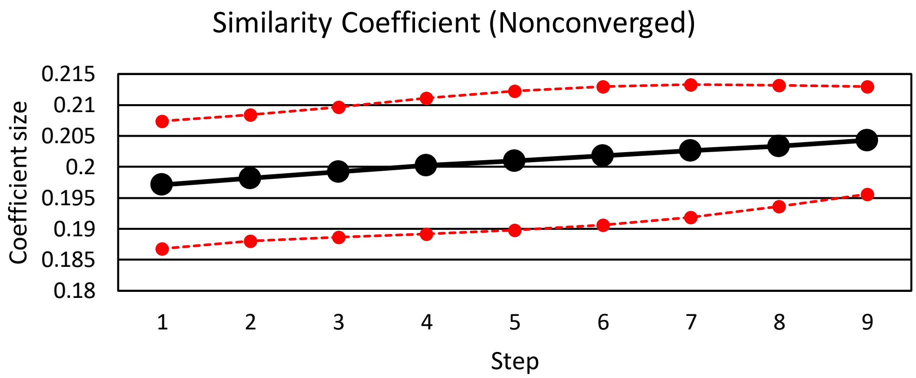

Figure 7. *Results of the symmetry analysis for set 2 of the 14-node network models. Every step on the x-axis represents a change of 10 percent, starting from 0 percent change at step 1. The dotted lines represent the upper and lower bounds of the $CI_{95\%}$.*

8-Node network models

Figure 8 gives the results of the symmetry analysis for the set of 8-node network models, in which 10 percent of the edges were removed per step. The similarity coefficient was calculated under convergence and nonconvergence. Whereas the nonconverged similarity coefficient fluctuated around .2, the converged similarity coefficient took on more plausible values. However, the correlations between the converged similarity coefficient and the correlation over edge weights ranged between -.43 and .84, and the correlations between the converged similarity coefficient and the sensitivity ranged between .29 and .86 (see Table 1, Appendix A). Thus, there was no agreement between the metrics that quantify the changes made to the static part of the model (sensitivity and correlation over edge weights) and the metric that quantifies the resulting change in the dynamic part of the model (converged similarity coefficient). In other words, there was no symmetry between the static and dynamic parts of the model under convergence. This lack of symmetry suggested that, under convergence, the model estimated by LBP was unstable.

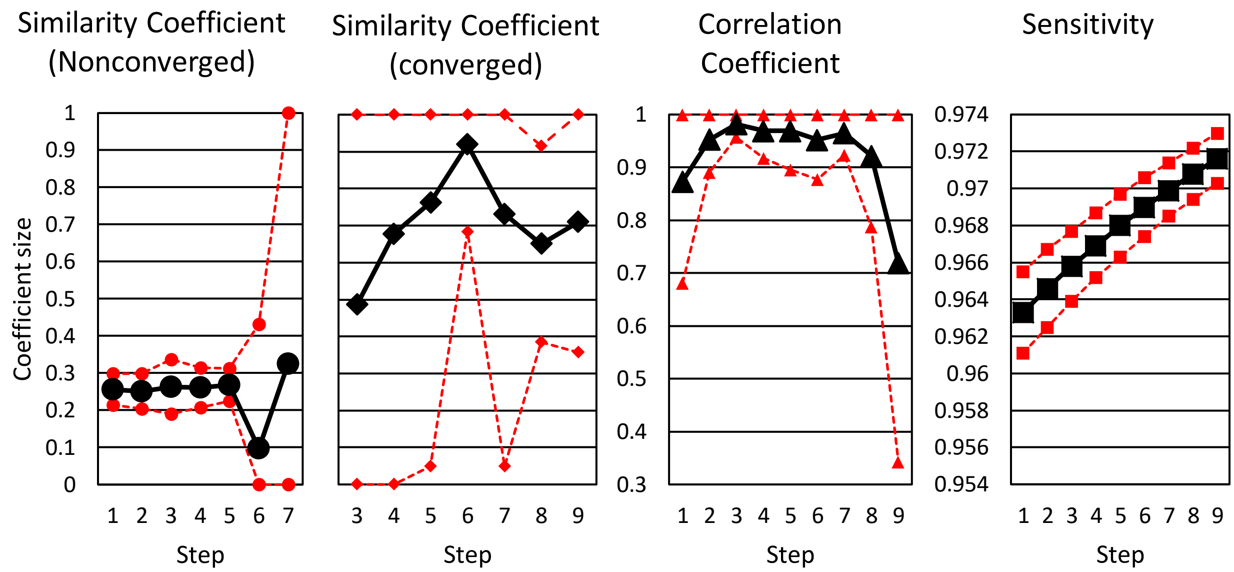

Figure 8. *Results of the symmetry analysis for the set of 8-node network models. Every step on the x-axis represents a change of 10 percent, starting from 0 percent change at step 1. The dotted lines represent the upper and lower bounds of the $CI_{95\%}$. Note that if a bound hits 1 or 0 it actually exceeds it, but this was not shown in order to reduce the range of the y-axis. Missing steps on the x-axis indicate a lack of data at these steps.*

Study Three: Similarity Estimation Method

Simulation, editing, and analyses.

Simulated data.

For study 3, 2 sets of 5 pairs of datasets were simulated from different PTSD samples. The base samples for a pair were randomly chosen without replacement, such that a pair of simulated datasets was always based on two different samples. Network models were then fitted, resulting in 2 sets of 5 pairs of network models. Changes were made to the adjacency matrices of each first network model out of a pair. The changes were made to increase the differences between the network models in a pair. Specifically, the number of edges (set 1) and the edge weights (set 2) were modified. Changes were made in 8 steps of 10 percent, resulting in 9 versions for each first network model in a pair (10% to 90% change). Note that the unchanged versions of the simulated network models were ignored because they were very similar to the empirical network models that were also used to investigate the similarity estimation method. The whole procedure was performed for the original network models (16 nodes) and the network models of reduced complexity (10 nodes). The method for creating reduced complexity network models was the same as for study 1: 6 out of the 16 nodes were randomly removed as well as 50 percent of the remaining edges. If convergence was not observed for the second 10-node network model out of a pair, the amount of edges or values of edge weights (set 1 and 2, respectively) were reduced by 10 percent until convergence was observed. The resulting version of the second network model then became the new second network model in the pair. This procedure ensured that the similarity estimation method could be tested on 2 10-node network models that both showed convergence. As in study 1 and 2, set 2 of the 10-node network models was not analyzed because the complexity reduction did not lead to the intended result (higher convergence rates).

The similarity estimation method was used to compare the final beliefs of the network models in a pair. The sensitivity of unweighted adjacency matrices and Pearson's correlation over edge weights between pairs were calculated for set 1. These metrics were then correlated to the similarity coefficient to check for agreement between metrics based on the static (sensitivity and correlation of edge weights) and dynamic (similarity coefficient) parts of the model. Correlations between metrics were computed separately for convergence and nonconvergence. For all metrics, 95 percent confidence intervals ($CI_{95\%}$) were computed based on the variance over different pairs of network models from a set at a specific step. As in study 2, the sensitivity and correlation over edge weights were not computed for set 2: These metrics would have been constant and could therefore not be correlated with the similarity coefficient. Set 2 was still analyzed to see how the similarity coefficient responded to changes made to the edge weights.

Empirical data.

In order to study the similarity estimation method using empirical data, network models were estimated for each of the 4 PTSD samples before estimating LBP-based network models. Similarity coefficients were then generated. Sensitivity of unweighted adjacency matrices and Pearson's correlation over edge weights were calculated and correlated to the similarity coefficient.

Results.

Simulated data.

16-Node network models.

Figure 9 gives the results of the similarity estimation method analysis for set 1. In this set, 10 percent of the edges were removed per step for 1 network model out of a pair. Only the similarity coefficient under nonconvergence could be calculated for this set. The sensitivity steadily increased per step and the correlation coefficient steadily decreased per step. No such obvious pattern was observed for the similarity coefficient. The similarity coefficient seemed to fluctuate around .2 with relatively low variance except for a dip at step 9 (90% of edges removed). Therefore unsurprisingly, agreement between metrics was not indicated by the correlations between the similarity coefficient and the correlation coefficient (r = .07 to r = .99) and the similarity coefficient and the sensitivity (r = -.69 to r = .86; see Table 2, Appendix B). Thus, under nonconvergence, the similarity estimation method provided seemingly invalid results.

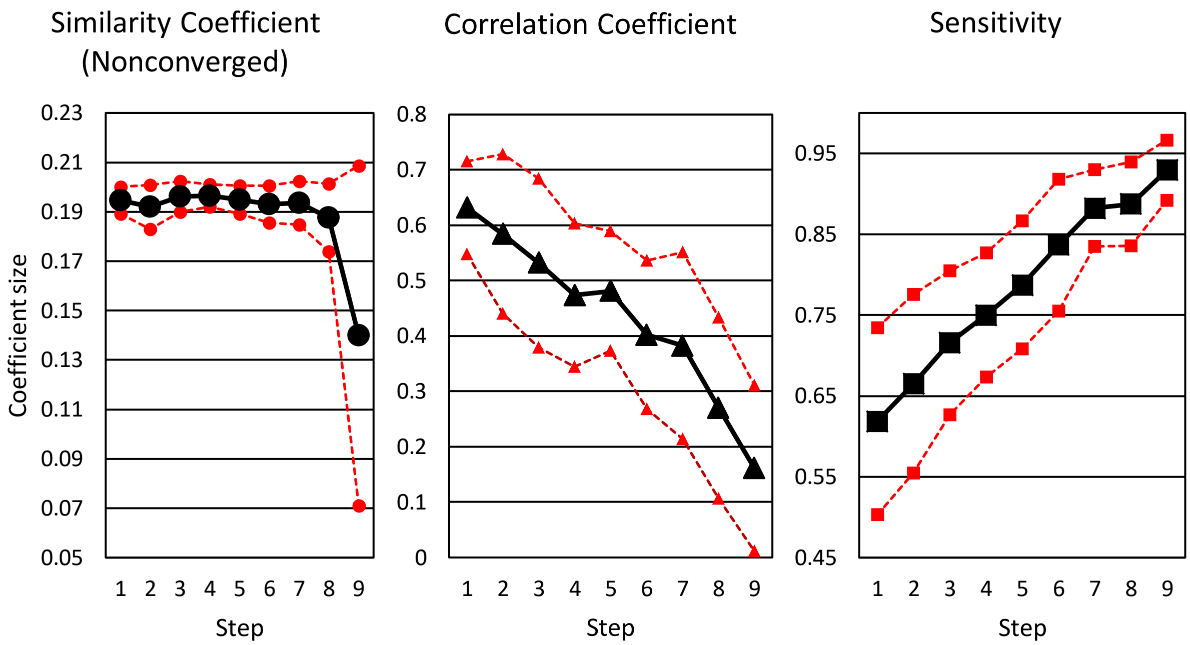

Figure 9. *Results of the similarity estimation method analysis for set 1 of the 16-node network models. Every step on the x-axis represents a change of 10 percent, starting from 10 percent change at step 1. The dotted lines represent the upper and lower bounds of the $CI_{95\%}$.*

Figure 10 gives the results of the similarity estimation method analysis for set 2, in which the edge weights were decreased by 10 percent per step for 1 network model out of a pair. Note that as with set 1, convergence was not observed, so the similarity coefficient could only be calculated under nonconvergence. The similarity coefficient changed by less than .003 between step 1 (edge weights decreased by 10%) and step 9 (edge weights decreased by 90%). As such, the nonconverged similarity coefficient did not seem to be affected by changes made to the edge weights.

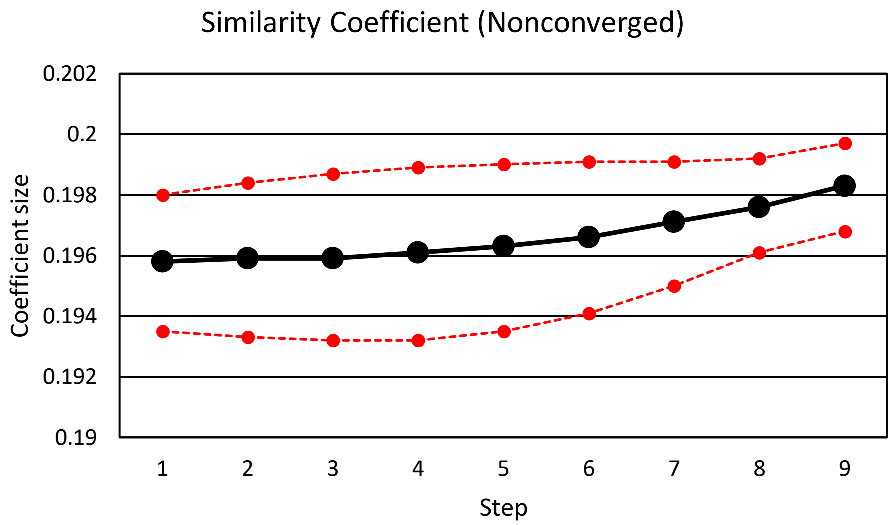

Figure 10. *Results of the similarity estimation method analysis for set 2 of the 16-node network models. Every step on the x-axis represents a change of 10 percent, starting from 10 percent change at step 1. The dotted lines represent the upper and lower bounds of the $CI_{95\%}$.*

10-Node network models

Figure 11 gives the results of the similarity estimation method analysis for the set of 10-node network models, in which 10 percent of the edges were removed per step for 1 network model out of a pair. Because instances of convergence and nonconvergence occurred, the similarity coefficient was calculated under convergence and nonconvergence. Whereas the nonconverged similarity coefficient fluctuated around .2, the converged similarity coefficient took on more plausible values. However, agreement between metrics was not strongly indicated by the correlations between the converged similarity coefficient and the correlation coefficient (r = .30 to r = .91) and the converged similarity coefficient and the sensitivity (r = .27 to r = .75; see Table 2, Appendix B). Thus, under convergence, the similarity estimation method provided seemingly invalid results.

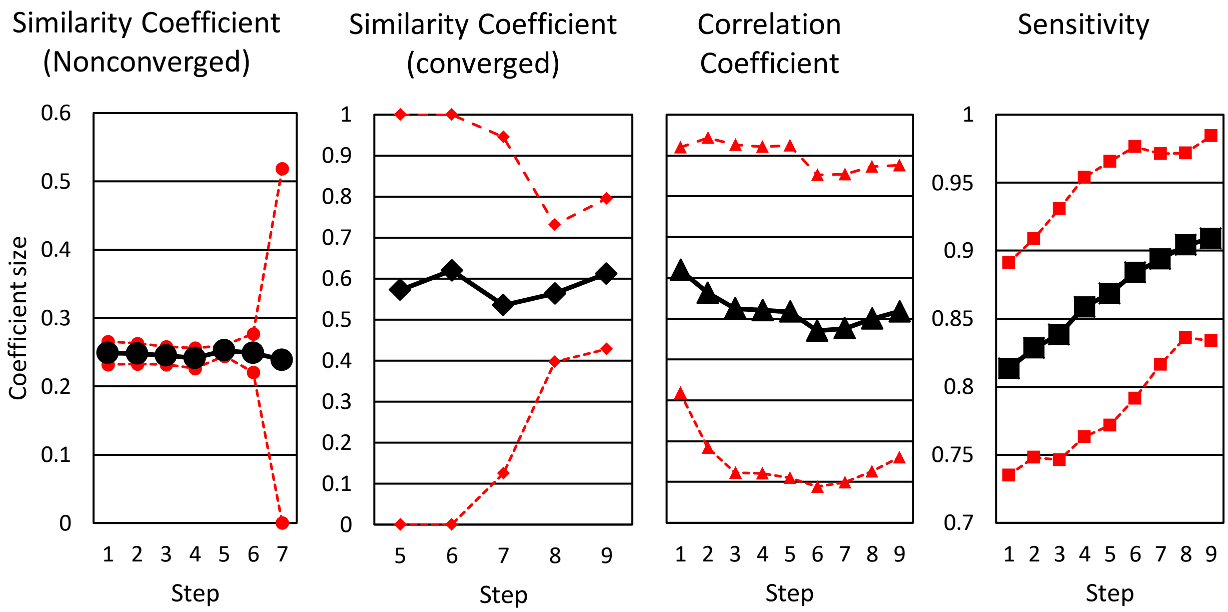

Figure 11. *Results of the similarity estimation method analysis for the 10-node network models. Every step on the x-axis represents a change of 10 percent, starting from 10 percent change at step 1. The dotted lines represent the upper and lower bounds of the $CI_{95\%}$. Note that if a bound hits 1 or 0 it actually exceeds it, but this was not shown in order to reduce the range of the y-axis. Missing steps on the x-axis indicate a lack of data at these steps.*

Empirical data.

Convergence was not seen for any of the empirical network models. As discussed earlier, under nonconvergence, the similarity estimation method provided invalid results. As such, the results for this part of the analysis were not interpreted.

Discussion

The present thesis focused on the use of LBP for creating dynamic network models using cross-sectional data on psychopathology networks. To this end, the thesis was split up into three different studies. In study one I investigated the practical usability of LBP by looking at scalability and the convergence rate. The results indicated linear scaling, meaning that the run time should not realistically blow up to infinity. Unfortunately, the convergence rate was low. Substantial reductions in complexity (i.e. removal of nodes and/or edges) were required to reach acceptable convergence rates, severely limiting the practical usability of LBP. However, it should be noted that complexity was likely not the only factor influencing the convergence rate. Certain topological features can influence the convergence rate (Gatterbauer et al., 2015). In absence of these features convergence could occur for relatively complex network models.

In study two I investigated whether there was symmetry between the static and dynamic parts of the network models. Symmetry was defined as a rough (but not perfect) proportionality between changes to the static and dynamic parts of the network models. This symmetry between the static part (based on conventional modeling) and LBP-based dynamic part of the network models is required to ensure that the LBP-based network models are stable. There was no evidence of symmetry, suggesting that LBP-based network models are unstable. This lack of symmetry means that any analysis performed on LBP-based network models, and thus also any conclusions based on these analyses, would be unreliable. For example, slight differences between network models estimated for the same population on consecutive days could lead to large differences in the LBP-based network models. Thus, any conclusions based on LBP-based network models would be unreliable as they could change on a daily basis.

In study three I investigated an extension of LBP-based network modeling: Estimating a similarity coefficient for nested network models. There was no strong indication of agreement between similarity metrics based on the static part of the network model and the similarity coefficient. Thus, the metrics based on the static part of the network models and the similarity coefficient are likely to provide conflicting results. In light of these conflicting results, the similarity estimation method should be rejected in favor of the more established metrics based on the static part of the network models.

The lack of evidence for symmetry and the lack of support for the similarity estimation method is surprising, as Koutra and colleagues (2011a) drew the opposite conclusions from their study. In the following two sections I will offer possible explanations for this difference in conclusions.

One possible explanation for the difference in conclusions comes from the inability to generalize beyond the range of inputs (Good & Hardin, 2012). For the present thesis, the range of inputs can be defined as the range of the amount of nodes and edges. Substantial reductions in complexity (i.e. removal of nodes/edges) were needed to reach acceptable convergence rates. As such, the range of inputs on which the main conclusions of this thesis are based differed strongly from the range of inputs used by Koutra and colleagues (2011a) and the intended range of inputs (network models for psychopathology networks). The results of the current thesis therefore cannot be generalized to the range of inputs used by Koutra and colleagues and the intended range of inputs. This idea is nicely illustrated by the sensitivity over nonweighted adjacency matrices for the 8-node network models of the symmetry analysis. The sensitivity seemed to be biased towards 1, as it never went below .96. However, this bias could be explained by the high number of empty cells in the 8-node network models. For the original, 14-node network models, this empty cell problem would not occur. Thus, the bias in the sensitivity cannot be generalized beyond the 8-node network models. The same argument holds for the conclusions regarding symmetry and the similarity estimation method.

Additional explanation can be identified for the difference in conclusions between this thesis and the study by Koutra and colleagues (2011a). First, LBP and/or the similarity estimation method may not be well suited for estimating/analyzing dynamic network models for psychopathology networks. There could be several reasons for this mismatch, including possibly unacceptable assumptions made by LBP. As a result of these unacceptable assumptions, LBP-based network models would not offer a valid representation of psychopathology networks. If conventional network modeling does provide a valid representation of psychopathology networks, then a lack of symmetry between conventional modeling and LBP-based network models is unsurprising. This lack of symmetry means that the analysis of LBP-based network models would be unreliable, thereby explaining the lack of support for the similarity estimation method. More research on the use of BP algorithms for dynamic modeling of psychopathology network models and a discussion about the assumptions are required. Second, perhaps LBP cannot be used, but different BP algorithms or approximations of LBP can be used. Koutra and colleagues used an approximation of LBP to increase the convergence rate. The use of this approximation leads to differences in the output and possibly also assumptions. These differences may be required for BP algorithms to be useful for dynamic modeling of psychopathology networks.

In sum, the results of this study are not promising, but also not conclusive. More research is needed to see if LBP can be used to estimate dynamic network models using cross-sectional data on psychopathology networks. If future studies continue to find similar results, a different (BP) algorithm will have to fill the gap between network theory and cross-sectional network modeling of psychopathology networks.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Bayati, M., Borgs, C., Chayes, J., & Zecchina, R. (2007). *Belief propagation for weighted b-matchings on arbitrary graphs and its relation to linear programs with integer solutions.* Technical Report, Microsoft Research.

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

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry, 16*(1), 5-13. doi:10.1002/wps.20375

Borsboom, D., Fried, E. I., Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Van Borkulo, C. D., & Van der Maas, H. L. J. (2017). False alarm? A comprehensive reanalysis of "Evidence that psychopathology symptom networks have limited replicability" by Forbes, Wright, Markon, and Krueger (2017). *Journal of Abnormal Psychology, 126*(7), 989-999. doi:10.1037/abn0000306

van Borkulo, C. D., Borsboom, D., Epskamp, S., Blanken, T. F., Boschloo, L., Schoevers, R. A., & Waldorp, L. J. (2014). A new method for constructing networks from binary data. *Scientific Reports, 4*(5918), 1–10. doi:10.1038/srep05918

Epskamp, S. (2020). Psychometric network models from time-series and panel data. *Psychometrika, 85*(1), 206-231. doi:10.1007/s11336-020-09697-3

Epskamp, S., Borsboom, D., & Fried, E. I. (2017). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods, 50*(1), 195-212. doi:10.3758/s13428-017-0862-1

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software, 48*(4), 1-18. doi:10.18637/jss.v048.i04

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

Fava, M., Rush, A. J., Trivedi, M. H., Nierenberg, A. A., Thase M. E., Sackeim, H. A. … Kupfer, D. J. (2003). Background and rationale for the sequenced treatment alternatives to relieve depression (STAR*D) study. *Psychiatric Clinics of North America, 26*, 457–494.

Fried, E. I., & Cramer, A. O. K. (2017). Moving forward: Challenges and directions for psychopathological network theory and methodology. *Perspectives on Psychological Science*. doi:10.1177/1745691617705892

Fried, E. I., Eidhof, M. B., Palic, S., Constantini, G., Huisman-van Dijk, M. H., Bockting, C. L. H., … Karstoft, K. (2018). Replicability and generalizability of posttraumatic stress disorder (PTSD) networks: A cross-cultural multisite study of PTSD symptoms in four trauma patient samples. *Clinical Psychological Science,* 1-17. doi:10.1177/2167702617745092

Fried, E. I., Epskamp, S., Nesse, R. S., Tuerlinckx, F., & Borsboom, D. (2016a). What are 'good' depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis. *Journal of Affective Disorders, 189,* 314-320. doi: 10.1016/j.jad.2015.09.005

Fried, E. I., Van Borkulo, C, D., Epskamp, S., Schoevers, R. A., Tuerlinckx, F., & Borsboom, D. (2016b). Measuring depression over time … or not? Lack of unidimensionality and longitudinal measurement invariance in four common rating scales of depression. *Psychological Assessment*. doi:1040-3590/16/$12.00

Foa, E. B., Cashman, L., Jaycox, L., & Perry, K. (1997). The validation of a self-report measure of posttraumatic stress disorder: The Posttraumatic Diagnostic Scale. *Psychological Assessment, 9*(4), 445–451. doi:10.1037/1040-3590.9.4.445

Gatterbauer, W., Günnemann, S., Koutra, D., & Faloutsos, C. (2015). Linearized and single-pass belief propagation. *Proceedings of the VLDB Endowment, 8*(5).

Good, P. I., & Hardin, J. W. (2012). *Common Errors in Statistics* (4th ed.). Hoboken, NJ: Wiley.

Koutra, D., Parikh, A., Ramdas, A., & Xiang, J. (2011a). *Algorithms for graph similarity and subgraph matching*. Pittsburgh, PA: Technical Report Carnegie-Mellon-University

Koutra, D., Ke, T., Kang, U., Chau, D. H., Pao, H. K., & Faloutsos, C. (2011b). Unifying guilt-by-associations approaches: Theorems and fast algorithms. In D. Gunopulos, T. Hofmann, D. Malerba, & M. Vazirgiannis (Eds.), *Machine Learning and Knowledge Discovery in Databases, Part II* (pp. 245-260). Berlin/Heidelberg, Germany: Springer.

Kübler U. (2013) Structured Clinical Interview for DSM-IV (SCID). In: Gellman M.D., Turner J.R. (eds) *Encyclopedia of Behavioral Medicine*. New York, NY: Springer

McAuley, J. J., Caetano, T. S., & Barbosa, M. S. (2008). Graph rigidity, cyclic belief propagation and point pattern matching. *IEEE Transactions on Patten Analysis and Machine Intelligence, 30*(11), 2047-2054.

Mollica, R., Caspi-Yavin, Y., Bollini, P., Truong, T., Tor, S., & Lavelle, J. (1992). The Harvard Trauma Questionnaire: Adapting a cross-cultural instrument for measuring torture, trauma and posttraumatic stress disorder in Iraqi refugees. *The Journal of Nervous & Mental Disease, 180*(2), 111–116.

Pearl, J. (1982). Reverend Bayes on inference engines: A distributed hierarchical approach. In American Association for Artificial Intelligence, *Proceedings: Second National Conference on Artificial Intelligence* (pp. 133-136), Pittsburgh, PA.

Pearl, J. (1986). Fusion, propagation, and structuring in belief networks. *Artificial intelligence, 29*, 241-288.

R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ron, J., Fried, E. I., & Epskamp, S. (2019). Psychological networks in clinical populations: A tutorial on the consequences of Berkson's Bias. *Preprint.* doi:10.31234/osf.io/5t8zw

Robinaugh, D. J., Hoekstra, R. H. A., Toner, E. R., & Borsboom, D. (2019). The network approach to psychopathology: A review of the literature 2008–2018 and an agenda for future research. *Psychological Medicine, 50,* 353-363. doi:10.1080/01559982.2019.1584953

Rush, A. J., Fava, M., Wisniewski, S. R., Lavori, P. W., Trivedi, M. H., Sackeim, H. A., … Niederehe, G. (2004). Sequenced treatment alternatives to relieve depression (STAR*D): Rationale and design. *Controlled Clinical Trials*, 25, 119–142.

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., … Keller, M. B. (2003). The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), Clinician rating (QIDS-C), and Self-Report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *Biological Psychiatry, 54*, 573–583. doi:10.1016/S0006-3223(02)01866-8

Sanchez, D., Solarte, O., Bucheli, V., & Ordonez, H. (2018) Evaluating the scalability of big data frameworks. *Scalable Computing: Practice and Experience, 19*(3), 301-307. doi:10.12694/scpe.v19i3.1402

Weathers, F. W., Litz, B. T., Herman, D. S., Huska, J. A., & Keane, T. M. (1993). The PTSD Checklist (PCL): Reliability, validity, and diagnostic utility. Poster session presented at the annual convention of the International Society for Traumatic Stress Studies, San Antonio, TX. Retrieved from: https://www.researchgate.net/ publication/ 291448760_The_PTSD_Checklist_PCL_Reliability_validity_and_diagnostic_utility

Wu, L. (2019). CRF: Conditional random fields. Retrieved March 28, 2019, from https://cran.r-project.org/web/packages/CRF/index.html

Yedidia, J. S., Freeman, W. T., & Weiss, Y. (2001). *Understanding Belief Propagation and its Generalizations*. Cambridge, MA, Mitsubishi Electric Research Laboratories Inc.

Appendix A

Table 1. *Symmetry analysis: Correlations of the similarity coefficient with other metrics*

		Nonconverged		Converged	
		Pearson Correlation	Sensitivity	Pearson Correlation	Sensitivity
Set 1[a]	Network 1	-.32	-.17	NA[c]	NA[c]
	Network 2	.22	.78	NA[c]	NA[c]
	Network 3	-.16	.54	NA[c]	NA[c]
	Network 4	-.24	.57	NA[c]	NA[c]
	Network 5	-.44	.75	NA[c]	NA[c]
Set rc1[b]	Network 1	.11	-.76	NA[c]	NA[c]
	Network 2	-.10	.21	NA[c]	NA[c]
	Network 3	NA[c]	NA[c]	-.43	.29
	Network 4	.20	.44	.84	.44
	Network 5	NA[c]	NA[c]	-.42	.86

[a] = Original complexity

[b] = Reduced complexity

[c] = Metric not computed because there were less than 3 datapoints or the standard deviation could not be computed

Appendix B

Table 1. *Similarity estimation method analysis: Correlations of the similarity coefficient with other metrics*

		Nonconverged		Converged	
		Pearson Correlation	Sensitivity	Pearson Correlation	Sensitivity
Set 1[a]	Network 1	.86	-.69	NA[c]	NA[c]
	Network 2	.99	.86	NA[c]	NA[c]
	Network 3	.07	.25	NA[c]	NA[c]
	Network 4	.46	.36	NA[c]	NA[c]
	Network 5	.48	.52	NA[c]	NA[c]
Set rc1[b]	Network 1	.29	.49	.30	.75
	Network 2	.43	.92	.56	.27
	Network 3	.17	.14	NA[c]	NA[c]
	Network 4	.23	.07	NA[c]	NA[c]
	Network 5	.65	.84	.91	NA[c]

[a] = Original complexity

[b] = Reduced complexity

[c] = Metric not computed because there were less than 3 datapoints or the standard deviation could not be computed