
Network Estimation and Sparsity

Abstract

Network models, in which psychopathological disorders are conceptualized as a complex interplay of psychological and biological components, have become increasingly popular in the recent psychopathological literature (Borsboom et al., 2011). These network models often contain significant numbers of unknown parameters, yet the sample sizes available in psychological research are limited. As such, general assumptions about the true network are introduced to reduce the number of free parameters. Incorporating these assumptions, however, means that the resulting network will always reflect the particular structure assumed by the estimation method—a crucial and often ignored aspect of psychopathological networks. For example, observing a sparse structure and simultaneously assuming a sparse structure does not imply that the true model is, in fact, sparse. To illustrate this point, we discuss a recently published paper that reveals a high-dimensional network of psychopathological symptoms (Boschloo et al., 2015). Furthermore, we show the effect of the assumption of sparsity in three simulation studies.

4.1 Introduction

Recent psychological literature has focused on a network approach to model many different psychological phenomena (Schmittmann et al., 2013). Such networks can be high-dimensional structures (i.e., the number of unknown parameters is much larger than the available data), which are hard to estimate without making general assumptions about the underlying true model structure. If the true model is assumed to be sparse, thus containing a small number of connections relative to the number of possible connections, a methodology can be applied that will return

This chapter has been adapted from: Epskamp, S., Kruis, J., and Marsman, M. (in press). Estimating psychopathological networks: be careful what you wish for. *PlosOne*.

such a sparse network structure. In other words, assuming a sparse network structure results in estimating a sparse network structure, which means that certain conclusions cannot be drawn from observing such a structure. In this chapter, we argue that care should be taken in interpreting the obtained network structure because the estimation procedure may pollute the results. We will illustrate this by showing examples of networks obtained when sparse networks are estimated even when the true network structure is dense.

4.2 Network Psychometrics

The network approach has been particularly promising in the field of psychopathology. Within this framework, symptoms (e.g., insomnia, fatigue, and concentration problems) are no longer treated as interchangeable indicators of some latent mental disorder (e.g., depression). Instead, symptoms play an active causal role. For example, insomnia leads to fatigue, fatigue leads to concentration problems, and so forth (Borsboom & Cramer, 2013). Psychopathological disorders, then, are not interpreted as the common cause of observed symptoms but rather as emergent behaviors that result from a complex interplay of psychological and biological components. To grasp such a complex structure, a network model can be used in which variables such as symptoms or moods are represented by nodes. Nodes are connected by edges that indicate associations between nodes. This line of research has led to intuitive new insights about various psychopathological concepts such as comorbidity (Borsboom et al., 2011; Cramer et al., 2010), the impact of life events (Cramer, Borsboom, Aggen, & Kendler, 2012; Fried et al., 2015), and sudden life transitions (e.g., sudden onset of a depressive episode; van de Leemput et al., 2014; Wichers, Groot, Psychosystems, ESM Group, & EWS Group, 2016).

The growing popularity of the network perspective on psychological phenomena has culminated in the emergence of a new branch of psychology dedicated to the estimation of network structures on psychological data—network psychometrics (Epskamp, Maris, Waldorp, & Borsboom, in press). This field focuses on tackling the problem of estimating network structures involving large numbers of parameters in high-dimensional models. When cross-sectional data are analyzed, the most popular models that are used are the Gaussian Graphical Model (GGM; Lauritzen, 1996) for continuous data and the Ising model (Ising, 1925) for binary data. Both the GGM and the Ising model fall under a general class of models called *Markov Random Fields*. These models represent variables as nodes which are connected by edges but only if the variables are *conditionally independent*. The strength of an edge (i.e., its absolute deviance from zero) demonstrates the strength of the association between two variables after conditioning on all other variables in the network; this is also termed *concentration* (Cox & Wermuth, 1993). In the GGM, edges directly correspond to partial correlation coefficients. The Ising model does not allow for such standardization, but edge weights can be similarly interpreted. A more detailed conceptual introduction of these models can be found in Chapter 1. A technical introduction to the GGM will follow in Chapter 6, and a technical introduction to the Ising Model will follow in Chapter 8.

In both models, we must estimate a weight matrix that contains $P(P - 1)/2$

number of parameters, where P is the number of nodes, in order to encode the network structure. These parameters encompass the conditional relationship between two nodes after conditioning on all other nodes in the network and can be shown to be quite instable with relatively low sample sizes (see Chapter 3). “Relatively low sample sizes,” is a loose description and has not yet been well-defined. A general rule would be to have at least as many observations as the number of parameters. But, as will be shown later, this general rule still results in unstable estimates. A common solution to overcome the problem of estimating many parameters is to reduce this number by using some form of regularization or penalization. A particularly promising technique is to apply the ‘least absolute shrinkage and selection operator’ (LASSO; Tibshirani, 1996) to the edge weights of the network. The LASSO penalizes the sum of absolute parameter values such that the estimated values shrink to zero. That is, the absolute parameter estimates will be small and will often equal exactly zero. Therefore, the resulting model is almost always sparse; only a relatively few number of parameters will be estimated to be nonzero. The use of LASSO typically leads to better performance in cross-validations (i.e., overfitting is prevented) and results in more easily interpretable models compared to nonregularized Ising models. Most important is that if the true network structure is sparse, the LASSO performs well in estimating this network structure and, more specifically, in estimating fewer edges to be nonzero that are actually zero in the true network (i.e., fewer false positives).

The LASSO uses a tuning parameter that controls the sparsity, which can be chosen to minimize some criterion such as the Extended Bayesian Information Criterion (EBIC; Chen & Chen, 2008). This methodology has been shown to work well for both the GGM (Foygel & Drton, 2010) and the Ising model (Foygel Barber & Drton, 2015; van Borkulo et al., 2014), has been implemented in easy-to-use software (e.g., Epskamp et al., 2012; van Borkulo & Epskamp, 2014), and has been utilized in an increasing number of publications (e.g., Dalege et al., 2016; Isvoranu, van Borkulo, et al., 2016; Kossakowski et al., 2016; Langley et al., 2015; Rhemtulla et al., 2016; van Borkulo et al., 2015). For a more thorough introduction to this methodology, we recommend reading van Borkulo et al. (2014) and Chapter 2. To date, the largest application of a psychological network estimated using the LASSO was carried out by Boschloo et al. (2015), who measured 120 psychiatric symptoms in 34,653 subjects and modeled these with an Ising model. We use their work in this chapter to illustrate our concerns regarding the interpretation of network structures that are the result of applying network methodology to data.

4.3 A Sparse Network Model of Psychopathology

The network of Boschloo et al. (2015) shows a network structure in which symptoms representative of a disorder strongly cluster together. Although Boschloo and her colleagues admitted that the found network structure closely represents the structure that is imposed by the Diagnostic and Statistical Manual of Mental Disorders (DSM; American Psychiatric Association, 2013), they concluded that the found structure indicates that symptoms are not interchangeable, which is presumed in the DSM. Commonly, a DSM diagnosis requires an individual to have X

out of Y symptoms, regardless of which specific symptoms. This means that two people with vastly different symptoms can be assigned the same diagnosis. This interchangeability results from an underlying causal notion of unobserved diseases causing symptoms rather than symptoms having an active causal role on each other—a notion more formally known as the common cause model (Schmittmann et al., 2013). Boschloo and her colleagues concluded that the network structure shows that symptoms are not interchangeable, mainly due to apparent differences in the number of connections and the strength of the connections between symptoms (a relative small number of pathways between disorders and the presence of some negative connections).

Although we do not necessarily disagree with the notion that symptoms play an active causal role in psychopathology, we wish to point out that the conclusion that symptoms are not interchangeable is difficult to ascertain from a sparse approximated network structure alone. This is because the LASSO relies on the assumption that the true network structure is sparse; the LASSO will always search for a model in which relatively few edges and paths explain the co-occurrence of all nodes. As a result, the LASSO can have a low sensitivity (i.e., not all true edges are detected) but always has a high specificity (i.e., fewer false positives; van Borkulo et al., 2014). It is this reason why network analysts prefer the LASSO; edges that are estimated by the LASSO are likely to represent true edges. Moreover, the LASSO returns a possible explanation of the data using only a few connections that can be interpreted as causal pathways (Lauritzen, 1996; Pearl, 2000). That the LASSO yields a possible explanation, however, does not mean that the LASSO provides the only explanation, nor does it indicate that other explanations are false. In the case of Boschloo et al., the sparse explanation found by the LASSO can give great insight regarding a possible way in which psychopathological symptoms interact with each other. However, merely finding a sparse structure does not mean that other explanations (e.g., a latent variable model with interchangeable symptoms) are disproved. Simply stated, using the LASSO always returns a sparse structure, that is what the LASSO does.

4.4 The Bet on Sparsity

The LASSO is capable of retrieving the true underlying structure but only if that true structure is sparse. Any regularization method makes the assumption that the true structure can be simplified in some way (e.g., is sparse) because otherwise too many observations are needed to estimate the network structure. This principle has been termed the *bet on sparsity* (Hastie et al., 2001). But what if the truth is not sparse, but dense?

Such a case would precisely arise if the true model were a latent common cause model in which one or several latent variables contribute to scores on completely interchangeable indicators. This is a feasible alternative because the Ising model can be shown to be mathematically equivalent to a certain type of latent variable model: the multidimensional item response model (MIRT; Reckase, 2009), with posterior normal distributions on the latent traits (Marsman, Maris, Bechger, & Glas, 2015; see also Chapter 8). The corresponding Ising model is a low-rank

network that will often be dense (i.e., all possible edges are present). Intuitively, this makes sense because the Ising model parameterizes conditional dependencies between items after conditioning on all other items, and no two items can be made conditionally independent if the common cause model is true. A low-rank weighted network will show indicators of a latent variable as clusters of nodes that are all strongly connected with each other. Therefore, if a common cause model is the true origin of the co-occurrences in the dataset, the corresponding Ising model should show the indicators to cluster together. Then if LASSO regularization is used, the corresponding network would likely feature sparsity but the nodes would still be clustered together—much like the results shown by Boschloo et al.

It is this relationship between the Ising model and MIRT that has led researchers to estimate the Ising model using a different form of regularization, by estimating a low-rank approximation of the network structure (Marsman et al., 2015). Such a structure is strikingly different than the sparse structure returned by LASSO estimation. Whereas the LASSO will always yield many edge parameters to be exactly zero, a low-rank approximation generally estimates no edge to be exactly zero. Thus a low-rank approximation will typically yield a dense network. On the other hand, this dense network is highly constrained by the eigenvector structure, leading many edge parameters to be roughly equivalent to each other rather than compared to the strongly varying edge parameters LASSO estimation allows. For example, the data can always be recoded such that a Rank 1 approximation only has positive connections. These are key points that cannot be ignored when estimating a network structure. Regardless of the true network structure that underlies the data, the LASSO will always return a sparse network structure. Similarly, a low-rank approximation will always return a dense low-rank network structure. Both methods tackle the bet on sparsity in their own way—sparsity in the number of nonzero parameters or sparsity in the number of nonzero eigenvalues—and both can lose the bet.

4.5 Estimating an Ising Model When the Truth Is Dense

Here we illustrate the effect that the estimation procedure has on the resulting Ising model in two examples. First, we simulated 1,000 observations from the true models shown in Figure 4.1. The first model is called a Curie-Weiss model (Kac, 1966), which is fully connected and in which all edges have the same strength (here set to 0.2). This network is a true Rank 1 network, which has been shown to be equivalent to a unidimensional Rasch model (Marsman et al., 2015). The Rasch model is a latent variable model in which all indicators are interchangeable. Figure 4.2 shows the results using three different estimation methods—sequential univariate logistic regressions for unregularized estimation (see Chapter 8), LASSO estimation using the *IsingFit*¹ R package (van Borkulo & Epskamp, 2014), and a Rank 2 approximation (Marsman et al., 2015)—on the first n number of rows in the simulated dataset. It can be seen that the unregularized estimation shows many spurious differences in edge strength, including many negative edges. The LASSO

¹All LASSO analyses in this chapter make use the default setup of *IsingFit*, using a hyperparameter (γ) value of 0.25 as well as the AND-rule.

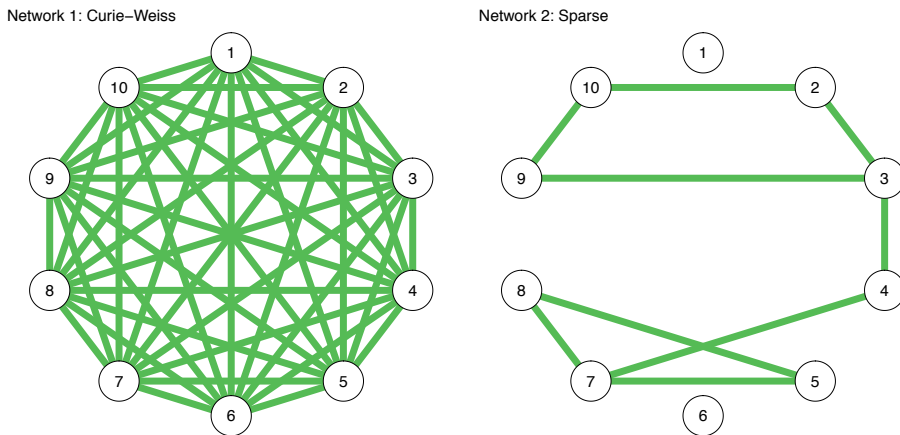


Figure 4.1: True network structures used in simulation study. The first network is a Curie-Weiss network: a fully connected network in which all edges have the same strength. The second network is a random sparse network. All edge weights are 0.2.

performs better but estimates a sparse model in which edge weights vary and in which many edges are estimated to be exactly zero. The Rank 2 approximation works best in capturing the model, which is not surprising because the true model is a Rank 1 network. At high sample sizes, all methods perform well in obtaining the true network structure.

The second model in Figure 4.1 corresponds to a sparse network in which 20% of randomly chosen edge strengths are set to 0.2 and in which the remaining edge strengths are set to 0 (indicating no edge). As Figure 4.3 shows, the LASSO now performs very well in capturing the true underlying structure. Because both the unregularized estimation and the Rank 2 approximation estimate a dense network, they have a very poor specificity (i.e., many false-positive edges). In addition, the Rank 2 approximation retains spurious connections even at high sample sizes (choosing a higher rank will lead to a better estimation). Thus, this example serves to show that the LASSO and low-rank approximations only work well when the assumptions on the true underlying model are met. In particular, using a low-rank approximation when the truth is sparse will result in many false positives, whereas using a LASSO when the truth is dense will result in many false negatives. Even when the true model is one in which every node represents an interchangeable symptom, the LASSO would still return a model in which nodes could be interpreted to not be interchangeable.

For the second example, we simulated data under the latent variable model as shown in Figure 4.4, using an MIRT model (Reckase, 2009). In this model, the symptoms for dysthymia and generalized anxiety disorder (GAD) were taken from the supplementary materials of Boschloo et al. (2015), with the exception of the GAD symptom “sleep disturbance,” which we split in two: insomnia and

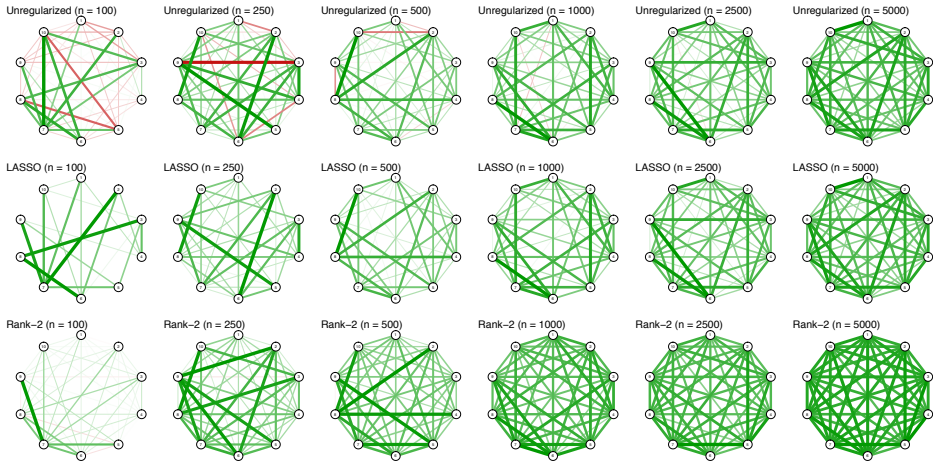


Figure 4.2: Examples of estimated network structures when the true network is a Curie-Weiss network, using different sample sizes and estimation methods. Graphs were drawn using the *qgraph* package without setting a maximum value (i.e., the strongest edge in each network has full saturation and width).

hypersomnia. The item discriminations of each symptom were set to 1, indicating that symptoms are interchangeable, and item difficulties were set to 0. All latent variables were simulated to be normally distributed with a standard deviation of 1, and the correlation between dysthymia and GAD was set to 0.55—similar to the empirically estimated comorbidity (Kessler, Chiu, Demler, & Walters, 2005). Nodes 2 and 3 in dysthymia and nodes 6 and 7 in GAD are mutually exclusive, which we modeled by adding orthogonal factors with slightly higher item discriminations of 1.1 and -1.1. Furthermore, nodes 7, 8, 9, and 10 of dysthymia are identical to nodes 6, 7, 8, and 9 of GAD respectively, which we modeled by adding orthogonal factors with item discriminations of 0.75. These nodes are typically not identical because a skip structure is imposed in datasets such as the one analyzed by Boschloo et al. That is, if someone does not exhibit the symptom “low mood,” that person is never asked about insomnia in the depression scale because he or she is assumed to not have this symptom. We did not impose a skip structure to keep the simulation study simple. Such shared symptoms are termed *bridge symptoms* in network analysis because they are assumed to connect the clusters of disorders and explain comorbidity (Borsboom et al., 2011; Cramer et al., 2010). In sum, the model shown in Figure 4.4 generates data that are plausible given the latent disease conceptualization of psychopathology.

Figure 4.5 shows the simulated and recovered network structures. First we simulated 10 million observations from this model and estimated the corresponding Ising model using nonregularized estimation by framing the Ising model as a log-linear model (Agresti, 1990, see Chapter 8) (the estimation was done using the *IsingSampler* package, Epskamp, 2014). Panel A shows the results, which give a good proxy of the true corresponding Ising structure. It can be seen that the true

4. NETWORK ESTIMATION AND SPARSITY

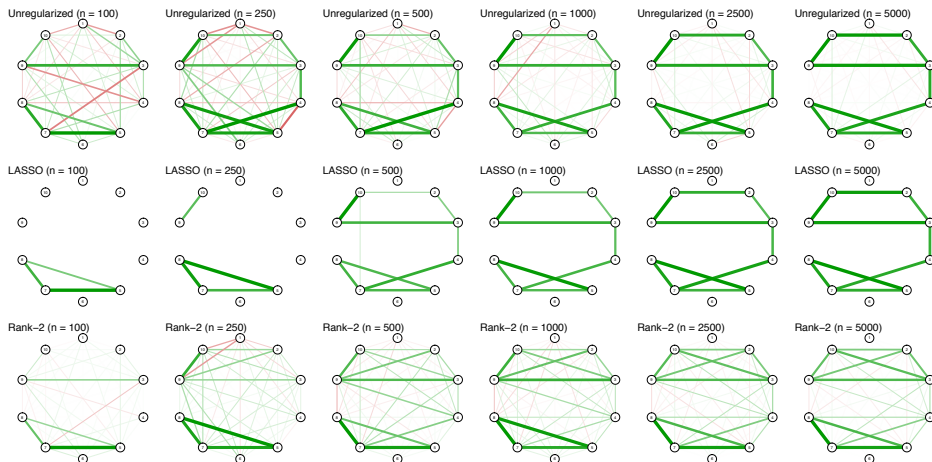


Figure 4.3: Examples of estimated network structures when the true network is sparse, using different sample sizes and estimation methods. Graphs were drawn using the *qgraph* package without setting a maximum value (i.e., the strongest edge in each network has full saturation and width).

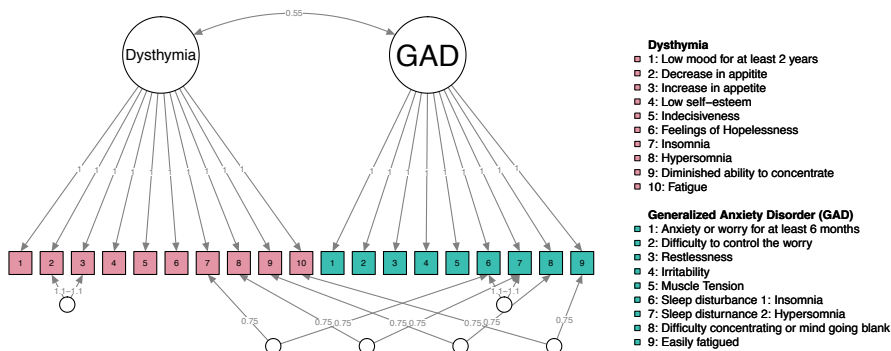


Figure 4.4: A multidimensional IRT model (MIRT) used in simulating data. All latent variables were normally distributed with standard deviation of 1 and all symptoms were binary. The edges in this model correspond to item discrimination parameters.

model is dense, meaning that indicators of the disorders cluster together. Two negative connections are formed between the mutually exclusive indicators, and bridging connections are formed between the shared indicators. Next, we simulated 1,000 observations from the model in Figure 4.4 and estimated the Ising model in various ways. Panel B shows unregularized estimation via a log-linear model and shows many spurious strong connections, including many more neg-

ative connections than present in the true model. As such, Panel B highlights our need to regularize—even in a sizable dataset of 1,000 observations for a 19-node network. The simulated data has 22.2 observations for every parameter, far more than the 4.9 observations per parameter in Boschloo et al. (2015). Thus, even with a high sample size (34,653 subjects) and even when more subjects are measured than there are parameters present, it can still be advisable to use some form of regularization, as is done by Boschloo et al. (Boschloo et al., 2015) in using the LASSO. Panel C shows the result from using the LASSO, using the *IsingFit* package (van Borkulo et al., 2014). In this model, the clustering is generally retrieved—two of the bridging connections are retrieved and one negative connection is retrieved. However, the resulting structure is much more sparse than the true model, and interpreting this structure could lead to the same conclusions determined by Boschloo and her colleagues: The number of connections differed across symptoms, connection strengths varied considerably across symptoms, and relatively few connections connected the two disorders. Finally, Panel D shows the result of a Rank 2 approximation, which is equivalent to a two-factor model. Here, it can be seen that although a dense structure is retrieved that shows the correct clustering, violations of the clustering (the negative and bridging edges) are not retrieved. The online supplementary materials² show that with a higher sample size ($n = 5,000$) the estimation is more accurate and that the unregularized and LASSO estimations result in similar network structures.

Different Estimation Techniques

In light of the examples discussed in this chapter, researchers may wonder when they should and should not use a particular estimation method. For example, low-rank estimation is more suited in the example demonstrated in Figure 4.2, whereas LASSO estimation fits better in the example shown in Figure 4.3. These conclusions, however, depend on knowing the true network structure as shown in Figure 4.1—something a researcher will not know in reality. The choice of estimation method, therefore, is not trivial. Choosing the estimation method depends on three criteria: (1) the prior expectation of the true network structure, (2) the relative importance the researcher attributes to sensitivity (discovery) and specificity (caution), and (3) the practical applicability of an estimation procedure. When a researcher expects the true network to be low rank (e.g., due to latent variables), low-rank estimation should be preferred over LASSO regularization. On the other hand, when a researcher expects the network to be sparse, LASSO regularization should be used. In addition, LASSO regularization should be preferred when a researcher aims to have high specificity (i.e., to refrain from estimating an edge that is missing in the true model). Finally, practical arguments can play a role in choosing an estimation procedure as well. LASSO, particularly in combination with EBIC model selection, is relatively fast even with respect to large datasets. As a result, researchers could apply bootstrapping methods to the estimation procedure to further investigate the accuracy of parameter estimation (see Chapter 3), which may not be feasible for slower estimation procedures.

²http://sachaepskamp.com/files/S2_EstimatedNetworks.pdf

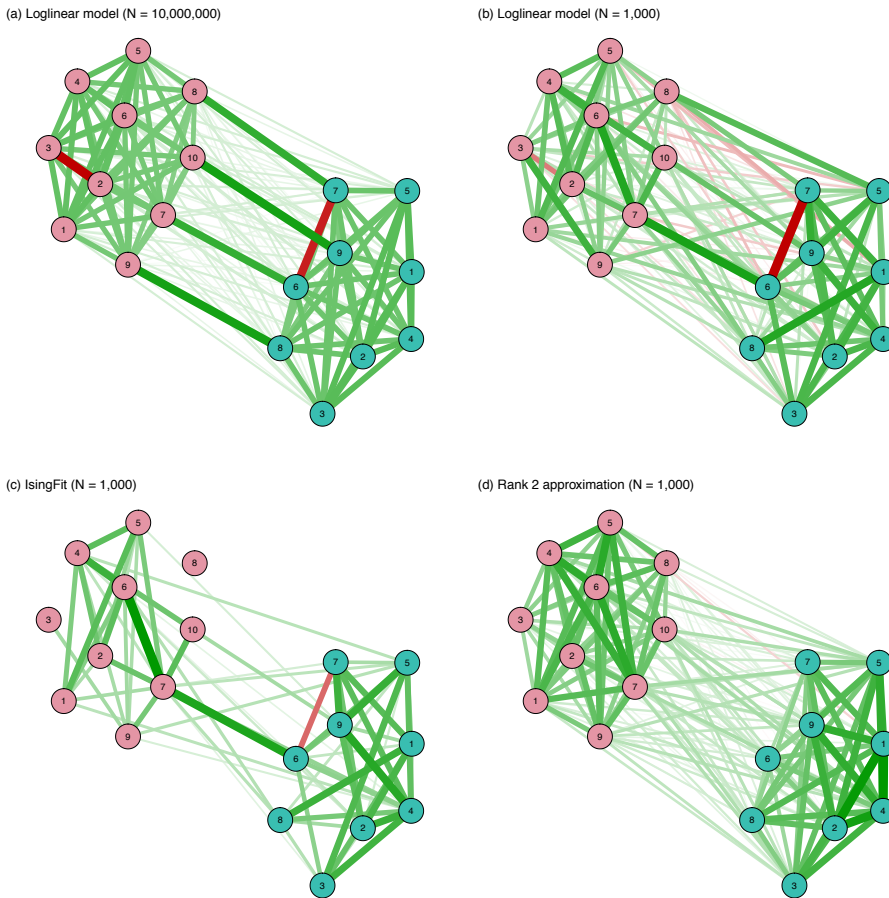


Figure 4.5: Estimated network structures based on data generated by the MIRT model in Figure 4.4.

We focused the argumentation on LASSO regularization and low-rank approximation because these are the main methodologies that have been applied in psychological literature and present two extreme cases of a range of different network structures that can be estimated. Because these methods lie on the extreme ends of sparsity relative to dense networks, they best exemplify the main point of this chapter: In small sample sizes, some assumptions of the true model must be made (e.g., the true model is sparse), and these assumptions influence the resulting network structure (e.g., the obtained network is sparse). This does not mean that LASSO and low-rank approximation are the only methods available. An alternative, for example, is to use *elastic-net* estimation, which mixes LASSO regularization with ridge regression (penalizing the sum of squared coefficients). The *elasticIsing* package (Epskamp, 2016) can be used to accomplish this; it uses

cross-validation in selecting the tuning parameters. The online supplementary materials show an example of elastic-net applied to the data analyzed in Figure 4.5. It is noteworthy that the elastic-net procedure selected a dense network (i.e., ridge regression) over LASSO regularization, indicating that data-driven evidence can be garnered to argue whether or not LASSO regularization should be used. The obtained network, like the unregularized network in Figure 4.5 (Panel B), also shows many connections which were falsely estimated to be negative; this raises the question of whether its result should or should not be preferred over LASSO regularized estimation. The online supplementary materials also contain examples of LASSO regularization using different tuning arguments (e.g., BIC selection instead of EBIC selection), which improves sensitivity (i.e., more edges are detected) in this particular case. Doing so, however, will result in less specificity when the true model is sparse (van Borkulo et al., 2014).

4.6 Conclusion

Network estimation has grown increasingly popular in psychopathological research. The estimation of network structures, such as the Ising model, is a complicated problem due to the fast growing number of parameters to be estimated. As a result, the sample size typically used in psychological science may be insufficient to capture the true underlying model. Although a large sample size network estimation typically goes well regardless of the estimation method used, Figures 4.2, 4.3, and 4.5 show that estimating an Ising model with sample sizes commonly used in psychological research results in poor estimates without the use of some form of constraint on the parameter space. Two such constraints involve limiting the size and number of nonzero parameters (LASSO) or reducing the rank of a network (low-rank approximation). It is important to realize that using such estimation methods makes an assumption on the underlying true model structure: The LASSO assumes a sparse structure whereas low-rank approximation assumes a dense but low-rank structure. Investigating the results of the estimation methods cannot validate these assumptions. The LASSO always yields a sparse structure, which does not mean that the true underlying structure could not have been dense. On the other hand, low-rank approximations rarely produce sparse structures, but that does not mean that the true underlying structure could not have been sparse.

Figure 4.2 illustrates this point by showing that LASSO estimation when the true network structure is a Curie-Weiss model still results in a sparse structure. This means that observing any of the sparse structures shown in Figure 4.2 does not mean that the nodes in the network could not represent interchangeable indicators of a single latent trait. Figure 4.5 illustrates this point again in a plausible scenario in psychopathology and also shows that when the true network structure is complicated and neither sparse nor low rank, as is the case here, all regularization methods partly fail even when using a relatively large sample size. As such, interpreting the sparsity of such a structure is questionable; the LASSO resulting in a sparse model gives us little evidence for the true model being sparse because a low-rank approximation returning a dense model seems to indicate that the true model is dense. Those characteristics from the networks we obtain are a

consequence of the method used to estimate a network structure (specifically the assumptions made by the employed method about the data-generating network structure) and often pollute the resulting estimated model (Kruis & Maris, 2015).

Recently it has been demonstrated that three, statistically indistinguishable, representations of the Ising model exist that explain observed associations between binary variables either through a common cause (latent variable), through the reciprocal effect between variables (network), or through the conditioning on a common effect (collider variable; Epskamp et al., in press; Marsman et al., 2015; Kruis & Maris, 2016). Consequently, when a model from one of these frameworks can sufficiently describe the associative structure of the measured variables, there exists an alternative representation for other frameworks that can also accurately represent the structure of the data. For example, Boschloo et al.’s sparse network structure (Boschloo et al., 2015), resulting from the LASSO being applied to the data, can also be described by a multidimensional latent variable model (with a single latent variable for each clique in the network) and residual correlations. As such, obtaining sufficient fit for a statistical network model cannot be regarded as evidence for the theoretical model, where a network structure acts as the causal mechanism from which associations between variables emerge. We therefore advise, in general, to tread carefully when drawing inferences about the theoretical causal mechanisms that generate the data from statistical model fit.

Network models show great promise in mapping out and visualizing relationships present in the data and are useful to comprehend high dimensional multivariate relationships. In addition, network models can be powerful tools to estimate the backbones of potential causal relationships—if those relationships are assumed to exist. Using the LASSO to estimate such network structures is a powerful tool in performing fast high-dimensional model selection that results in fewer false positives, and interpreting network structures obtained from the LASSO can illuminate the strong relationships present in the dataset. Important to realize is that using LASSO estimation will result in a sparse structure, and similarly, using a low-rank approximation will result in a dense low-rank result. Our aim here is not to argue against using the LASSO or to argue that estimating network structures is wrong. Our aim is to clarify that choosing the estimation method is not trivial and can greatly impact both the estimated structure as well as any conclusions drawn from that structure.