Chapter 12

Discussion: The Road Ahead

12.1 Introduction

This dissertation provided an overview of network models applicable to psychological data as well as descriptions of how these methods relate to general psychometrics. The visualization methods outlined in the final part of this dissertation are based on the oldest publications and relate to the state-of-the-art when this PhD project started. At the start of this PhD project, 4 years ago, network estimation in psychology consisted of not much more than drawing networks based on marginal correlation coefficients. This can be shown in publications from this period. Cramer et al. (2010) marks the first psychological network estimated from data and shows a network in which edges are based on associations. The *qgraph* package was based on this and, for the first time, provided psychologists with a simple method for constructing networks based on correlations (Epskamp et al., 2012). Key publications of that time mostly outlined conceptual and theoretical implications of the network perspective and often relied on correlation networks to showcase what such a network could possibly look like (e.g., Cramer, Sluis, et al., 2012; Borsboom & Cramer, 2013; Schmittmann et al., 2013). Partial correlation networks were proposed and published (e.g., Epskamp et al., 2012; Cramer, Sluis, et al., 2012) but were not yet worked out in enough detail to provide the powerful visualizations now used in psychology.[1] In addition, time-series models showed promise (e.g., Borsboom & Cramer, 2013) but had not yet been worked out in detail and implemented in easy-to-use software.

The use of network estimation on psychological data has come a long way since then. In fact, the achieved progress warrants the birth of a new field of research: *network psychometrics*. This progress has been marked by a gradual increase in the understanding of both the interpretation and the applicability of network models as well as by key turning points in the development of the methodology. Some of these key turning points came with the emergence of new software routines that

[1] In retrospect, the original promise of partial correlation networks might have been taken too strong. For example, we now know that the partial correlation network shown by Epskamp et al. (2012) consists of far too many nodes compared to the number of observations to likely lead to stable results.

make network estimation accessible to psychological researchers. In particular, the development of the *IsingFit* package (van Borkulo et al., 2014) and new versions of the *qgraph* package[2] changed network psychometrics from a conceptual framework to a concrete methodology that psychologists could readily apply. More recently, network psychometrics has further matured by the development of software packages that include data-driven statistical procedures which assess the properties of the estimated networks, such as comparing network structures of different samples (*NetworkComparisonTest*; van Borkulo, 2016), and assess the accuracy and differences in network properties, such as centrality indices (*bootnet*; Epskamp, Borsboom, & Fried, 2016; see Chapter 3). In addition, promising new software packages became available that allow for network estimation on time-series data, on multiple subjects (*mlVAR*; see Chapter 6), in clinical practice (*graphicalVAR*; see Chapter 5 for an example), and in datasets using variables of different distributions without the assumption of stationarity (*mgm*; Haslbeck & Waldorp, 2016a). Finally, the *lvnet* package (Epskamp, Rhemtulla, & Borsboom, 2016; see Chapter 7) marks the first software package that combines undirected network models with latent variable modeling.

We have come a long way, but there is still a long road ahead. As more and more technical details and conceptual interpretation of these models are worked out, more and more questions emerge. Network psychometrics is now being fleshed out as its own field of research—a field that many talented researchers are entering. Therefore, I wish to conclude this dissertation with an overview of open questions and potential future directions for this young field of science.

12.2 Open Questions in Network Psychometrics

Handling Missing Data and Ordinal Data

The methods used in network psychometrics mostly come from the fields of statistical learning, statistical physics, and econometrics. Data from such fields are very different from data typically found in psychology. Two properties of psychological data especially do not often occur in other fields of science: data with missing values and data on an ordinal scale of measurement. Network estimation can learn from a long history of handling such problems in psychometrics. As such, network psychometrics should focus on the two problems noted above. Both problems are far from trivial and will require substantive future research.

Missing data. In psychology, missing data is usually not the exception but rather the norm. Network estimation methods, however, are not yet capable of handling missing data in an appropriate way. When estimating a GGM, an estimate of the variance–covariance matrix is used as input to the graphical LASSO (see Chapter 2). When data are missing, such an estimate can be obtained by deleting cases containing one missing value or by pairwise estimation. Typically,

[2] Version 1.2.5 (revamping the choice of cutoff selection in visualizing networks and introducing standardized centrality plots) and version 1.3 (introducing EBIC model selection of glasso networks; Foygel & Drton, 2010).

pairwise estimation is used; however, this relies on certain assumptions on why the data are missing (i.e., are the data missing at random or not) and might result in variance–covariance matrices that are not positive definite. In addition, it is questionable what the sample size (e.g., for EBIC model selection) is when the variance–covariance matrix is pairwise estimated. When estimating the Ising model, mixed graphical models, or time-series models, researchers often delete full cases as a method for handling missing data, but they lose a significant amount of information in the process.

Psychometricians have worked out in detail many ways of handling missing data in various modeling frameworks (Enders, 2001). Powerful methods involve (multiple) imputation techniques and full-information maximum likelihood (FIML) estimation. Such methods could, in theory, also be applied to network estimation but further research is needed. Klaiber, Epskamp, and van der Maas (2015) proposed imputation techniques to estimate the Ising model iteratively. First the model is fit to the data, then the data are imputed given the Ising model, then the model is fit to the imputed data, and so on until the parameter estimates are stable. In theory, FIML is possible for estimating the GGM (the GGM can be framed in terms of a typical SEM model; see Chapter 7), but this could only work for confirmatory models. Usually, regularization techniques such as the LASSO are applied in the estimation. Perhaps a penalized version of FIML can be worked out in future research, combining the strengths of FIML with LASSO estimation.

Ordinal data. Another well-known problem in psychometrics is the scale of measurement on which items are assessed (Stevens, 1946). Researchers seek to measure concepts that are not directly observable, such as the severity of a person's rumination, using psychological items. Such items are frequently measured on Likert scales and cannot readily be treated as continuous (Rhemtulla, Brosseau-Liard, & Savalei, 2012). This problem is especially prominent in data on psychopathological symptoms, often measured on a 4-point scale (e.g., Fried, van Borkulo, et al., 2016), ranging from 0 (*not present*) to 3 (*severe problems*). Often, these data are highly skewed (i.e., many people report 0, especially when a general population sample is used).

Although network psychometrics is often applied to ordinal data, the handling of such data should also be a topic of future research. Currently, no method of appropriately handling ordinal data exists. There are four methods often applied to handle such data, all of which can be problematic:

1. The method most commonly used is to compute polychoric correlations and to use these as input to the EBIC model selection of GGM networks using the graphical LASSO (see Chapter 2). This methodology, however, is not without problems. First, researchers employ the methodology to estimate the model in two steps, first by computing the polychoric variance–covariance matrix and next by treating this as the sample variance–covariance matrix of continuous variables in computing the likelihood. Even though simulation studies show that this works well, it is not the most appropriate way of handling such data (e.g., in SEM, the thresholds of the polychoric correlations are estimated at the same time as the SEM model). Second, this method-

ology assumes an underlying normally distributed variable, which might be problematic because zero usually means the absence of any symptoms (a strict boundary). Third, polychoric correlations can lead to strange results (see Chapter 2 for an overview) when pairwise marginal crosstabulations of items contain zeroes, which could be expected in highly skewed ordinal data.

2. Data can be dichotomized, and the Ising model can be computed. Although setting the cutoff between 0 and 1 seems appropriate and is defensible, doing so will lose information on the severity of items.

3. Mixed graphical models can be used, in which case the variables are treated as categorical. This method takes all responses into account but loses information pertaining to the order of responses (e.g., 3 is higher than 2, and 2 is higher than 1) and instead treats each response as a categorical outcome.

4. Ordinal data can be ignored and treated as continuous. This method is not recommended because simulation studies have shown that doing so has a lower sensitivity than when using polychoric correlations and also features an inflated Type 1 error rate when statistically comparing centrality indices.

Future researchers should focus on better estimation methods for graphical models on ordinal data. Such estimation methods will likely come from psychometrics because ordinal data has long been handled in many ways. Because the GGM can be included in the SEM framework (see Chapter 7), handling ordinal data in the same manner as in SEM (e.g., by using weighted least squares estimation; Muthén, 1984) seems a logical first step. However, extending such methodology to include high-dimensional model selection will be challenging.

Evidence for Sparsity

As strongly argued in Chapter 4, using the LASSO estimation leads to sparsity (edge weights of zero) in the corresponding network model. As such, observing zeroes is not evidence that the true network is sparse. The same is true when edges are thresholded for significance (as in Chapter 6) or when step-wise model search is used (as in Chapter 7). The goal of these methods is to maximize *specificity* (see Chapter 2). Closely related to null hypothesis testing: removing an edge is not evidence that the edge weight is zero (i.e., the null-hypothesis being true); an edge might also be removed because the data are too noisy. Classical tests, LASSO regularization, and frequentist model search cannot differentiate between noisy data and the null-hypothesis being true (Wagenmakers, 2007).

The question whether a missing edge is due to the null hypothesis being true, however, is a very important one. An edge weight of *zero* in a pairwise Markov random field, such as the GGM or the Ising model, indicates that two variables are *conditionally independent*. This is important for two reasons. First, as already outlined in this dissertation, conditional independence plays a crucial role in causality (Pearl, 2000). For example, the causal structure $A \rightarrow B \rightarrow C$ implies that A and C are conditionally independent given B. Second, when the latent common cause model is true, no two variables should be conditionally independent given any other variable in the dataset. Conceptually, this implies that the only variable on which one could condition to make observed variables independent is the latent variable. Network models only show conditional associations after conditioning on observed variables. As such, when we find strong evidence that several pairs of variables become conditionally independent given a third, the common cause model will not be true.

In network psychometrics, we are interested in finding *conditional independence* in addition to finding strong conditional *dependencies*. However, the methods used today only allow for the latter. Future researchers should aim to develop methods in which both can be found. That is, for every edge, we should want to know the evidence for that edge existing (strong relationship) and the evidence for that edge not existing (conditional independence). This is difficult to accomplish in the frequentist framework, typically used in network psychometrics, but it is possible in a *Bayesian* framework. In recent years, Bayesian analysts have worked out the Bayes factor (Kass & Raftery, 1995; Ly, Verhagen, & Wagenmakers, 2016) as a default method for quantifying both the evidence for the null and for the alternative hypothesis. Such Bayes factors can possibly be computed for every edge in the graph, allowing a researcher to identify which edges are likely present, which edges are likely zero, and the edges whose data are too noisy to make such a distinction. The Bayes factors for partial correlations have been worked out (Wetzels & Wagenmakers, 2012). Node-wise estimation of graphical models could possibly also be used to obtain two Bayes factors per edge (Gelman, Jakulin, Pittau, & Su, 2008), using regular regression for the GGM and logistic regression for the Ising model. Finally, the work of Mulder (2014) on testing constraints on correlation matrices could possibly be extended to testing constraints on partial correlation matrices. An additional challenge will be to combine such methods with high-dimensional model selection, such as the LASSO, for which the Bayesian LASSO could possibly be used (Park & Casella, 2008).

Should Graph Theory Be Used to Analyze Probabilistic Graphical Models?

In this dissertation I have presented several methods for estimating network structures on psychological data. As nodes represent variables and edges are typically unknown, all of these models belong to a class known as *probabilistic graphical models* (Koller & Friedman, 2009; Lauritzen, 1996). These models aim to characterize the joint likelihood of observed variables, and allow for results to be represented through networks. Although graphically depicting these models as networks is a powerful technique for communicating such high-dimensional analysis results, it is questionable if measures from graph theory, such as centrality indices, could be readily applied to the networks estimated this way. Meaning, can probabilistic graphical models be interpreted in the same way as, say, a railroad network?

As described several times in this dissertation, typical methodology for analyzing weighted networks—such as computing centrality measures (Opsahl et al., 2010)—are often used on models obtained in network psychometrics. In this methodology, each edge is first transformed into a *length* (i.e., the inverse absolute value of an edge weights), then the resulting network is analyzed as, for example, a railroad network would. The distance between two nodes in a network is

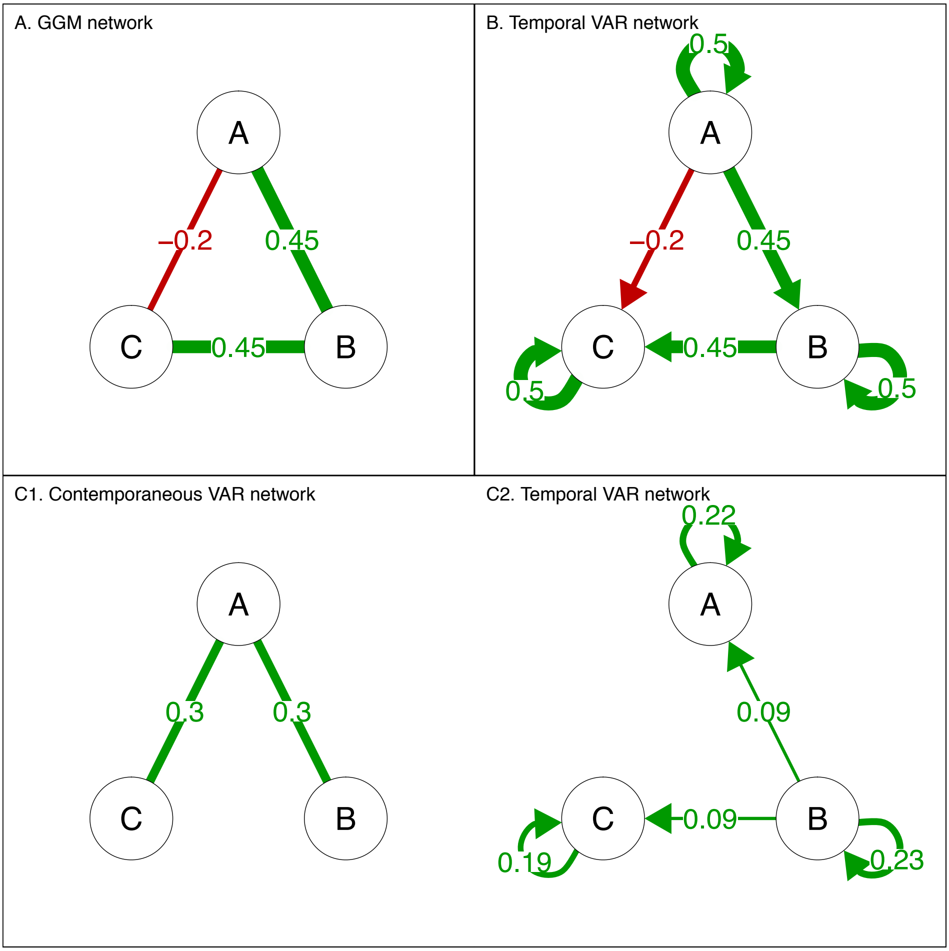

Figure 12.1: Three hypothetical graphical models for which computing network descriptives might be problematic. Panel A shows a Gaussian graphical model (a network of partial correlation coefficients), Panel B shows the temporal structure of a VAR analysis and Panel C shows both the contemporaneous and temporal structure of a VAR analysis.

defined through computing *shortest paths*, an important property in computing both closeness node centrality (i.e., the inverse sum of such distances) and the betweenness node centrality (i.e., measuring how often a node lies on the shortest path). Such methods are insightful when researchers can interpret these shortest paths to be sensible (e.g., passengers using a rail network will probably only travel through the shortest path), but it might make less sense in the context of probabilistic graphical models.

Figure 12.1 illustrates some examples of where interpreting such structures as networks can go wrong. In Panel A, a three-node GGM is depicted. When investigating the distance between nodes A and C, a typical network analysis would indicate that the *shortest path* is the path $A - B - C$, resulting in B having the highest *betweenness* and the direct path between A and C being ignored. However, such a result would not take all the information of the model into account. In fact, A and C are marginally independent; the correlation between A and C is exactly zero, indicating that knowing A contains no information on C and vise versa. Such a structure could emerge if B is a common effect of A and C, in which case disturbing A can, in no way, have any effect on C. As such, it is questionable what it means for B to have a high betweenness if no causal effect goes through B.

Panel B shows a vector auto-regression (VAR) temporal model in which the contemporaneous structure is ignored. This network would lead to a similar conclusion because of the network in Panel A: The shortest path from A to C goes via B. In this network, however, edges indicate Lag-1 effects. This means that the path $A \to B \to C$ indicates a Lag-2 effect, whereas the direct path $A \to B$ indicates a Lag-1 effect. Such paths are not even comparable because they indicate completely different temporal structures. Finally, Panel C shows both model matrices obtained from a VAR model. Suppose a researcher is interested in identifying which node is best able to predict all nodes at later measurement. In this case, only investigating the temporal structure would lead to the conclusion that the most important node is B. However, such an analysis would not take the contemporaneous network, in which Node A is highly central, into account.

Information theory. A potential solution for such problems is to *not* interpret probabilistic graphical models as networks, but rather, for what they are: full characterizations of the joint likelihood. In this line of thinking, the graphical representation is only useful for visualizing the statistical results but should not be over interpreted. The estimated model, nonetheless, is extremely powerful because it captures the associational structure of a dataset without the need for underlying theory on the causal mechanisms. A possible solution for inference methods then lies in the use of *information theory* (Cover & Thomas, 2012), which has shown to be a promising gateway to understanding the full complexity of such systems (Quax, Apolloni, & Sloot, 2013; Quax, Kandhai, & Sloot, 2013).

In information theory, we can make use of the *Shannon entropy* (Cover & Thomas, 2012) of a set of random variables, $\boldsymbol{Y}$, which denotes the average amount of bits of information needed to communicate a discrete outcome. When dealing with continuous variables, as we do in the GGM and VAR models, we can define the *differential entropy*:

$$h(\boldsymbol{Y}) = -\mathbb{E}\left[\log_2 f(\boldsymbol{Y})\right],$$

in which $f(\boldsymbol{y})$ denotes the density function of $\boldsymbol{Y}$. This measure can be computed for any number of variables and quantifies their volatility—in the case of a single continuous variable, the entropy is directly related to the variance. Now, divide $\boldsymbol{Y}$ in two subsets $\boldsymbol{Y}^{(1)}$ and $\boldsymbol{Y}^{(2)}$. We can then quantify the association in the two subsets using the *mutual information*:

$$I\left(\boldsymbol{Y}^{(1)}; \boldsymbol{Y}^{(2)}\right) = h\left(\boldsymbol{Y}^{(1)}\right) + h\left(\boldsymbol{Y}^{(2)}\right) - h\left(\boldsymbol{Y}^{(1)}, \boldsymbol{Y}^{(2)}\right).$$

This measure can act as a general measure for strength of association between any set of variables to any other set of variables.

In the network perspective, we treat strongly associated variables as densely connected and usually take the analogy of such strongly connected variables to be *close* to one another. Mutual information can then be seen as a new form of quantifying *closeness* between nodes or sets of nodes—the inverse of distance. Therefore, mutual information is an alternative to the shortest path length. This measure not only takes the shortest paths into account but all other paths as well. For example, the mutual information of two variables or nodes (we often use these terms interchangeably), $I(Y_i; Y_j)$, can be taken as a measure of how close these two nodes are to each other. The mutual information between two sets, $I\left(\boldsymbol{Y}^{(1)}; \boldsymbol{Y}^{(2)}\right)$—for example, in which Set (1) contains the symptoms of depression and Set (2) contains the symptoms of generalized anxiety—can be taken as a measure of closeness between two groups of nodes. Furthermore, the mutual information of one variable, (Y_1), with respect to all other variables, $(Y_{-(i)})$ and $I(Y_1; Y_{-(i)})$, can be used as a centrality measure. Finally, when temporal information is present, only computing the information one node has on all nodes at the next measurement, $I(Y_{t1}; Y_{t+1})$, can be taken as a temporal centrality measure, which takes into account both the contemporaneous and temporal network (see Chapter 5 and Chapter 6).

When $\boldsymbol{Y}$ has a multivariate normal distribution with size P and variance–covariance matrix $\boldsymbol{\Sigma}$, as is the case in both the GGM and graphical VAR model, the differential entropy becomes (Cover & Thomas, 2012, p. 250):

$$h(\boldsymbol{Y}) = \frac{1}{2}\log_2\left((2\pi e)^P \mid \boldsymbol{\Sigma}|\right).$$

As can be seen, this measure is a direct function of the size of the variance–covariance matrix $\boldsymbol{\Sigma}$. This expression allows us to compute all mutual informations described above. For example, the mutual information between two variables can be computed as:

$$I(Y_i; Y_j) = -\frac{1}{2}\log_2\left(1 - \sigma_{ij}^2\right).$$

As can be seen, this measure is a direct property of the explained variance σ_{ij}^2 between two variables. The mutual information of one variable with all other variables becomes:

$$I(Y_1; Y_{-(i)}) = \frac{1}{2}\log_2\left(\frac{|\boldsymbol{\Sigma}_{-(i)}|}{|\boldsymbol{\Sigma}|}\right),$$

in which $\boldsymbol{\Sigma}_{-(i)}$ denotes the variance–covariance matrix without row and column i. Finally, in a stationary time series of multivariate normal data (e.g., a Lag-1 VAR model), the temporal closeness above becomes:

$$I\left(y_{t,j}; \boldsymbol{y}_{t+1}\right) = \frac{1}{2}\log_2\left(\frac{\sigma_{jj} \mid \boldsymbol{\Sigma} \mid}{\mid \boldsymbol{\Sigma}_{j,+}^{\mathrm{TP}} \mid}\right),$$

in which $\boldsymbol{\Sigma}_{j,+}^{\mathrm{TP}}$ denotes a subsetted Toeplitz matrix:

$$\boldsymbol{\Sigma}_{j,+}^{\mathrm{TP}} = \mathrm{Var}\left(Y_{ti}, \boldsymbol{Y}_{t+1}\right),$$

which can be obtained from a VAR analysis.

Applying these metrics to the networks shown in Figure 12.1 leads to strikingly different interpretations. In Panel A, Nodes A and C are now shown to be independent; thus, Node B does not have a problematic interpretation of having a high betweenness. Now in Panel C,[3] Node A is shown to have slightly more information over the next time point than Node B, even though Node B has more temporal connections. Information theory is a promising gateway to analyzing the network models obtained. However, future researchers must thoroughly test and validated these metrics on psychological data.

The Importance of Intercepts

The network models outlined in this dissertation are all models of second-order moments. That is, they model variances and covariances but not expected values. The parameters that do model the expected value in the GGM, Ising model, and VAR model are the *intercepts*. When drawing a network, these are ignored. As such, when using the network structure to compute centrality, for example, intercepts are not taken into account. The problem with this approach is that links are formed between variables that may have largely different intercepts. Particularly in models of binary variables, links may be formed between variables that never have high entropy at the same time. For example, whenever a person is in danger of suffering from suicidal ideation (high entropy), we might expect that person to always experience sadness (low entropy). If we apply a virus-spreading analogy (Borsboom et al., 2011), such nodes would never "infect" each other; the link would never be used.

Figure 12.2, Panel A, shows an example of a network we might estimate on educational data. This network is a fully connected Ising model, also called a Curie-Weiss model, which is known to be equivalent to the IRT model shown in Panel B (Marsman et al., 2015; see also Chapter 8). As such, IRT models are often used and work well on educational data, the Ising model of Panel A is not unreasonable. We can see a link between the items "1 + 1" and "0.07692 + 0.3409." This link represents a very plausible predictive relationship. Knowing someone can answer "0.07692 + 0.3409" tells us that person can also answer "1 +

[3]Panel B only shows half the information needed to characterize the full likelihood; see Chapter 6.

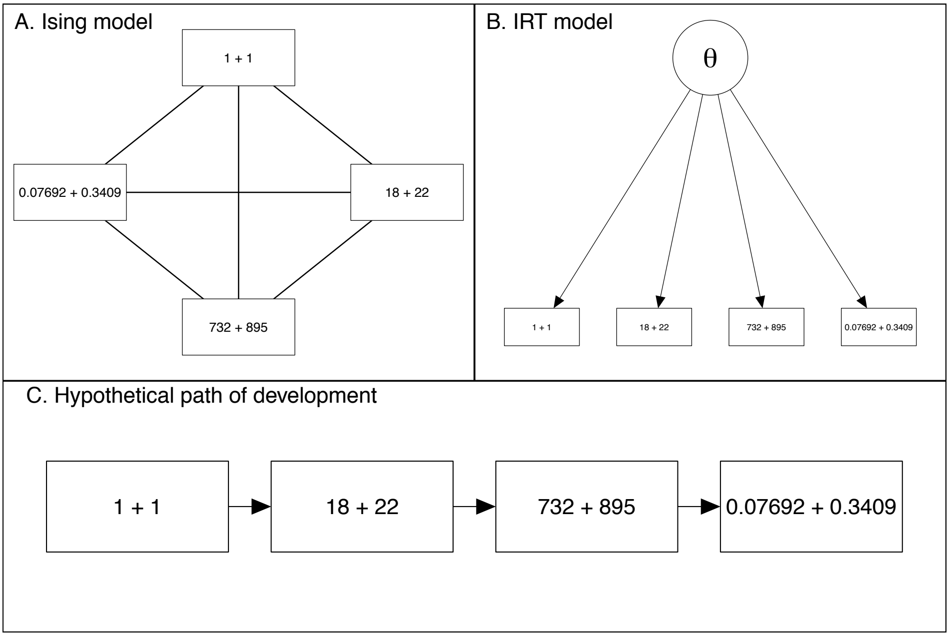

Figure 12.2: Three potential network structures on educational data. Panels A and B are equivalent models that do not show the difficulty of an item. In Panel C, items are ordered according to their difficulty, pointing from the easier item to the more difficult item.

1" (because the person is able to answer the much harder question as well), and knowing someone cannot answer "1 + 1" also tells us that person cannot answer "0.07692 + 0.3409" (because that person cannot answer the simpler question). These items likely do not have high entropy at the same time in any person's life. This means that when a person is struggling with "1 + 1," and may or may not answer this item correctly (high entropy), it is highly likely that this person will never be able to answer "0.07692 + 0.3409" correctly (always an incorrect response; low entropy). Conversely, whenever a person correctly answers the item "0.07692 + 0.3409," that person will likely always answer the item "1 + 1" correctly.

The network perspective would lead to the following interpretation of Panel A, that one could influence the probability of correctly answering "0.07692 + 0.3409" by training someone on the question "1 + 1." However, this seems unlikely. Teaching a person the techniques needed to answer "1 + 1" would not prepare that person to answer "0.07692 + 0.3409," which requires knowledge of decimal points, counting over tens, and so forth. The latent variable model in Panel B would implicate that training someone on one of the items would never help that person answer other items correctly. Although I do not wish to argue against mathematical ability, I do think that such an assumption might also be too strict. Children learn by making items, and learning how to make one item helps a child make another item. This learning, however, does not jump wildly as would be expected from Panel A, rather it follows a straight path. Learning “1 + 1” helps to answer the item “2 + 2,” which helps to answer the item “3 + 3”, and so forth.

Panel C shows a network that is based only on intercepts rather than covariances. Here, each item points to the first harder item. This very different network structure shows that people first learn “1 + 1,” then “18 + 22,” and so forth. I term such a network structure a *path of development*. The network shown in Panel C is merely a hypothetical example of what a network that also takes intercepts into account could look like. In perfect unidimensional cases, it might look like Panel C, whereas in multidimensional cases, one could envision, for example, parallel paths or the path splitting. I do not seek to propose a new modeling framework in this section but merely wish to highlight that taking intercepts into account could lead to different ways of investigating the phenomena of interest.

Complexity

The network models, as outlined in this dissertation, are but one of the many consequences that may come from a more general *hypothesis of complexity*. Psychological behavior plausibly is the result of emergent behavior in a complex system of interacting psychological, biological, and sociological components. Simply stated, psychology is complex. People’s behavior is dictated by their brains, which consist of billions of neurons, formed by many years of development. As such, every person is a mini universe of complexity. These universes, in turn, interact with one another in complicated social networks. Perhaps, psychology is one of the hardest fields to tackle. It is, in my opinion, only logical that many behaviors have no simple explanation.

The network model is but one attempt at grasping this complexity; we should not get sidetracked by believing it is the only possible attempt. The hypothesis of complexity is not limited to the expectation that data are generated due to an underlying (sparse) network model of, at most, second-order moments. This hypothesis reaches further, with many more implications. This point of view can take psychological research in many different directions—rather than merely the estimation of network models. For example, long-term predictions can be made on the effects of interventions, without understanding the true underlying causal mechanisms. Also of particular importance is the work done by van de Leemput et al. (2014) and Wichers et al. (2016) on identifying early warning signals for phase transitions in psychology, such as the onset of depression. I think that the hypothesis of complexity has much to offer in the years to come and will change psychological research in ways we cannot imagine now.

12.3 Conclusion

In this discussion, I outlined various topics for future research which can be tackled in network psychometrics: improving centrality measures, handling missing and ordinal data, quantifying evidence for sparsity in the network, and incorporating intercepts in inference on these models. This is just a highlight of several future directions; many more can be conceived, such as tackling heterogeneity, improving multilevel estimation of contemporaneous effects, handling a mixture of observational and experimental data, and extending networks to nonlinear dynamics. Finally, I noted a far more general field of research—complexity in psychology—of which network modeling is merely a small part. The network models proposed in this dissertation add much to the toolbox of psychological and psychometric researchers. Network psychometrics, however, is still a young field of research with many unanswered questions. The full utility of these methods and their place in psychological research and psychometrics will be determined on the road ahead.