Psychological Methods

On Ising Models and Algorithms for the Construction of Symptom Networks in Psychopathological Research

Michael J. Brusco, Douglas Steinley, Michaela Hoffman, Clintin Davis-Stober, and Stanley Wasserman

Online First Publication, October 7, 2019. http://dx.doi.org/10.1037/met0000207

American Psychological Association

Psychological Methods

1082-989X/19/$12.00

http://dx.doi.org/10.1037/met0000207

On Ising Models and Algorithms for the Construction of Symptom Networks in Psychopathological Research

Michael J. Brusco
Florida State University

Douglas Steinley
University of Missouri—Columbia

Michaela Hoffman
Medical University of South Carolina

Clintin Davis-Stober
University of Missouri—Columbia

Stanley Wasserman
Indiana University, Bloomington

Abstract

During the past 5 to 10 years, an estimation method known as *eLasso* has been used extensively to produce symptom networks (or, more precisely, symptom dependence graphs) from binary data in psychopathological research. The *eLasso* method is based on a particular type of Ising model that corresponds to binary pairwise Markov random fields, and its popularity is due, in part, to an efficient estimation process that is based on a series of l_1-regularized logistic regressions. In this article, we offer an unprecedented critique of the Ising model and *eLasso*. We provide a careful assessment of the conditions that underlie the Ising model as well as specific limitations associated with the *eLasso* estimation algorithm. This assessment leads to serious concerns regarding the implementation of *eLasso* in psychopathological research. Some potential strategies for eliminating or, at least, mitigating these concerns include (a) the use of partitioning or mixture modeling to account for unobserved heterogeneity in the sample of respondents, and (b) the use of co-occurrence measures for symptom similarity to either replace or supplement the covariance/correlation measure associated with *eLasso*. Two psychopathological data sets are used to highlight the concerns that are raised in the critique.

Translational Abstract

We present the first principled critique of the Ising model and the associated *eLasso* method. The Ising model is a form of a pairwise Markov random field that is used to model binary data. The current approach utilizes a set of logistic regressions to obtain parameter estimates, with the final model being chosen with the so-called *eLasso* method to mitigate against overfitting. Throughout, we highlight why application of this model to data with traits that are commonly found in psychopathological data may be problematic due to issues such as sample size, density, asymmetric parameter weights, and so on. Specifically, the necessary sample size for fitting these models is dependent on the density (e.g., the number of connections) of the network, with a strong preference being given to networks that are not dense. Unfortunately, several (if not the vast majority of) psychopathology disorders have been conceptualized as cohesive traits, and as such, methods used to measure them assume that all of the items arise from a common factor. This standard assumption leads to items that are highly intercorrelated, resulting in networks that are dense—exactly the situation that is not desired for the predominant estimation techniques of the Ising model. In turn, we recommend a partitioning approach that may help address some of these issues. Beyond this issue, we assess the appropriateness of averaging parameter asymmetric parameter estimates to achieve symmetric parameter estimates—a practice that could result in the loss of information. Finally, we show that it is likely that much of the data analyzed with these methods do not arise from the same homogeneous population; as such, we introduce a partitioning method to aid in determining if several "subnetworks" are being averaged together when they are fit together as one big network, resulting in global findings that are not consistent with any local structures.

Keywords: binary pairwise Markov random fields, Ising models, dependence graphs, two-mode networks, partitioning methods

Editor's Note. Kristopher J. Preacher served as the action editor for this article that was submitted and reviewed before Douglas Steinley started his term as Incoming Editor of *Psychological Methods*.—LLH

Michael J. Brusco, Department of Business Analytics, Information Systems and Supply Chain, Florida State University; Douglas Steinley, Department of Psychological Sciences, University of Missouri—Columbia; Michaela Hoffman, Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina; Clintin Davis-Stober, Department of Psychological Sciences, University of Missouri—Columbia; Stanley Wasserman, Department of Statistics, Indiana University, Bloomington.

Correspondence concerning this article should be addressed to Douglas Steinley, Department of Psychological Sciences, University of Missouri, 210 McAlester Hall, Columbia, MO 65211. E-mail: steinleyd@missouri.edu

Psychopathology data are often binary, representing whether a list of symptoms is present (i.e., recorded as a 1) or absent (i.e., recorded as a 0) for each participant in the data set. This is known as "two-mode" data because it includes information about (a) individuals, represented by the rows of the data set; and (b) symptoms, represented by the columns. This data structure is very common in statistics and arises as the matrix of affiliations in an affiliation network (see Wasserman & Faust, 1994, Chapter 8). It is also the basis of much of multivariate statistics (in both binary and nonbinary form)—a standard $n \times p$ matrix, $\mathbf{X} = [x_{ij}]$, of p variables measured on n individuals.

During the past several years, an extraordinarily popular method of analysis for affiliation networks in psychopathology has been the *eLasso* procedure (van Borkulo et al., 2014).[1] The *eLasso* itself is an adaptation of a method designed for high-dimensional graphical model selection within the framework of binary pairwise Markov random fields, which is a particular type of Ising model (Anandkumar, Tan, Huang, & Willsky, 2012; Banerjee, El Ghaoui, & d'Aspremont, 2008; Barber & Drton, 2015; J. Guo, Cheng, Levina, Michailidis, & Zhu, 2015; J. Guo, Levina, Michailidis, & Zhu, 2010; Höfling & Tibshirani, 2009; S.-I. Lee, Ganapathi, & Koller, 2007; Loh & Wainwright, 2013; Ravikumar, Wainwright, & Lafferty, 2010; Santhanam & Wainwright, 2012; Wainwright, Ravikumar, & Lafferty, 2007; Xue, Zou, & Cai, 2012). The *eLasso* procedure seeks to estimate a $p \times p$ network matrix for the symptoms along with a corresponding network graph. Although many researchers who study *eLasso* and related methods apply the term, the use of the word "network" in this context is problematic because there are no measured relational ties between the symptoms. The object under study by researchers using the *eLasso* method is more precisely known as a *dependence graph*.[2]

Early work discussed the potential for the network analysis of models of psychological disorders (e.g., Borsboom, 2008; Borsboom & Cramer, 2013; Schmittmann et al., 2013), focusing on dependence graphs estimated using bivariate correlations, odds ratios, or even self-reported subjective associations. Once again, in this research, there was no interest in the individuals represented by the rows of **X**, which is a crucial assumption if the situation is to be viewed as a network. In fact, the rows are assumed to be merely a sample of uninteresting individuals, completely independent, with no gathering of relevant relational information on them (in contrast, in traditional network analysis, information is gathered regarding the relation between individuals).

With the development of the *eLasso* method specifically for binary data, as well as its associated R package (IsingFit), there has been a sharp increase in applications (both correct and incorrect) of graphical analysis in psychological research. The algorithm has been implemented in a variety of articles examining the structures of both single disorder dependence graphs, such as substance abuse (e.g., Rhemtulla et al., 2016) or major depressive disorder (e.g., Cramer et al., 2016; Kendler, Aggen, Flint, Borsboom, & Fried, 2018), and dependence graphs, including multiple disorders (e.g., Afzali et al., 2017; Bekhuis, Schoevers, van Borkulo, Rosmalen, & Boschloo, 2016; Boschloo et al., 2015).

Despite the popularity of *eLasso* for constructing symptom dependence graphs in psychopathology, some authors have raised concerns regarding the saliency of the results it produces. For example, Forbes, Wright, Markon, and Krueger (2017) recently questioned the replicability of the *eLasso* model for psychopathological data. Their skepticism was supported by the results of Steinley, Hoffman, Brusco, and Sher (2017), who showed that the marginal distributions of the variables modeling the underlying data imposed a significant degree of structure. In their rebuttal to the findings of Forbes et al., Borsboom et al. (2017) reanalyzed the data from the Forbes et al. study and concluded that *eLasso* and other dependence graph approaches produced replicable results.

Expanding on this current debate, we posit that there are a number of legitimate concerns regarding the use of the *eLasso* method in psychopathological research. In part, these concerns stem from differences between the types of data for which the binary pairwise Markov random field (Ising) model is most suitable and the characteristics of data structure in psychopathology. Although clinical researchers have been informed of some of the merits of the Ising model and the estimation procedures used in the *eLasso* procedure, the assumptions, conditions, and limitations of these models and methods have not, in our opinion, been adequately articulated. We seek to rectify this problem by providing a careful presentation of key issues that pertain to the Ising model in general, such as the emphasis on high-dimensional sparse graphs, the assumption that observations are random samples from a single underlying distribution, and the capitalization on correlation between symptoms that may accrue because of the absence of symptoms for a large number of individuals. We also address issues that are more specific to the particular estimation process used by *eLasso*, such as the dependency and incoherence assumptions observed by Wainwright et al. (2007) and the forced symmetry of the edge weights that is required by the *eLasso* method (Höfling & Tibshirani, 2009).

Of particular concern is the potential for heterogeneity in binary affiliation networks, which has recently been studied by J. Guo et al. (2015). The *eLasso* treats the rows of **X** as random samples from an underlying distribution based on the Ising model. This is a common procedure in network analysis but is not appropriate because of the assumption that the individuals are completely lacking interest, with the focus being placed primarily on the variables. However, this indeed may not be the situation. We posit that the rows of **X** are likely obtained from several subpopulations with different (yet not necessarily readily apparent) symptomatic structures. This type of *unobserved* heterogeneity has been recog-

[1] van Borkulo et al. (2014, p. 2) coined the term *eLasso* for their methodology, which they implemented in within the R software program IsingFit. There is also an R software package named 'elasso,' which was developed by P. Guo (2015) for variable selection in logistic regression. Throughout our article, we refer to *eLasso* as the methodology developed by van Borkulo and implemented in the IsingFit package.

[2] Although some researchers will maintain that the debate over the terms *network* vs. *dependence graph* is merely one of semantics, we believe there is a salient distinction. The edge weights obtained from the output of *eLasso* are not relational ties between symptoms but derived measures of association or similarity. By this definition, a correlation matrix could also be viewed as a network. In fact, a matrix of Euclidean distances between pairs of observations (such as found in many cluster analysis and multidimensional scaling routines) could be viewed as a network, meaning that every data set could be viewed as a network as long as an appropriate similarity measure was constructed. This interpretation can be problematic when researchers seek to apply standard measures of social network theory, such as centrality, to the edge weights, which may not be appropriate. In any case, whether or not the use of the term *network* is proper is not central to our critique of *eLasso*.

nized as a serious concern within the context of several other multivariate statistical methods, such as regression (Brusco, Cradit, Steinley, & Fox, 2008; DeSarbo & Cron, 1988; Späth, 1979), structural equation modeling (Jedidi, Jagpal, & DeSarbo, 1997; Muthén, 1989), partial least squares (Sarstedt & Ringle, 2010), exponential random graph models (Steinley, Brusco, & Wasserman, 2011), decision making (Brown, Park, Steinley, & Davis-Stober, 2018), and multidimensional scaling (Brusco & Cradit, 2005; DeSarbo, Manrai, & Manrai, 1994).

One possible approach for addressing the heterogeneity problem is a two-step process whereby the individuals are partitioned and then *eLasso* is used to construct the symptom dependence graphs for each of the subgroups of the partition. We evaluate this two-step procedure using *p*-median (also known as *K*-median) clustering for the first step of the process (Brusco, Shireman, & Steinley, 2017; Köhn, Steinley, & Brusco, 2010). This was accomplished in R using the partitioning around medoids method (Kaufman & Rousseeuw, 2005), which is available via the 'pam' software program in the 'cluster' library. Although the two-step approach can afford some benefits, we also identify some serious limitations and recommend that heterogeneity might be better accommodated by the development of mixture modeling approaches similar to those available for regression (DeSarbo & Cron, 1988), structural equation modeling (Jedidi et al., 1997), and partial least squares (Sarstedt & Ringle, 2010). We also acknowledge that addressing the heterogeneity issue does not correct all of the concerns associated with the Ising model and *eLasso* method.

In the next section, we provide a brief overview of the Ising model and its application to the study of data and networks in contexts outside the physical sciences. This is followed by a more focused analysis on graphical model selection for high-dimensional binary pairwise Markov random fields as well the solution approach based on l_1-regularized logistic regression and a precise description of the *eLasso* method. This is immediately followed by a critique of the Ising model and the *eLasso* estimation method. Results stemming from the application of *eLasso* to data from the National Comorbidity Survey—Replication (NCS-R; Forbes et al., 2017), which pertain to major depressive and generalized anxiety symptoms, are then used to highlight key points of the critique. These issues are then reinforced with a second application associated with alcohol use disorder, using data obtained from a national representative sample (National Epidemiologic Survey of Alcohol and Related Conditions [NESARC]; Grant et al., 2004). The article concludes with a brief summary as well as discussion of limitations and suggestions for future research.

The Ising Model

The Ising model (Ising, 1925) was originally formulated within the framework of ferromagnetism and later generalized to the study of other phase transitions that arise in the physical universe. The key underlying principle of Ising models is that small measurable changes in a system parameter can suddenly result in dramatic qualitative change in the state of that system. Commonly cited examples include water temperature decreasing to its freezing point (changing the state from liquid to solid) or increasing to its boiling point (changing the state from liquid to gas). Cipra (1987) provides an excellent mathematical overview of Ising models and their inherent combinatorial structure, and Wasserman (1978) describes how such models can be applied to networks.

In recent years, there have been attempts to expand the application of Ising models beyond the study of large-scale physical systems in physics, chemistry, and biology so as to address topics pertinent to social network analysis. As mentioned, Wasserman (1978) was the first to point out this extension (see also Kindermann & Snell, 1980). Galam and Moscovici (1991) used Ising model principles in their study of attitude changes and consensus in groups. Galam (1997) also investigated the use of Ising models within contexts of rational group decision making and opinion formation. Similar studies were later undertaken by Grabowski and Kosinski (2006) and Liu, Ying, and Shakkottai (2010). In their study of patterns of interaction and reciprocity in the open-source software community, Oh and Jeon (2007) used simulation analysis to validate their Ising theory-based propositions. Laciana and Rovere (2011) used an Ising model approach to study technology diffusion. The Ising model has also been applied to the study of tax evasion fluctuations in agent-based communities (Lima, 2015).

Clearly, the Ising model has been implemented for a broad range of applications in the social sciences. Through the development of l_1-regularized logistic regression methods for fitting Ising models and the incorporation of such methods in software programs such as *eLasso*, the application of Ising models in psychopathology has been especially prevalent in recent years. In the next section, we carefully describe the model, methods, assumptions, and conditions that serve as the foundation for these applications.

Binary Pairwise Markov Random Fields (Ising Model)

The Model

The underlying model of interest is a special case of graphical loglinear models or Markov fields (Lauritzen, 1996). In particular, of interest is the pairwise Markov random field model for binary data, which is a type of Ising model. Conceptually, the p variables (e.g., symptoms) corresponding to the columns of the data matrix ($\mathbf{X}$) are considered as vertices of a graph. The n rows of the matrix are considered as independent random samples from an underlying distribution and are used to estimate the presence or absence of edges between pairs of vertices (e.g., objects/symptoms). A formal presentation of the Ising model of interest is provided in Appendix A.

The underlying optimization problem is one of estimating a $p \times p$ matrix of pairwise interaction parameters, $\mathbf{\Theta} = [\theta_{st}]$, so as to maximize a penalized log-likelihood function. It is almost universally assumed in the literature that $\mathbf{\Theta}$ is symmetric (i.e., $\theta_{st} = \theta_{ts}$). In many contexts, the estimates are simply used to identify the presence or absence of edges of a graph, $G = (V, E)$. In a psychopathology context, the vertex set, V, consists of the p symptoms. An edge between symptom s and symptom t (where $s < t$) exists if the estimate for θ_{st} is nonzero. Accordingly, the symptom pair $\{s, t\}$ is a member of the edge set E (that is, $\{s, t\} \in E$). Alternatively, the θ_{st} estimates can be considered as measures of similarity between symptoms s and t and be applied directly as edge weights for the resulting graph.

One of the most underdeveloped issues in the study of graph estimation using binary Ising models is the issue of edge weights. In the description of their method for graph estimation for the Ising

model using p distinct l_1-regularized logistic regressions, Wainwright et al. (2007, p. 1466) indicate that the direct estimation of the θ_{st} parameters was of "secondary concern." This is also true for most of the studies that followed (Barber & Drton, 2015; Höfling & Tibshirani, 2009; Ravikumar et al., 2010; Santhanam & Wainwright, 2012). For example, Ravikumar et al. (2010, p. 1291) explicitly stated that "the goal of graphical model selection is to infer the edge set E." Moreover, the emphasis was on signed-edge recovery. In other words, for an estimate $\theta_{st} > 0$, a positive edge between vertices s and t is assumed to be present in E. However, if $\theta_{st} < 0$, then a negative edge is assumed. An estimate of $\theta_{st} = 0$ suggests that there is no edge associated with vertices s and t.

The emphasis on recovery of the edge set notwithstanding, Höfling and Tibshirani (2009, p. 887) noted that the Wainwright et al. (2007) method could also be applied directly to estimate the off-diagonal parameters of $\mathbf{\Theta}$, and that the intercepts for each of the p l_1-regularized logistic regression models could be placed along the main diagonal of $\mathbf{\Theta}$. Santhanam and Wainwright (2012) also explicitly refer to the elements of $\mathbf{\Theta}$ as edge weights. It seems, therefore, that there is not broad consensus on whether to treat the interaction parameters (θ_{st}) directly as edge weights or only to use them as indicators of the absence or presence of signed edges. Additionally, there appears to be a dearth of studies that explicitly evaluate the capabilities of extant methods for recovering the interaction parameters. Instead, the emphasis has been on the recovery of the edges of *high-dimensional sparse* graphs (Banerjee et al., 2008; Barber & Drton, 2015; Ravikumar et al., 2010; Wainwright et al., 2007; Xue et al., 2012). Sparsity suggests that the total number of edges in the graph is small in comparison to the possible number of edges that could exist, but it is also often extended to include an assumption that the *maximum degree* (d) is small in comparison to p. The *degree* of a vertex is the number of edges that it shares with other vertices, and the *maximum degree* is the largest degree across all vertices.

There are a number of assumptions and conditions that affect the computational viability of graph recovery algorithms for binary pairwise Markov random fields (Ising models). *Graph recovery* generally pertains to the ability of an algorithm to perfectly identify the correct (or true) edge set for a graph. Perfect identification requires that all of the true edges are estimated and also that no superfluous (untrue) edges are estimated. In circumstances in which the edge weights themselves are also of interest, graph recovery can be extended to include measures of how well the estimated θ_{st} values approximate the true parameter values.[3]

When considering graph recovery, we believe that there are three aspects that pertain to the Ising model that should not be overlooked: (a) the model has most typically been evaluated with respect to the recovery of high-dimensional *sparse* graphs, (b) it is assumed that the rows of the data matrix are independent and identically distributed random samples from a single underlying distribution, and (c) when considering two columns (e.g., symptoms) of the binary data matrix, matches of zeros are of equal importance to matches of ones. We expand on each of these three issues in the following subsections and emphasize their implications for the analysis of psychopathological data.

High-Dimensional Sparse Graphs

The importance of the emphasis on high-dimensional *sparse* graphs cannot be understated. In their pioneering work on high-dimensional graph estimation for Gaussian models, Meinshausen and Bühlmann (2006, p. 1441) note that "the main assumption is the sparsity of the graph." Although high dimensionality and sparsity are not assumptions for the Ising model, these properties are consistently emphasized in the literature pertaining to the model and *eLasso* (Barber & Drton, 2015; Höfling & Tibshirani, 2009; Ravikumar et al., 2010; Santhanam & Wainwright, 2012; Xue et al., 2012). Regrettably, clear definitions of precisely what constitute high dimensionality and sparsity are often lacking.

There is a body of existing research that examines the effect of sparsity on the potential for graph recovery. Wainwright et al. (2007) addressed assumptions associated with the growth rates of n, p, and the maximum degree of the graph, $d = \max_{s \in V}\{d(s)\}$. Wainwright et al. and Ravikumar et al. (2010) showed that the sample size requirement for graph recovery grows logarithmically with p but exponentially as a function of d. Santhanam and Wainwright (2012) later expanded the conditions for p and d to include the minimum interaction coefficient and the maximum neighborhood size.[4] The minimum interaction coefficient (λ) is the absolute value of the nonzero off-diagonal element in $\mathbf{\Theta}$ that is closest to zero. The maximum neighborhood size (ω) is the value of the largest row (or column) sum (excluding the main diagonal) in $\mathbf{\Theta}$. Santhanam and Wainwright derived sample size lower bounds necessary for graph recovery (with a probability of at least δ) as a function of p, d, λ, and ω. It is important to note that the sample size conditions derived by Santhanam and Wainwright are information-theoretic limits that pertain to the sufficiency of *some* graph recovery algorithm and, therefore, represent a best-case scenario. The conditions for any particular estimation algorithm are potentially more demanding. The formulae for the sample size lower bounds are provided in Appendix B.

Most of the literature on the use of binary pairwise Markov random fields for graphical model selection pertains to high-dimensional sparse data with small values of maximum degree d in comparison to the number of vertices p (Barber & Drton, 2015; Höfling & Tibshirani, 2009; Ravikumar et al., 2010; Santhanam & Wainwright, 2012; Wainwright et al., 2007; Xue et al., 2012). When considering the simulation studies that have been completed to analyze graph recovery, the emphasis has almost always been on high-dimensional sparse graphs (Barber & Drton, 2015; Epskamp, Borsboom, & Fried, 2018; Höfling & Tibshirani, 2009; Ravikumar et al., 2010; van Borkulo et al., 2014), yet there are some noteworthy differences across studies. For example, Ravikumar et al. (2010) and Barber and Drton (2015) use levels of $p = 64$, $p = 100$, and $p = 225$ in their simulation studies, whereas van Borkulo et al.

[3] In practical applications, the true graphical structure is seldom known with certainty. Therefore, graph recovery is commonly examined in a controlled (simulation) context in which the true structure is known. Nevertheless, it is possible that, in some contexts, researchers might have strong hypotheses regarding what the true structure should be for a given application.

[4] The minimum interaction coefficient is computed as $\lambda = \min_{\{s,t\} \in E}\{|\theta_{st}|\}$ and the maximum neighborhood size is computed as $\omega = \max_{s \in V}\{\sum_{t \in \eta(s)} \theta_{st}\}$.

(2014) use mostly lower dimensionality levels of $p = 10$, $p = 20$, $p = 30$, and $p = 100$. Moreover, with respect to sparsity, Ravikumar et al. and Barber and Drton controlled *maximum degree* in their design conditions through the use of nearest-neighbor lattices and star graphs, whereas the random data sets in the van Borkulo et al. study control *average* degree via tie probabilities.

Given that the Ising model has been tested primarily for the estimation of *sparse* graphs, the key question for the psychopathological researcher using *eLasso/IsingFit* is, Is the symptom graph associated with the particular psychopathological research study sparse? Although the "true" graphical structure is typically unknown in psychopathological research, it is our experience that the estimated graphs are often not sparse, having both a high maximum degree and maximum neighborhood size.

Indeed, one of the hallmarks of psychopathology research is the positing of the existence of "disorders." The disorders are often conceptualized as latent variables, in which the measured items are imperfect approximations of those latent variables. Although there has been some debate as to whether the latent variables are more appropriately conceptualized as continuous or categorical (see Steinley & McDonald, 2007, for a discussion), they are frequently assessed with common factor models. Often, each disorder is represented by a single factor. Focusing on the single-factor model can be instructive. One assumption of the common factor model is that all correlations among a set of variables are represented by the common factor. The variance of each variable can be broken down into two parts: (a) variance related to the common factor (and, hence, related to the other variables in the model), and (b) variance that is "unique" to the variable and unrelated to the other variables. The common information is usually represented by the factor loading (e.g., how strongly the variable is related to the common factor), while the unique information is relegated to the residual of the variable (e.g., the information that cannot be explained by the common factor). Generally, when measures are developed to assess disorders, it is recommended that a set of items are created to capture the underlying unobservable, latent disorder. Because each item is an imperfect representation of the latent factor, a set of items is used to measure the latent variable more reliably. All of this is to say that we expect all pairs of items measuring the disorder to have moderate to high intercorrelations. As such, the corresponding network would be a very dense, if not, complete graph (e.g., note that a complete graph is when all vertices in the network are connected).

A Single Underlying Distribution

The general data assumption is that the rows of $\mathbf{X}$ consist of n independent identically distributed (i.i.d.) realizations from the Ising model (Xue et al., 2012). A key underlying aspect of nearly all of the extant simulation analyses is that they generate data from an Ising model, and then examine how well a method based on the Ising model recovers the simulated graphical structure. The primary measures of performance are typically based on (a) positive selection—true edges in the graph that are recovered, and (b) false discovery—the edges estimated by the method that are not present in the underlying graph. In addition, when graph recovery is expanded to include the actual estimation of the edge weights, van Borkulo et al. (2014) consider a measure of correlation between the estimated graph parameters and the true underlying parameters. To an extent, the strong performance of methods in these studies is tautological because of the relationship between the data generation process and the estimation method.

In psychopathological research, it seems implausible that the rows of $\mathbf{X}$ consist of n i.i.d. realizations from the Ising model. Instead, it is more likely that there is considerable heterogeneity in the sample. That is, the data are obtained by sampling from several different subpopulations exhibiting different symptomatic patterns. A failure to consider the heterogeneity of the sample in the Ising model context may lead to the estimation of a network that does not actually represent *any* of the individual respondents (or, possibly, a very small subset of the respondents). This effect, which is conceptually similar to the Condorcet paradox in voting theory (Gehrlein, 2002), is attributable to the process of averaging across data sampled from several different subpopulations. Moreover, the adverse consequences of this averaging effect have been well-documented in another area of graphical statistics, namely, multidimensional scaling (Ashby, Maddox, & Lee, 1994; Brusco & Cradit, 2005; M. D. Lee & Pope, 2003; Siegler, 1987).

The need for the flexibility to accommodate this type of heterogeneity in graphical models has recently been emphasized for both the Gaussian (J. Guo, Levina, Michailidis, & Zhu, 2011) and Ising (J. Guo et al., 2015) cases. To address the heterogeneity issue and potentially improve the performance of graphical models, we explore the possibility of partitioning the sample using p-median clustering (Kaufman & Rousseeuw, 2005; Köhn et al., 2010) and, subsequently, fitting a graphical model to each cluster. A similar type of two-step approach has been proposed and tested in the context of multidimensional scaling (Brusco & Cradit, 2005).

Covariance/Correlation Measures Versus Co-Occurrence Measures

The primary output of *eLasso* and other methods for fitting the Ising model is a $p \times p$ similarity matrix ($\mathbf{\Theta}$) that is established from the estimated coefficients. Although this is called an average weighted adjacency matrix and discussed in a network context, it is essentially a matrix of pairwise similarity measures. Accordingly, an alternative perspective would be to view *eLasso* as a possible approach for collapsing an $n \times p$ binary matrix into a $p \times p$ weighted similarity matrix. From this perspective, the *eLasso* method is properly placed outside the domain of network analysis and is more analogous to the scores of binary similarity coefficients that are pervasive in the classification literature (for reviews, see Choi, Cha, & Tappert, 2010; Sneath & Sokal, 1973; Sokal & Sneath, 1963; Todeschini et al., 2012).

The Ising model and *eLasso* produce a similarity matrix based on covariance among the symptoms and, in this context, the output is similar to correlation-based binary similarity coefficients such as Pearson correlation, tetrachoric correlation, and Yule's Q (Yule, 1900). A key observation is that when establishing a similarity coefficient between a pair of symptoms, all of these measures treat matches of zeros (i.e., absence of both symptoms) as equally important to matches of ones (i.e., presence of both symptoms). Nevertheless, during the past century, many of the binary similarity coefficients that have been proposed either ignore or give less weight to zero matches (Faith, 1983; Jaccard, 1901; Kulczynski, 1927; Ochiai, 1957).

Our goal is not to advocate in favor of similarity measures that ignore or down-weight zero matches. However, we do believe it is important for psychopathological researchers to recognize the issue. To illustrate, consider four symptoms: S1, S2, S3, and S4. Suppose that S1 and S2 have very high base rates and S3 and S4 have very low base rates. Because of their high base rates, it is likely that S1 and S2 will have far more matches of ones (i.e., co-occurrence of symptom presence) than S3 and S4. Likewise, because of their low base rates, S3 and S4 are apt to have far more matches of zeros (i.e., symptom absence) than S1 and S2. If the matches of zeros for S3 and S4 generally outpace the matches of ones for S1 and S2, then correlation and covariance measures will tend to judge S3 and S4 as more similar than S1 and S2. But are S3 and S4 really more similar than S1 and S2? It is possible that S3 and S4 might seldom co-occur in individuals, yet they are judged similar simply because their base rates are so low and they have many zero matches. By contrast, S1 and S2 frequently co-occur in individuals, yet they are judged less similar than S3 and S4. The binary similarity coefficients that ignore zero matches would certainly judge S1 and S2 more similar than S3 and S4.

The distinction between symptom presence matches and symptom absence matches would seem to be particularly relevant in circumstances in which the psychopathological researcher is trying to establish causality. A high covariance/correlation between a pair of symptoms that arises because they have low base rates but seldom, if ever, co-occur together cannot provide much information with respect to one symptom leading to another. However, by studying the presence (co-occurrence) of symptoms, a foundation for causality might be possible. For example, regardless of whether the base rates for a pair of symptoms is low or high, if one symptom is only observed when the other is present, then at least a potential basis for causality might be established.

Solution Method

Exact Versus Approximate Method

Höfling and Tibshirani (2009) outline several solution methods for estimating the θ_{st} parameters. These methods can be classified into *exact* and *approximate* categories, which are distinguished by the fact that methods in the former are guaranteed to identify a global maximum whereas the latter are not. Exact procedures consider probability functions aggregated over all $\{s, t\} \in E$, and seek to maximize the l_1-penalized log-likelihood function. S.-I. Lee, Ganapathi, et al. (2007) developed an exact method that operates by applying standard nonlinear programming methods to the penalized log-likelihood function for a gradually expanded set of variables. A more efficient exact procedure was subsequently designed by Höfling and Tibshirani. Their method is based on repeated iterations of a pseudolikelihood approach that converge to the exact solution for the regular likelihood. Höfling and Tibshirani indicate that the pseudolikelihood estimations yield efficiency benefits relative to the conjugate gradient approaches in related methods (see S.-I. Lee, Lee, Abbeel, & Ng, 2007).

Perhaps the most widely used method for solving the penalized log-likelihood function is an approximate method based on l_1-regularized logistic regression. Rather than tackle the estimation of $\mathbf{\Theta}$ in its entirety, this approach subdivides the problem into p distinct l_1-regularized logistic regressions. The method was originally described by Wainwright et al. (2007), with an extended treatment provided by Ravikumar et al. (2010). The process for estimating $\mathbf{\Theta}$ is neighborhood-based l_1-regularized logistic regression. Each vertex (s) is considered in turn, and all other vertices ($t \in V \backslash s$) are used as predictors. It is then a simple process of setting $\theta_{ss} = \beta_{s0}$ (the intercept of the l_1-regularized logistic regression solution), and $\theta_{st} = \beta_{st}$ (the coefficient of the l_1-regularized logistic regression solution associated with variable t when used to predict s). Repeating this process p times, using each $s \in V$ as the dependent variable in turn, populates $\mathbf{\Theta}$. The lasso constraint in the l_1-regularized logistic regression model tends to produce some parameter estimates (model coefficients) of zero, whereas other coefficients will be positive or negative.

Although the approach of running p separate l_1-regularized logistic regressions developed by Wainwright et al. (2007) and Ravikumar et al. (2010) performed reasonably well in experiments conducted by Höfling and Tibshirani (2009), subsequent comparisons have shown that the method is systematically outperformed by methods that estimate the interaction coefficients in a single (unified) optimization problem (J. Guo et al., 2010; Xue et al., 2012). More specifically, the superior performance of the unified approaches had been demonstrated with respect to both edge recovery and estimation of the interaction parameters. The *eLasso* method incorporated in the IsingFit software program is a particular adaptation of the Ravikumar et al. procedure and is described more precisely in the next subsection.

The *eLasso* Method

The primary characteristics of the *eLasso* method proposed by van Borkulo et al. (2014) and implemented in the IsingFit software package for estimation of psychopathology dependence graphs can be succinctly summarized as follows:

1. The underlying model for the *eLasso* is a binary pairwise Markov random field (or Ising model), as described in Equation A.1.

2. The estimation procedure is based on the fitting of p l_1-regularized logistic regression models, as originally described by Wainwright et al. (2007) and, subsequently, Ravikumar et al. (2010).

3. The model selection process is facilitated by generating, for each of the p vertices, candidate l_1-regularized logistic regression models for different values of the penalty parameter. Following the work of Barber and Drton (2015), model selection is accomplished for each vertex using the EBIC criterion (Chen & Chen, 2008).

4. The focus of *eLasso* emphasizes edge weights more than edge recovery because the principal output, $\mathbf{\Theta}$, is a symmetric weighted matrix. The method for populating $\mathbf{\Theta}$ most closely resembles the description of Höfling and Tibshirani (2009); however, the rule for establishing symmetry is slightly different. van Borkulo et al. (2014) use the AND rule (e.g., an edge is present if both estimates are nonzero) to determine whether the edge weight for the vertex pair $\{s, t\}$ should be nonzero, and then average θ_{st} and θ_{ts} if the decision is affirmative.

van Borkulo et al. (2014, p. 8) describe the *eLasso* method as a four-step process. The first part of Step 1 is the random selection of a variable (say, s) to serve as the dependent variable. The rationale for random selection of the first variable is unclear. It should be possible to take each of the p variables in turn as the dependent variable without the need for any random selection. Step 1a runs l_1-regularized logistic regression for 100 different values of the lasso penalty parameter (ρ). The EBIC for each value of ρ is computed in Step 1b, and a selection based on this index is made in Step 1c. In Step 1d, the intercept of the selected model is placed on the main diagonal of $\mathbf{\Theta}$, and the other l_1-regularized logistic regression coefficients populate the remaining elements in column s of $\mathbf{\Theta}$. Step 1e is to repeat Steps 1a through 1d for all $t \in \{V \backslash s\}$. Step 2 is to define the edge set as follows: $\{s, t\} \in E \Leftrightarrow (\theta_{st} \neq 0 \wedge \theta_{ts} \neq 0)$, where $\wedge$ is the conjunction operator. In other words, an undirected edge is present between variables s and t if and only if both θ_{st} and θ_{ts} are nonzero. Step 3 defines $\theta_{st} = \theta_{ts} = (\theta_{st} + \theta_{ts})/2$ if $\{s, t\} \in E$ based on Step 2; otherwise, $\theta_{st} = \theta_{ts} = 0$. Step 4 creates a graph of the off-diagonal elements of $\mathbf{\Theta}$.

Concerns Specific to *eLasso*

In addition to the general assumptions pertaining to the Ising model, there are two considerations unique to the estimation process associated with *eLasso*. The first of these corresponds to assumptions that are required for consistent estimation. The second, and potentially more problematic, consideration is associated with the need to coerce asymmetric interaction coefficients to symmetry.

Wainwright et al. (2007) and Ravikumar et al. (2010) identified two conditions that are necessary for consistent estimation via the l_1-regularized regression process used by *eLasso*: dependency and incoherence. The implications of the violation of these conditions are analogous to estimation problems that arise in the presence of high correlation among variables in regression analysis (Meinshausen & Bühlmann, 2006). In the logistic regression context, both of these conditions are linked to the Fisher information matrix ($\mathbf{Q}$: the Hessian associated with the conditional log-likelihood function). The first assumption, the *dependency* condition, corresponds to bounded eigenvalues for the submatrix of $\mathbf{Q}$ ($\mathbf{Q}^{\mathrm{R}}$) associated with the relevant variables (those part of the edge set for a given vertex). The goal is to assure that there is not an overdependence among these variables. The second assumption, the *incoherence* condition, focuses on the relationship between $\mathbf{Q}^{\mathrm{R}}$ and the submatrix of $\mathbf{Q}$ ($\mathbf{Q}^{\mathrm{C}}$) corresponding to the irrelevant variables. The focus of the assumption is the assurance that the irrelevant variables do not exert too strong an influence on the relevant ones.

To the best of our knowledge, the dependency and incoherence assumptions have never been mentioned in any applications of *eLasso* in the psychopathological literature. In light of the fact that psychopathology data can have a very large proportion of null rows (i.e., asymptomatic individuals) in the sample, as well as a nontrivial proportion of individuals showing all symptoms, the implications of these properties for the tenability of the model assumptions at least warrants some consideration. In fairness, however, there is not a lot of practical guidance available for assessing whether or not the dependency and incoherence conditions are satisfied. Perhaps as a general guideline, extremely high bivariate or tetrachoric correlations among subsets of variables might give some pause for concern regarding the estimated interaction coefficients.

Regardless of whether the emphasis is on the recovery of edges or edge weights, there is an asymmetry that typically occurs with the use of p separate l_1-regularized logistic regressions. Ideally, the estimates would be such that $\theta_{st} = \theta_{ts}$, yet this seldom occurs. In fact, it is generally recognized in the literature (Barber & Drton, 2015; Höfling & Tibshirani, 2009; Wainwright et al., 2007) that several situations are possible, including (a) θ_{st} and θ_{ts} are nonzero and of the same sign, but of different magnitude; (b) θ_{st} is zero and θ_{ts} is nonzero, or vice versa; and (c) θ_{st} and θ_{ts} are nonzero, but of different signs. Several strategies are available to address the asymmetry issue, and they also differ depending on whether the goal is to estimate the edge set or interaction parameters (edge weights).

If the goal is to estimate the interaction parameters (weighted edges) directly, then two options discussed by Höfling and Tibshirani (2009) seem particularly appropriate for converting $\mathbf{\Theta}$ to a symmetric matrix. The first approach, the MAX rule, resets θ_{st} to θ_{ts} if $|\theta_{ts}| \geq |\theta_{st}|$, and resets θ_{ts} to θ_{st} otherwise. Likewise, the MIN rule resets θ_{st} to θ_{ts} if $|\theta_{ts}| \leq |\theta_{st}|$, and resets θ_{ts} to θ_{st} otherwise. For situations in which the goal is to estimate the edge set, Wainwright et al. (2007) considered two rules: (a) the AND rule, and (b) the OR rule. The AND rule requires both θ_{st} and θ_{ts} to be nonzero (and, presumably, the same sign) to conclude $\{s, t\} \in E$. The OR rule only requires either θ_{st} or θ_{ts} to be nonzero to conclude $\{s, t\} \in E$. The AND and OR rules are also discussed by Xue et al. (2012) and Barber and Drton (2015).

As noted previously, *eLasso* focuses on direct estimation of the edge weights (interaction parameters). The default rule (AND) is precisely the same as the intersection rule described by Xue et al. (2012, p. 1415). The alternative rule (OR) is very similar to the union rule described by Xue (p. 1415). In their simulation experiments, Xue et al. found that the union (OR) rule was generally superior for estimation of the true interaction parameter values, but that the intersection (AND) rule did a better job recovering the true edge set. Similar findings are reported by van Borkulo et al. (2014, p. 2). For the psychopathological researcher using *eLasso*, it is difficult to determine a priori whether the AND or OR rule is preferable for any given application. Therefore, it is our recommendation that *both* solutions be examined. If the OR rule results in the estimation of just a few more edges relative to the AND rule, then this can increase the confidence in the saliency of the estimated graph. Contrastingly, if the OR rule results in many more edges, then confidence in the graph is diminished.

An Illustration of the Cautions Associated With the Ising Model and *eLasso*

Data for Illustration Purposes

To effectively illustrate some of the cautions and concerns when using *eLasso* for estimating the edge weights of a graph, it is helpful to have illustrations using actual data that allow to us to highlight examples of these concerns. For this purpose, we consider a sample of $n = 9{,}282$ individuals measured on the presence or absence of $p = 18$ symptoms of major depression and gener-

alized anxiety, which were taken from the National Comorbidity Survey-Replication (NCS-R). The 18 symptoms are described in Table 1. These data were analyzed by Forbes et al. (2017) in their study pertaining to the limited replicability of symptom dependence graphs. They were also analyzed by Borsboom et al. (2017) and Steinley et al. (2017).

The *eLasso* method was applied to the NCS-R data using the IsingFit software program with the default settings. The estimated edge weights between each pair of the 18 symptoms are provided in Table 2. More specifically, the top panel of Table 2 contains the asymmetric edge weights ('asymm.weights' in the IsingFit output) and the bottom panel contains the symmetric edge weights ('wei-adj' in the IsingFit output) obtained via the AND rule.

As noted previously, psychopathology data may often be better described as a sample from multiple subpopulations rather than random samples from a single underlying distribution. For this reason, it seemed reasonable that potentially better performance might be realized by partitioning the individuals in the sample and then fitting *eLasso* for each of the clusters in the partition. We used the R program known as 'pam' (for partitioning-around-medoids, see Kaufman & Rousseeuw, 2005) to partition the individuals in the NCS-R data into two clusters. This program essentially obtains a partition based on *p*-median clustering (see Köhn et al., 2010, for a review), which recently outperformed both *K*-means and latent class analysis in a comparative study for partitioning observations based on binary measurements (Brusco et al., 2017).

Succinctly, the partitioning process distinguished between individuals with low and high symptom rates. The first cluster (Cluster 1) consisted of $n = 7{,}215$ individuals with comparatively low levels of reported symptoms. The density of the data matrix (e.g., the percentage of the elements of the matrix that are "one") for Cluster 1 was 8% and the average number of symptoms per individual was 1.46. By contrast, the second cluster (Cluster 2) consisted of $n = 2{,}067$ individuals with comparatively high levels of reported symptoms. The density of the data matrix for Cluster 2 was 65% and the average number of symptoms per individual was 11.77. The *eLasso* method was applied to the two clusters of the NCS-R data using the IsingFit software program with the default settings. The estimated edge weights between each pair of the 18 symptoms are provided for Clusters 1 and 2 in Tables 3 and 4, respectively.

Heterogeneity and Sparsity

Though there is some evidence that l_1-regularized regression does well recovering simulated graphs generated from an Ising model (although its performance does deteriorate as a function of degree), there is little compelling evidence that the Ising model is well-suited for identifying good graphical models of psychopathological data. There are two primary reasons for this. First, as noted previously, psychopathology data may not consist of n i.i.d. realizations from the Ising model. Instead, it is likely that they arise from the sampling of multiple subpopulations. Second, psychopathological graphical models are not necessarily sparse.

The extant simulation work, in conjunction with the sample size requirement implications outlined above (see also Ravikumar et al., 2010; Santhanam & Wainwright, 2012), suggest that the Ising model estimation algorithms such as *eLasso* are most conducive to effective graph recovery when dimensionality is high (i.e., p is large), maximum degree (d) is small, the minimum edge weight (λ) is not too small, and the maximum neighborhood size is small (ω). These conditions are not necessarily descriptive of psychopathology data, as evidenced by the high-density graph obtained by *eLasso* for the NCS-R data as shown in the bottom panel of Table 2. For these data ($p = 18$), the maximum degree in the estimated graph was $d = 13$, as shown by the 13 nonzero entries in the row (or column) for S7 in the bottom panel of Table 2. The minimum (absolute) edge weight was $\lambda = .090$ (between S16 and S7), and the maximum neighborhood weight was $\omega = 23.079$ (the sum of the row, or column, for S11 in the bottom panel of Table 2). Assuming a confidence probability of $\delta = .5$, plugging these

Table 1
Symptoms Associated With the NCS-R Study

Symptom number	Symptom type	Symptom abbreviation	Symptom description
1	Major depressive	depr	Depressed mood
2	Major depressive	inte	Loss of interest
3	Major depressive	weig	Weight problems
4	Major depressive	mSle	Sleep problems
5	Major depressive	moto	Psychomotor disturbances
6	Major depressive	mFat	Fatigue
7	Major depressive	repr	Self reproach
8	Major depressive	mCon	Concentration problems
9	Major depressive	suic	Suicidal ideation
10	Generalized anxiety	anxi	Chronic anxiety/worry
11	Generalized anxiety	even	Anxiety about more than one event
12	Generalized anxiety	ctrl	No control over anxiety
13	Generalized anxiety	edge	Feeling on edge
14	Generalized anxiety	gFat	Fatigue
15	Generalized anxiety	gCon	Concentration problems
16	Generalized anxiety	irri	Irritability
17	Generalized anxiety	musc	Muscle tension
18	Generalized anxiety	gSle	Sleep problems

Note. NCS-R = National Comorbidity Survey—Replication.

Table 2
Estimated Edge Weights for the (Full Sample) NCS-R Data

Symptom	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18
S1		.96	2.15	2.99	.90	2.12	.82	2.32	2.69	.51		.21						
S2	2.89		.76	.72	.42	1.17	.75	.95	.63			.08	**.05**			**.23**		
S3	3.86	.77		1.40	.69	.19	.06	.64	.30					**.14**				
S4	4.03	.65	1.38		.54	1.05	.06	.71	.44									.50
S5	2.74	.63	.82	.78		**.15**	.36	1.43		.06			**.21**	−.37	.10	.16	.16	
S6	3.71	1.19	.24	1.15			.38	1.16					**−.17**	1.31		−.16		−.19
S7	2.86	1.01	.25	.32	.38	.60		.42	.98	.25	**.03**	.08		.20		.04	.22	
S8	3.74	.99	.68	.86	1.28	1.18	.20		.33							.97		
S9	4.35	.62	.29	.47			.90	.29		.10	**.03**		**.07**	**.03**			**.11**	**.03**
S10	.55		**.40**		.13		.27		.18		7.54							
S11						**.05**		**.02**		6.51		4.88	1.52	.36	1.16	.61		.95
S12	.11	.24					.14	**.00**	**.04**		6.93		1.00	.52	.47	.41	.31	.48
S13											2.39	1.13		.55	1.95	1.50	.73	1.51
S14	**−.55**			**−.29**	−.33	1.44	.21				2.97	.80	.93		1.24	.85	1.19	.78
S15	**.02**		**.12**		.12	**.02**					3.71	.70	2.09	1.20		1.40	.26	.62
S16	**−.26**				.18	−.19	.14	.97			3.17	.68	1.69	.79	1.43		.29	1.19
S17					.15	**.03**	.26				**2.96**	.66	1.13	1.26	.42	.43		.98
S18				1.03		−.21		**−.35**			3.46	.69	1.71	.72	.65	1.19	.87	
S1		1.93	3.01	3.51	1.82	2.92	1.84	3.03	3.52	.53		.16						
S2	1.93		.76	.69	.52	1.18	.88	.97	.63			.16						
S3	3.01	.76		1.39	.76	.21	.15	.66	.29									
S4	3.51	.69	1.39		.66	1.10	.19	.79	.45									.77
S5	1.82	.52	.76	.66			.37	1.35		.09				−.35	.11	.17	.16	
S6	2.92	1.18	.21	1.10			.49	1.17						1.38		−.18		−.20
S7	1.84	.88	.15	.19	.37	.49		.31	.94	.26		.11		.21		.09	.24	
S8	3.03	.97	.66	.79	1.35	1.17	.31		.31							.97		
S9	3.52	.63	.29	.45			.94	.31		.14								
S10	.53				.09		.26		.14		7.03							
S11										7.03		5.90	1.96	1.67	2.43	1.89		2.20
S12	.16	.16					.11				5.90		1.07	.66	.58	.54	.49	.58
S13											1.96	1.07		.74	2.02	1.60	.93	1.61
S14					−.35	1.38	.21				1.67	.66	.74		1.22	.82	1.23	.75
S15					.11						2.43	.58	2.02	1.22		1.41	.34	.64
S16					.17	−.18	.09	.97			1.89	.54	1.60	.82	1.41		.36	1.19
S17					.16		.24					.49	.93	1.23	.34	.36		.92
S18				.77		−.20					2.20	.58	1.61	.75	.64	1.19	.92	

Note. The top panel contains the asymmetric edge weights as estimated by IsingFit (values in bold found are unreciprocated edges). The bottom panel contains the symmetric edge weights using the AND rule. NCS-R = National Comorbidity Survey—Replication.

values into Equations B.1 and B.2 yields sample size requirements of 8.105×10^{25} and 1.198×10^{52} for the cases of known and unknown edge weights, respectively. Clearly, the issues pertaining to dimensionality, sparseness, and the sample size requirements needed for graph recovery cannot be ignored in psychopathology studies.

Assuming that heterogeneity in the sample is a potential concern, one approach for remedying this problem is the two-step process of partitioning the sample and, subsequently, applying *eLasso* to each cluster independently. For example, partitioning the NCS-R data into two clusters led to one large cluster (Cluster 1) with low symptom rates and one small cluster (Cluster 2) with high symptom rates. The bottom panels of Tables 3 and 4 report the symmetric edge weights for the large and smaller clusters, respectively. The results for Cluster 1 in Table 3 reveal a graph that is much less dense (46 edges, maximum degree of $d = 7$) than the one for the full sample (80 edges, maximum degree of $d = 13$) in Table 2. The graph for Cluster 2 in Table 4 is also less dense (51 edges, maximum degree of $d = 10$). At a minimum, the partitioning process has resulted in *eLasso*'s estimation of graphs for each cluster that exhibit greater sparseness than the one for the full sample, and this is more concordant with the type of graph for which the Ising model is designed.

It should be acknowledged that establishing a partition of the sample and running *eLasso* via IsingFit on each cluster must be implemented with caution. For example, we also generated a five-cluster partition of the NCS-R data, but IsingFit would not run on four of the clusters. The reason is that the partitioning process produced clusters such that some of symptoms would have values of all ones (or all zeros) in the clusters. A potential solution to this problem is to append an identity matrix to the each of the within-cluster data matrices, allowing the estimation of the associated network. This is similar to other, common corrections in categorical data analysis, such as adding a small value to all the cells of a contingency table for ensuring that each team has a win and loss when estimating the Bradley-Terry-Luce model for paired comparisons. Even then, it should be noted that clusters that produce symptom columns that are nearly all ones or zeros yield a warning message, "Nodes with too little variance (not allowed)," resulting in the symptoms that are nearly all ones/zeros not having edge weights and remaining on the

Table 3
Estimated Edge Weights for Cluster 1 for the NCS-R Data

Symptom	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18
S1			2.44	4.35	5.34	3.54	1.86	2.96	4.82									
S2	**2.97**		1.36	.40		2.01	.25	.61	1.34									
S3	4.34	.69		1.44	.32				**.39**									
S4	5.41	.16	1.45			.70		.34										
S5	4.68		.32															
S6	4.68	1.24		.45			.46	.58	**−.34**									
S7	3.66	.28				.50			**.45**									
S8	5.42	.29		.31		.46												
S9	4.66	.31																
S10								**−.12**			6.60							
S11		**−.87**	**−.17**	**−.17**		**−.10**		**−1.32**	**−.90**	7.69		6.02	2.05	1.41	1.17	1.27		2.10
S12	**−.84**										5.80		1.20	.51	.38	.54	.07	.16
S13											3.73	1.31		.82	2.19	1.43	1.11	1.47
S14						**1.85**					3.16	.80	1.20		1.46	.49	1.50	.85
S15											2.97	.61	2.31	1.31		1.49		.74
S16											3.21	.81	1.72	.43	1.58		.67	1.26
S17											**3.08**	.44	1.51	1.51	.02	.74		.83
S18											3.79	.41	1.70	.75	.79	1.20	.73	
S1			3.39	4.88	5.01	4.11	2.76	4.19	4.74									
S2			1.02	.28		1.62	.26	.45	.83									
S3	3.39	1.02		1.44	.32													
S4	4.88	.28	1.44			.57		.33										
S5	5.01		.32															
S6	4.11	1.62		.57			.48	.52										
S7	2.76	.26				.48												
S8	4.19	.45		.33		.52												
S9	4.74	.83																
S10											7.14							
S11										7.14		5.91	2.89	2.29	2.07	2.24		2.94
S12											5.91		1.25	.66	.50	.68	.26	.28
S13											2.89	1.25		1.01	2.25	1.57	1.31	1.58
S14											2.29	.66	1.01		1.39	.46	1.51	.80
S15											2.07	.50	2.25	1.39		1.53		.76
S16											2.24	.68	1.57	.46	1.53		.71	1.23
S17												.26	1.31	1.51		.71		.78
S18											2.94	.28	1.58	.80	.76	1.23	.78	

Note. The top panel contains the asymmetric edge weights as estimated by IsingFit (values in bold found are unreciprocated edges). The bottom panel contains the symmetric edge weights using the AND rule. NCS-R = National Comorbidity Survey—Replication.

periphery of the network. In this instance, the two-cluster solution sufficed and no warnings were produced.

Symmetry of the Interaction Coefficients

The *eLasso* procedure requires p distinct model selection decisions in Step 1 of the algorithm. In addition to the fact that the cumulative effects of suboptimal model selection decisions are apt to increase as a function of p, a serious limitation of the *eLasso* estimation process is the asymmetry of the interaction coefficients. The coefficients generated by *eLasso* are almost always such that $\theta_{st} \neq \theta_{ts}$ and, therefore, the undirected nature of the graph, $G = (V, E)$, along with the symmetric nature of the similarity matrix, $\mathbf{\Theta}$, must be forced by procedural steps. At this point, it is interesting to consider the possibility that symmetry should not be forced. That is, the *eLasso* output could be used to provide a directed graph without any ad hoc refinement to symmetry (see, e.g., Bryant et al., 2017; McNally et al., 2015). Asymmetry might be especially advantageous for applied researchers in circumstances in which the goal is to establish some type of causal relationship between symptoms. Causality almost certainly requires a directional relationship. Here, we maintain the focus on symmetric relations because that is the norm in the literature; however, asymmetry is an interesting possibility to consider for the future.

To convert an asymmetric similarity matrix to symmetry, several rules are applied. As noted previously, if $\theta_{st} \neq 0$ and $\theta_{ts} \neq 0$, then *eLasso* uses the average of these coefficients to set the edge weights. However, averaging does not really assess the adequacy of the coefficients. If $\theta_{st} = 0.8$ and $\theta_{ts} = 0.1$, then it is quite possible that one of the two is actually the more appropriate measure of tie strength between s and t and, therefore, averaging would not be the appropriate remedy. It is even possible that θ_{st} and θ_{ts} could have different signs. Second, van Borkulo et al. (2014) employ a conservative (AND) rule, whereby if either $\theta_{st} = 0$ or $\theta_{ts} = 0$, then the edge $\{s, t\} \notin E$. This rule will tend to produce greater sparseness in the final symmetric matrix, $\mathbf{\Theta}$, yet there is no guarantee that the zero coefficient was the more appropriate of the two coefficients. The bottom line is that any researcher using *eLasso* needs to pay close attention to the sym-

Table 4
Estimated Edge Weights for Cluster 2 for the NCS-R Data

Symptom	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18
S1																		
S2						.08	.61	.34										
S3				.52	.24													
S4		**.27**	.77		.44	.63	.19	.29	**.35**	**−.36**	**−.13**							.63
S5		**.10**	.32	.25			.22	.81								.06		
S6		.38		.65			.43	.54	**−.18**	**−.42**	**−.08**	**−.10**	**−.28**	1.45		−.08		−.18
S7	**1.35**	.99	**.19**	.22	.36	.51		.47	.94	**.37**	**.02**	**.09**		.20		.02	.20	
S8		.52	**.32**	.42	1.10	.68	.51			−.78						1.43		−.37
S9							.70											
S10								−.23			5.12							
S11			**−.27**							3.88		3.47	1.00		.37	.62		
S12	**.17**										3.95		.71	.47	.51	.23	.34	.85
S13											.82	.78		.26	1.69	1.54	.42	1.54
S14		**.19**	**.28**	**−.21**	**−.25**	1.68	.20				**1.75**	.83	.60		1.04	1.15	.95	.74
S15											1.24	.69	1.83	1.00		1.30	.57	.43
S16		**.38**	**.16**		.22	−.31	.16	1.51			2.72	.51	1.63	1.10	1.34			1.19
S17							.10					.71	.76	1.01	.69	**.17**		1.05
S18	**1.73**	**−.14**		1.21		−.27		−.57			**1.75**	1.11	1.77	.73	.47	1.22	1.00	
S1																		
S2						.23	.80	.43										
S3				.65	.28													
S4			.65		.34	.64	.21	.35										.92
S5			.28	.34			.29	.96								.14		
S6		.23		.64			.47	.61						1.56		−.20		−.23
S7		.80		.21	.29	.47		.49	.82					.20		.09	.15	
S8		.43		.35	.96	.61	.49			−.51						1.47		−.47
S9							.82											
S10								−.51			4.50							
S11										4.50		3.71	.91		.80	1.67		
S12											3.71		.75	.65	.60	.37	.53	.98
S13											.91	.75		.43	1.76	1.58	.59	1.66
S14						1.56	.20					.65	.43		1.02	1.13	.98	.73
S15											.80	.60	1.76	1.02		1.32	.63	.45
S16					.14	−.20	.09	1.47			1.67	.37	1.58	1.13	1.32			1.20
S17							.15					.53	.59	.98	.63			1.03
S18				.92		−.23		−.47				.98	1.66	.73	.45	1.20	1.03	

Note. The top panel contains the asymmetric edge weights as estimated by IsingFit (values in bold found are unreciprocated edges). The bottom panel contains the symmetric edge weights using the AND rule. NCS-R = National Comorbidity Survey—Replication.

metry of $\mathbf{\Theta}$. If θ_{st} and θ_{ts} are comparable in sign and magnitude, then there is greater confidence in taking an average of these values. However, sign reversals or large discrepancies in magnitude should signal that careful consideration of the findings is necessary.

To illustrate the issues pertaining to asymmetry, we return to the results for NCS-R data in Table 2. There were 26 *unreciprocated* edge weights estimated by *eLasso* for these data, which are highlighted in bold font in the top panel of Table 2. As an example, $\theta_{2,16} = .23$ but $\theta_{16,2} = 0$. The implication of the unreciprocated edges is that, if the OR rule had been applied instead of the AND rule, the number of estimated edges would increase by more than 30% (from 80 to 106). In the 80 instances in which there was reciprocity in the estimation of the edge weights, the disparity with respect to magnitude varied. In some instances, such as in the case of $\theta_{23} = .76$ and $\theta_{32} = .77$, there was tremendous consistency and averaging to produce symmetry presents no concerns. In other circumstances, such as $\theta_{11,14} = .36$ and $\theta_{14,11} = 2.97$, there is much less confidence in an averaging process.

Another critical aspect pertaining to the issue of symmetry is that, because the *eLasso* is based on logistic regression, the beta weights can be thought of as functions of the odds ratio between two binary variables. However, the averaging of odds ratios (e.g., beta weights) is unlikely the best measure when the symptoms exhibit such disparate base rates. Assume that there are two symptoms (say, *S1* and *S2*) with different base rates. It is unreasonable to assume that $P(S1|S2) = P(S2|S1)$; that is, the chances of someone having *S1* given they have *S2* is not likely the same as the probability of someone having *S2* given they have *S1*. The likeliness of the equality of these conditional probabilities decreases as $P(S1)$ and $P(S2)$ become more disparate. In fact, from Bayes's theorem (Bayes, 1764), the only way for those two conditional probabilities to be equal is if the two had the same base rate. We know from the psychopathological literature that this is generally not the case, as the scales that measure the items that comprise the symptoms are designed to cover a broad range of the latent trait (e.g., some symptoms are purposely chosen to be rare and some symptoms are purposely chosen to be common).

The final issue when creating symmetric edge weights from the estimated, asymmetric weights is to note that the original estimates are biased due to the *eLasso* procedure. Further, the regularization

parameter for the *eLasso* is estimated separately in each of the p logistic regressions. Consequently, the averaging process is combining biased parameters whose individual biases were created via different processes, leaving the final direction and degree of bias completely unknown.

A Comparison With a Co-Occurrence-Based Graph

We also investigated the potential differences that might stem from using a co-occurrence measure of similarity that excludes zero matches (i.e., symptom absence matches). The first step in this process was to obtain a similarity matrix consisting of Jaccard (1901) indices for each pair of symptoms. For each pair of symptoms, the Jaccard index is computed by dividing the total number of matches of ones (i.e., symptom presence matches) by the following quantity: total sample size (n) minus the number of zero matches. In the second step, the edges with the 80 largest Jaccard indices were selected for the graph. We selected 80 for direct comparison with the 80-edge IsingFit solution, while acknowledging that better methods for choosing the number of edges should be developed. The matrix of edge weights (Jaccard indices) is shown in Table 5.

The comparison of the edges in Jaccard graph in Table 5 with the IsingFit graph in Table 2 reveals some similarities. For example, a complete subgraph for anxiety symptoms S12 to S18 is present in both graphs.[5] The Jaccard graph in Table 5 shows another complete subgraph occurring among the depression symptoms S1 to S9. However, a complete subgraph among these symptoms is not present in the IsingFit solution in Table 2 because of the absence of edges among all pairs associated with Items S5, S6, and S9. The exclusion of these edges from a graphical model is difficult to justify, particularly in light of the fact that their Jaccard indices exceed those of many other selected pairs.

The Jaccard graph in Table 5 also establishes a clearer and more compelling bridge between depression and anxiety symptoms than the IsingFit graph in Table 2. This bridge is formed by symptoms S10, S11, and S12. These three symptoms (especially S11 and S12) provide a channel for linking the specific anxiety symptoms to depression symptoms S1, S2, S4, S6, and S8. Moreover, when comparing symptom subsets S1 to S9 and S13 to S18, there are only two edges ({S8, S13} and {S8, S16}) in the Jaccard graph connecting these two subsets to one another. By contrast, Table 2 shows that the IsingFit solution establishes 12 edges between these two subsets of symptoms. However, nine of these edges have small edge weights (less than .35 in absolute magnitude) and three of these nine edge weights are negative. In our opinion, these edges are likely a spurious artifact of the graph estimation process.

Both the results from the Ising model and the co-occurrence matrix of Jaccard coefficients are consistent with what would be expected from a two-factor model, in which the factors were correlated. Namely, there are nine depression items and nine anxiety items; if each set of items arose from a common factor, they should have 36 edges each, for a total of 72 edges (this would correspond to a complete graph for depression and a complete graph for anxiety). In comparison, the final network from the Ising model had 80 edges. Of those 80 edges, 33 were in the set of depression items (rather than 36 if it were a complete graph), 27 were in the set of anxiety items (again, 36 would be expected if it were a complete graph), and 20 were edges that fell between the two sets of items. These latter "bridging" items can be thought of as capturing the correlation that would be expected between the "depression" factor and the "anxiety" factor. Prima facie, the patterning in the weight matrix for the Ising model reflects what would be expected from a two-factor model, with correlations. For the matrix of Jaccard coefficients, the structure is more prominent—the depression items reflect a complete graph, with 36 out of 36 edges; the anxiety items have a near complete graph, with 31 out of 36 edges; the remaining 13 indicate the items driving the correlation between the two sets of items. In any case, the same general conclusion is reached—depression and anxiety are likely represented by common factors that are moderately correlated.

Using the same interpretation for the two clusters identified in the NCS-R data, the first cluster reflects individuals that perfectly segregate the items into "depression" and "anxiety" without any bridging edges (note, this would correspond to a two-factor model with uncorrelated factors). The second cluster has weaker connections within the depression and anxiety sets of items and more edges between the two sets.

A Second Example

The second example used the NESARC (Grant et al., 2004). NESARC assessed a very large, nationally representative sample that is racially diverse, with an oversampling of Blacks and Hispanics (25% Hispanics, 19% Blacks), as well as geographically diverse. For this example, we used Wave 2 data and focused on questions used to assess alcohol use disorders as assessed under the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; American Psychiatric Association, 2013). We also excluded those who reported abstaining from alcohol use in the past year, resulting in a final sample of size $N = 22{,}177$. The 11 criteria used to create the symptom network were *tolerance* (increased tolerance of alcohol), *cut down* (inability to cut down on use), *larger/longer* (used more alcohol than intended), *time spent* (spent a lot of time using or recovering from alcohol), *give up* (important activities given up because of use), *continue* (continued use despite physical or psychological problems), *withdrawal* (withdrawal after using alcohol), *role-interference* (failure to fulfill roles/obligations), *hazardous use* (use in physically hazardous situations), *social problems* (use despite social/interpersonal problems), and *craving* (craving or a strong desire or urge to use alcohol).

As with the NCS-R example, the data are arranged in a binary matrix that we analyze using the *eLasso* program. Given that the results are parallel to those of the NCS-R data, we provide a description of the results; however, we do not provide the full results in tabular form. The maximum degree of the graph for the 11-item set is $d = 10$ for both *time spent* and *continue*. The minimum (absolute) edge weight was $\lambda = .257$ (between *continue* and *tolerance*), and the maximum neighborhood weight was $\omega = 10.08$ (the sum of the row, or column, for *larger/longer*). Assuming a confidence probability of $\delta = .5$, plugging these values into Equations B.1 and B.2 yields sample size requirements of 3.359×10^{13} and 7.617×10^{26} for the cases of known and unknown edge

[5] A complete subgraph for symptoms S12 through S18 is evident from the fact that the submatrices in Tables 2 and 5 that correspond to these seven symptoms have nonzero elements for all pair of symptoms. That is, an edge exists between every pair of these seven symptoms.

Table 5
Jaccard Index Edge Weights for the NCS-R Data

Symptom	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18
S1		.826	.818	.884	.489	.821	.421	.857	.671	.356	.362	.357						
S2	.826		.751	.793	.498	.775	.449	.795	.642		.328	.340						
S3	.818	.751		.798	.500	.729	.417	.769	.617									
S4	.884	.793	.798		.500	.792	.426	.818	.647	.328	.338	.340						
S5	.489	.498	.500	.500		.478	.376	.523	.427									
S6	.821	.775	.729	.792	.478		.434	.790	.605		.324	.331						
S7	.421	.449	.417	.426	.376	.434		.435	.451									
S8	.857	.795	.769	.818	.523	.790	.435		.638		.337	.346	.326			.332		
S9	.671	.642	.617	.647	.427	.605	.451	.638										
S10	.356			.328							.781	.515	.343					
S11	.362	.328		.338		.324		.337		.781		.659	.439	.340	.399	.391		.379
S12	.357	.340		.340		.331		.346		.515	.659		.568	.451	.518	.509	.351	.494
S13								.326		.343	.439	.568		.613	.737	.718	.497	.691
S14											.340	.451	.613		.619	.600	.513	.583
S15											.399	.518	.737	.619		.691	.477	.637
S16								.332			.391	.509	.718	.600	.691		.477	.656
S17												.351	.497	.513	.477	.477		.497
S18											.379	.494	.691	.583	.637	.656	.497	

Note. NCS-R = National Comorbidity Survey—Replication.

weights, respectively. To put these sample size requirements in context, they are greater than the total number of individuals (1.08×10^{11}) who have ever lived on Earth.

In fact, the overall estimated network is quite dense, with 47 out of the 55 (85%) possible pairwise edges being estimated as "present." As described in the introduction, this is likely a reflection of AUD being better fit by a unidimensional factor model. Indeed, although the specific ordering of items on the latent trait has some variability, it is common to find that a single-factor model fits the data well (Lane, Steinley, & Sher, 2016). In that regard, the results are very similar to what was found in the NCS-R data, except here there is only one factor, whereas the prior data set had two distinct factors.

As with the NCS-R data set, we explored whether partitioning the data would result in more sparse graphs. In this case, the $K = 2$ cluster solution results in the first cluster having $n = 19{,}365$ observations and the second cluster having 2,812 observations. The respective densities of the two clusters are 2.8% and 29%. In the process of creating the clusters, there was perfect separation on one of the variables (e.g., *cut down*), in which the first cluster contains no individuals endorsing that symptom and, in the second cluster, every individual endorsing that symptom.

In this case, we explore using the corrective approach of appending an 11×11 identity matrix to each of the within-cluster data sets. First, we tested the sensitivity of the parameter estimates to such an approach by appending an 11×11 identity matrix to the full NESARC data set. A maximum change of .00013 occurred for the interaction between *time spent* and *role-interference*. From this example and prior experience, we do not believe that this corrective factor alters the resultant structure or interpretation of any network/graph that is derived pre- versus postcorrection.

Again, the clustering process results in subsequent within-cluster networks that are sparser than the original network on the entire data. The maximum degree is $d = 9$ for both the first and second cluster and the minimum edge weights are now .281 and .192 for the first and second clusters, respectively. The maximum neighborhood weight for the first cluster was virtually unchanged at 10.26, whereas there was a substantial decrease for the second cluster ($\omega = 6.79$).

One oddity that is related to the issue of how to measure the similarity between symptoms is that the cluster with the fewest symptom endorsements has the strongest edge weights, yet the cluster with the most symptom endorsements has substantially weaker edge weights. In fact, the first cluster has a data set that is 14.5 times sparser than the second cluster; however, its average edge weight is 49% stronger. Clearly, there is a disconnect between rate of endorsement and strength of connection between symptoms. This same phenomenon was observed in the NCS-R example, in which, comparing the bottom panels of Tables 3 and 4, it is evident that the edge weights for the cluster of individuals with comparatively low levels of reported symptoms (Cluster 1) are frequently much larger than those for the cluster of individuals with comparatively high levels of reported symptoms (Cluster 2).

Conclusions

Summary

Ising models have recently been applied in social network applications pertaining to opinion formation and technology diffusion. A particular type of Ising model based on binary pairwise Markov random fields has received considerable attention with respect to applications in psychopathology. In part, this is attributable to the availability of the *eLasso* program, which uses l_1-regularized logistic regression for estimation. The output of the *eLasso* program has commonly been identified as a "network"; however, a more descriptive term for the output under study is a "dependence graph."

To date, little concern has been reflected in the psychopathology literature with respect to the data assumptions and conditions that underlie the Ising model and/or the methods designed for its estimation. First, little attention has been paid to the issue of

sparseness, which has commonly been assumed in the statistical literature. Although not an assumption of the Ising model per se, sparseness has profound implications for the sample size necessary for graph recovery, when expressed as a function of maximum degree, maximum neighborhood size, and minimum edge weight.

Second, we note that the Ising model assumes the data are generated from a single underlying distribution rather than obtained from a sample of multiple subpopulations. Although there is the potential for the data to be meaningfully separated into known subpopulations based on prior knowledge of certain characteristics (Kossakowski et al., 2016; van Borkulo et al., 2015), our concern centers on the type of unobserved heterogeneity addressed in other multivariate statistical contexts (Ashby et al., 1994; Brusco & Cradit, 2005; DeSarbo & Cron, 1988; Jedidi et al., 1997; Sarstedt & Ringle, 2010). In light of the potential for unobserved heterogeneity, we evaluated a two-step process of *p*-median partitioning followed by the application of *eLasso* to each cluster. The two-step approach has both advantages and disadvantages. Partitioning can substantially reduce the level of heterogeneity and, therefore, circumvent that perils of averaging across samples from diverse subpopulations (Ashby et al., 1994; Brusco & Cradit, 2005; M. D. Lee & Pope, 2003). Moreover, in some circumstances, partitioning can produce large clusters of respondents with an obvious graphical structure that precludes the need to apply an algorithm such as *eLasso*. For example, a five-cluster *p*-median partition of the NCS-R data resulted in one sizable cluster in which each respondent exhibited most if not all of the symptoms. The obvious graphical structure for this cluster is a complete graph. Moreover, there was a very large cluster in which the vast majority of respondents exhibited no symptoms, such that the obvious graphical structure was a null graph. At the same time, partitioning has some serious drawbacks: (a) the inherent limitation of a process that requires two distinct steps, and (b) the observed potential for the infeasibility at the second step. For this reason, we recommend the development of mixture modeling approaches that can simultaneously estimate latent classes of respondents and fit graphical models for each class.

Third, it is noted that *eLasso* can perhaps best be characterized as one of many available approaches for collapsing the binary two-mode data into a one-mode, $p \times p$ similarity matrix among symptoms. Like other binary similarity coefficients based on covariance or correlation between symptoms (e.g., Pearson correlation, tetrachoric correlation, Yule's Q), *eLasso* considers matches of zeros (absence) between pairs of symptoms of equal importance to matches of ones (presence). Alternatively, there are many available coefficients that ignore or reduce the influence of matches of zeros (Jaccard, 1901; Kulczynski, 1927; Ochiai, 1957). Although there are arguments both for and against the inclusion of zero matches, the issue is one that psychopathological researchers should consider when selecting a similarity measure. To provide a preliminary investigation into this issue, we reported some graphical results for the Jaccard index for comparison to the results obtained by IsingFit for the NCS-R data. Both the Jaccard and IsingFit graphs were somewhat concordant with a two-factor model. However, the structure associated with the Jaccard graph was more pronounced and avoided the small and/or negative edge weights in the IsingFit graph that are quite possibly spurious. Another potential question that stemmed from this analysis is, are there really inherent advantages to these graphical approaches relative to what can be accomplished using a common factor model?

There are also concerns that pertain uniquely to *eLasso*, such as the need for multiple model selection decisions and the dependency and incoherence assumptions recognized by Wainwright et al. (2007). Of even greater concern is the need to pay careful attention to the asymmetry in the estimates of the edge weights that arise from the p l_1-regularized logistic regressions that are required with *eLasso*.

Limitations and Extensions

The purpose of this critique was to identify concerns associated with a particular model (binary pairwise Markov random fields—Ising) and method (*eLasso*), which have been widely used in the psychopathological literature in the absence of any critical evaluation of the assumptions and conditions that underlie their usage. Our critique is not designed to imply that published results obtained using *eLasso* are "wrong" or "misleading." To do so would require substantive knowledge that we do not possess. Moreover, it is important to clarify that our goal is not to deter researchers from pursuing the development of graphical models for psychopathological research. Indeed, there are many benefits to a parsimonious graphical representation. The refinement and development of binary graphical models for psychological applications is a worthwhile avenue for future research. In particular, models that capture heterogeneity in the sample are especially promising (J. Guo et al., 2015).

In addition, we believe that alternative models for the analysis of two-mode binary data merit consideration. These methods, which preserve information pertaining to both the symptoms and the individuals in the sample, can be divided into three categories: (a) permutation methods, (b) two-mode partitioning or blockmodeling methods, and (c) factorization methods.

Permutation methods seek to resequence the rows and columns of the binary data matrix so as to provide a better visual display of the patterning of zeros and ones in **X**. A common permutation criterion is the bond-energy index (McCormick, Schweitzer, & White, 1972), and methods based on this criterion (or similar criteria) have been developed by Arabie, Hubert, and Schleutermann (1990) and Brusco and Steinley (2006). Given that the number of symptoms is typically modest, a resequenced matrix obtained by a permutation method can provide a visualization of the data that might enable researchers to better understand the structure of the data. At the same time, permutation methods do not explicitly provide a simplified representation of the data and are commonly augmented with other (typically partitioning) approaches.

Two-mode partitioning methods seek to simultaneously establish K clusters of individuals and L clusters of variables. In the case of a binary affiliation network, two-mode partitioning is closely associated with *two-mode deterministic blockmodeling based on structural equivalence* (Doreian, Batagelj, & Ferligoj, 2004, 2005; Lorrain & White, 1971). When applying this approach to psychopathological data, the goal would be to partition the individuals into K clusters and symptoms into L clusters such that the submatrices formed by the clusters of individuals and symptoms are either complete (all ones) or null (all zeros) to the greatest extent possible. Essentially, this process can be used to provide a sim-

plified $K \times L$ representation (with an emphasis on clusters of symptoms and individuals) of the original $n \times p$ matrix. A variety of metaheuristics have been developed for deterministic two-mode blockmodeling (Brusco & Doreian, 2015; Brusco, Doreian, Lloyd, & Steinley, 2013; Brusco & Steinley, 2007, 2011).

A potential criticism of deterministic two-mode blockmodeling is that it imposes a discrete structure on the model-fitting process. By contrast, *stochastic blockmodeling* allows for a probabilistic, model-based approach to the fitting process. Building on earlier work in the sociological literature that framed blockmodeling within the scope of random graphs (Holland, Laskey, & Leinhardt, 1983; Holland & Leinhardt, 1976, 1977, 1981), Nowicki and Snijders (2001) provided the foundation for model-based estimation in stochastic blockmodeling. Subsequently, there were a number of other contributions pertaining to stochastic blockmodeling (Airoldi, Blei, Fienberg, & Xing, 2008; Bickel & Chen, 2009; Handcock, Raftery, & Tantrum, 2007), and a good review of the literature up through the early 2000s is provided by Goldenberg, Zheng, Fienberg, and Airoldi (2010, Section 3.8). More recently, based on the pioneering work of Hoff, Raftery, and Handcock (2002), there has been a renewed emphasis on continuous, latent space representations in stochastic blockmodeling (Sussman, Tang, Fishkind, & Priebe, 2012; Suwan et al., 2016).

There are a variety of alternative matrix factorization approaches for analyzing two-mode binary data. Perhaps the most straightforward approach is a singular value decomposition of $\mathbf{X}$, whereby a low-rank approximation based on eigenvectors from $\mathbf{X}\mathbf{X}'$ and $\mathbf{X}'\mathbf{X}$ is obtained (Eckart & Young, 1936; Everett & Borgatti, 2013). A closely related approach is nonnegative matrix factorization (Brusco, 2011; D. D. Lee & Seung, 1999, 2001; Fogel, Hawkins, Beecher, Luta, & Young, 2013), which forces the factors to have only nonnegative values, thereby avoiding improvement in fit via cancellations and preserving a "sum of the parts" interpretation. Both singular value decomposition and nonnegative matrix factorization assure a continuous representation of individuals and symptoms. There are also factorization methods for two-mode binary data that emphasize a discrete structure based on Boolean vectors and lattices (Chubb, 1986; Pattison, 1993; Pattison & Bartlett, 1982; Pattison & Breiger, 2002).

In summary, there are several alternative methodological approaches for analyzing two-mode binary psychopathological data. These approaches should not be viewed as *substitutes* for IsingFit because they have different goals and constraints. However, they are viable competitive modeling paradigms that warrant serious consideration by researchers in light of the inherent problems with IsingFit and its implementation in psychopathological research. Accordingly, we see two important directions for future research. One avenue is the development of better graphical models and methods for binary data. Such models and methods should afford remedies for the problems that plague IsingFit (e.g., ignoring unobserved heterogeneity in the sample, assuming that symptom absence matches and symptom presence matches are of equal importance, forcing symmetry in an ad hoc manner). The second avenue is the pursuit of the alternative modeling approaches that address different, yet important, research questions.

References

Afzali, M. H., Sunderland, M., Teesson, M., Carragher, N., Mills, K., & Slade, T. (2017). A network approach to the comorbidity between posttraumatic stress disorder and major depressive disorder: The role of overlapping symptoms. *Journal of Affective Disorders, 208,* 490–496. http://dx.doi.org/10.1016/j.jad.2016.10.037

Airoldi, E. M., Blei, D. M., Fienberg, S. E., & Xing, E. P. (2008). Mixed membership stochastic blockmodels. *Journal of Machine Learning Research, 9,* 1981–2014.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Anandkumar, A., Tan, V. Y. F., Huang, F., & Willsky, A. S. (2012). High-dimensional structure estimation in Ising models: Local separation criteria. *Annals of Statistics, 40,* 1346–1375. http://dx.doi.org/10.1214/12-AOS1009

Arabie, P., Hubert, L., & Schleutermann, S. (1990). Blockmodels from the bond energy approach. *Social Networks, 12,* 99–126. http://dx.doi.org/10.1016/0378-8733(90)90001-P

Ashby, F. G., Maddox, W. T., & Lee, W. W. (1994). On the dangers of averaging across subjects when using multidimensional scaling or the similarity-choice model. *Psychological Science, 5,* 144–151. http://dx.doi.org/10.1111/j.1467-9280.1994.tb00651.x

Banerjee, O., El Ghaoui, L., & d'Aspremont, A. (2008). Model selection through sparse maximum likelihood estimation. *Journal of Machine Learning Research, 9,* 485–516.

Barber, R. F., & Drton, M. (2015). High dimensional Ising model selection with Bayesian information criteria. *Electronic Journal of Statistics, 9,* 567–607. http://dx.doi.org/10.1214/15-EJS1012

Bayes, T. (1764). An essay toward solving a problem in doctrine of chances. *Philosophical Transactions of the Royal Society of London, 53,* 370–418.

Bekhuis, E., Schoevers, R. A., van Borkulo, C. D., Rosmalen, J. G. M., & Boschloo, L. (2016). The network structure of major depressive disorder, generalized anxiety disorder and somatic symptomatology. *Psychological Medicine, 46,* 2989–2998. http://dx.doi.org/10.1017/S0033291716001550

Bickel, P. J., & Chen, A. (2009). A nonparametric view of network models and Newman-Girvan and other modularities. *Proceedings of the National Academy of Sciences of the United States of America, 106,* 21068–21073. http://dx.doi.org/10.1073/pnas.0907096106

Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of Clinical Psychology, 64,* 1089–1108. http://dx.doi.org/10.1002/jclp.20503

Borsboom, D., & Cramer, A. O. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology, 9,* 91–121. http://dx.doi.org/10.1146/annurev-clinpsy-050212-185608

Borsboom, D., Fried, E. I., Epskamp, S., Waldorp, L. J., van Borkulo, C. D., van der Maas, H. L. J., & Cramer, A. O. 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,* 989–999. http://dx.doi.org/10.1037/abn0000306

Boschloo, L., van Borkulo, C. D., Rhemtulla, M., Keyes, K. M., Borsboom, D., & Schoevers, R. A. (2015). The network structure of symptoms of the diagnostic and statistical manual of mental disorders. *PLoS ONE, 10,* e0137621. http://dx.doi.org/10.1371/journal.pone.0137621

Brown, N., Park, S., Steinley, D., & Davis-Stober, C. P. (2018). Modeling between-subject variability in decision strategies via statistical clustering: A p-median approach. *Journal of Behavioral Decision Making, 31,* 250–264. http://dx.doi.org/10.1002/bdm.1957

Brusco, M. (2011). Analysis of two-mode network data using nonnegative matrix factorization. *Social Networks, 33,* 201–210. http://dx.doi.org/10.1016/j.socnet.2011.05.001

Brusco, M. J., & Cradit, J. D. (2005). ConPar: A method for identifying groups of concordant subject proximity matrices for subsequent multidimensional scaling analyses. *Journal of Mathematical Psychology, 49,* 142–154. http://dx.doi.org/10.1016/j.jmp.2004.11.004

Brusco, M. J., Cradit, J. D., Steinley, D., & Fox, G. L. (2008). Cautionary remarks on the use of clusterwise regression. *Multivariate Behavioral Research, 43,* 29–49. http://dx.doi.org/10.1080/00273170701836653

Brusco, M., & Doreian, P. (2015). A real-coded genetic algorithm for two-mode *KL*-means partitioning with application to homogeneity blockmodeling. *Social Networks, 41,* 26–35. http://dx.doi.org/10.1016/j.socnet.2014.11.007

Brusco, M., Doreian, P., Lloyd, P., & Steinley, D. (2013). A variable neighborhood search method for a two-mode blockmodeling problem in social network analysis. *Network Science, 1,* 191–212. http://dx.doi.org/10.1017/nws.2013.5

Brusco, M. J., Shireman, E., & Steinley, D. (2017). A comparison of latent class, *K*-means, and *K*-median methods for clustering dichotomous data. *Psychological Methods, 22,* 563–580. http://dx.doi.org/10.1037/met0000095

Brusco, M., & Steinley, D. (2006). Inducing a blockmodel structure for two-mode binary data using seriation procedures. *Journal of Mathematical Psychology, 50,* 468–477. http://dx.doi.org/10.1016/j.jmp.2006.05.005

Brusco, M., & Steinley, D. (2007). A variable neighborhood search method for generalized blockmodeling of two-mode binary matrices. *Journal of Mathematical Psychology, 51,* 325–338. http://dx.doi.org/10.1016/j.jmp.2007.07.001

Brusco, M., & Steinley, D. (2011). A tabu-search heuristic for deterministic two-mode blockmodeling of binary network matrices. *Psychometrika, 76,* 612–633. http://dx.doi.org/10.1007/s11336-011-9221-9

Bryant, R. A., Creamer, M., O'Donnell, M., Forbes, D., McFarlane, A. C., Silove, D., & Hadzi-Pavlovic, D. (2017). Acute and chronic posttraumatic stress symptoms in the emergence of posttraumatic stress disorder: A network analysis. *Journal of the American Medical Association Psychiatry, 74,* 135–142. http://dx.doi.org/10.1001/jamapsychiatry.2016.3470

Chen, J., & Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika, 95,* 759–771. http://dx.doi.org/10.1093/biomet/asn034

Choi, S.-S., Cha, S.-H., & Tappert, C. C. (2010). A survey of binary similarity and distance measures. *Systemics, Cybernetics, and Informatics, 8,* 43–48.

Chubb, C. (1986). Collapsing binary data for algebraic multidimensional representation. *Journal of Mathematical Psychology, 30,* 161–187. http://dx.doi.org/10.1016/0022-2496(86)90012-X

Cipra, B. (1987). An introduction to the Ising model. *The American Mathematical Monthly, 94,* 937–959. http://dx.doi.org/10.1080/00029890.1987.12000742

Cramer, A. O., van Borkulo, C. D., Giltay, E. J., van der Maas, H. L., Kendler, K. S., Scheffer, M., & Borsboom, D. (2016). Major depression as a complex dynamic system. *PLoS ONE, 11,* e0167490. http://dx.doi.org/10.1371/journal.pone.0167490

DeSarbo, W. S., & Cron, W. L. (1988). A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification, 5,* 249–282. http://dx.doi.org/10.1007/BF01897167

DeSarbo, W. S., Manrai, A. K., & Manrai, L. A. (1994). Latent class multidimensional scaling: A review of recent developments in the marketing and psychometric literature. In R. P. Bagozzi (Ed.), *Advanced methods of marketing research* (pp. 190–222). Cambridge, MA: Blackwell.

Doreian, P., Batagelj, V., & Ferligoj, A. (2004). Generalized blockmodeling of two-mode network data. *Social Networks, 26,* 29–53. http://dx.doi.org/10.1016/j.socnet.2004.01.002

Doreian, P., Batagelj, V., & Ferligoj, A. (2005). *Generalized blockmodeling*. Cambridge, UK: Cambridge University Press.

Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika, 1,* 211–218. http://dx.doi.org/10.1007/BF02288367

Epskamp, S., Borsboom, D., & Fried, E. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods, 50,* 195–212. http://dx.doi.org/10.3758/s13428-017-0862-1

Everett, M. G., & Borgatti, S. P. (2013). The dual-projection approach for two-mode networks. *Social Networks, 35,* 204–210. http://dx.doi.org/10.1016/j.socnet.2012.05.004

Faith, D. P. (1983). Asymmetric binary similarity measures. *Oecologia, 57,* 287–290. http://dx.doi.org/10.1007/BF00377169

Fogel, P., Hawkins, D. M., Beecher, C., Luta, G., & Young, S. S. (2013). A tale of two matrix factorizations. *The American Statistician, 67,* 207–218. http://dx.doi.org/10.1080/00031305.2013.845607

Forbes, M. K., Wright, A. G. C., Markon, K. E., & Krueger, R. F. (2017). Evidence that psychopathology symptom networks have limited replicability. *Journal of Abnormal Psychology, 126,* 969–988. http://dx.doi.org/10.1037/abn0000276

Galam, S. (1997). Rational group decision making: A random field Ising model at T = 0. *Physica A, 238,* 66–80. http://dx.doi.org/10.1016/S0378-4371(96)00456-6

Galam, S., & Moscovici, S. (1991). Towards a theory of collective phenomena: Consensus and attitude changes in groups. *European Journal of Social Psychology, 21,* 49–74. http://dx.doi.org/10.1002/ejsp.2420210105

Gehrlein, W. V. (2002). Condorcet's paradox and the likelihood of its occurrence: Different perspectives on balanced preferences. *Theory and Decision, 52,* 171–199. http://dx.doi.org/10.1023/A:1015551010381

Goldenberg, A., Zheng, A., Fienberg, S. E., & Airoldi, E. M. (2010). A survey of statistical network models. *Foundations and Trends in Machine Learning, 2,* 129–233. http://dx.doi.org/10.1561/2200000005

Grabowski, A., & Kosinski, R. (2006). Ising-based model of opinion formation in a complex network of interpersonal interactions. *Physica A, 361,* 651–664. http://dx.doi.org/10.1016/j.physa.2005.06.102

Grant, B. F., Dawson, D. A., Stinson, F. S., Chou, S. P., Dufour, M. C., & Pickering, R. P. (2004). The 12-month prevalence and trends in DSM-IV alcohol abuse and dependence: United States, 1991–1992 and 2001–2002. *Drug and Alcohol Dependence, 74,* 223–234. http://dx.doi.org/10.1016/j.drugalcdep.2004.02.004

Guo, J., Cheng, J., Levina, E., Michailidis, G., & Zhu, J. (2015). Estimating heterogeneous graphical models for discrete data with an application to roll call voting. *The Annals of Applied Statistics, 9,* 821–848. http://dx.doi.org/10.1214/13-AOAS700

Guo, J., Levina, E., Michailidis, G., & Zhu, J. (2010). *Joint structure estimation for categorical Markov networks* [Technical report]. Ann Arbor, MI: Department of Statistics, University of Michigan.

Guo, J., Levina, E., Michailidis, G., & Zhu, J. (2011). Joint estimation of multiple graphical models. *Biometrika, 98,* 1–15. http://dx.doi.org/10.1093/biomet/asq060

Guo, P. (2015). *eLasso: Enhanced least absolute shrinkage and selection operator regression model*. Retrieved from https://cran.r-project.org/web/packages/elasso/elasso.pdf

Handcock, M. S., Raftery, A. E., & Tantrum, J. (2007). Model-based clustering for social networks. *Journal of the Royal Statistical Society A, 170,* 301–354. http://dx.doi.org/10.1111/j.1467-985X.2007.00471.x

Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association, 97,* 1090–1098. http://dx.doi.org/10.1198/016214502388618906

Höfling, H., & Tibshirani, R. (2009). Estimation of sparse binary pairwise Markov networks using pseudo-likelihoods. *Journal of Machine Learning Research, 10,* 883–906.

Holland, P. W., Laskey, K. B., & Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks, 5,* 109–137. http://dx.doi.org/10.1016/0378-8733(83)90021-7

Holland, P. W., & Leinhardt, S. (1976). Local structure in social networks. *Sociological Methodology, 7,* 1–45. http://dx.doi.org/10.2307/270703

Holland, P. W., & Leinhardt, S. (1977). A dynamic model for social networks. *Journal of Mathematical Sociology, 5,* 5–20. http://dx.doi.org/10.1080/0022250X.1977.9989862

Holland, P. W., & Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs (with discussion). *Journal of the American Statistical Association, 76,* 33–50. http://dx.doi.org/10.1080/01621459.1981.10477598

Ising, E. (1925). Beitrag zur theorie des ferromagnetismus [Contribution to the theory of ferromagnetism]. *Zeitschrift fur Physik, 31,* 253–258. http://dx.doi.org/10.1007/BF02980577

Jaccard, P. (1901). Distribution de la flore alpine dans le Bassin des Dranses et dans quelques régions voisines [Distribution of the Alpine Flora in the Dranses Basin and in some neighboring regions]. *Bulletin de la Société des Sciences Naturelles, 37,* 241–272.

Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997). Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science, 16,* 39–59. http://dx.doi.org/10.1287/mksc.16.1.39

Kaufman, L., & Rousseeuw, P. J. (2005). *Finding groups in data: An introduction to cluster analysis* (2nd ed.). New York, NY: Wiley.

Kendler, K. S., Aggen, S. H., Flint, J., Borsboom, D., & Fried, E. I. (2018). The centrality of DSM and non-DSM depressive symptoms in Han Chinese women with major depression. *Journal of Affective Disorders, 227,* 739–744. http://dx.doi.org/10.1016/j.jad.2017.11.032

Kindermann, R., & Snell, J. L. (1980). *Markov random fields and their applications*. New York, NY: American Mathematical Society. http://dx.doi.org/10.1090/conm/001

Köhn, H.-F., Steinley, D., & Brusco, M. J. (2010). The *p*-median model as a tool for clustering psychological data. *Psychological Methods, 15,* 87–95. http://dx.doi.org/10.1037/a0018535

Kossakowski, J. J., Epskamp, S., Kieffer, J. M., van Borkulo, C. D., Rhemtulla, M., & Borsboom, D. (2016). The application of a network approach to Health-Related Quality of Life (HRQoL): Introducing a new method for assessing HRQoL in healthy adults and cancer patients. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation, 25,* 781–792. http://dx.doi.org/10.1007/s11136-015-1127-z

Kulczynski, S. (1927). Zespoly roslin w Pieninach [Sciences Naturelles]. *Bulletin International de l'Académie Polonaise des Sciences et des Lettres, Classe des Sciences Mathématiques et Naturelles, Série B*(Suppl. 2), 57–203.

Laciana, C. E., & Rovere, S. L. (2011). Ising-like agent-based technology diffusion model: Adoption patterns vs. seeding strategies. *Physica A, 390,* 1139–1149. http://dx.doi.org/10.1016/j.physa.2010.11.006

Lane, S. P., Steinley, D., & Sher, K. (2016). Meta-analysis of DSM alcohol use disorder criteria severities: Structural consistency is only "skin deep." *Psychological Medicine, 46,* 1769–1784. http://dx.doi.org/10.1017/S0033291716000404

Lauritzen, S. L. (1996). *Graphical models*. New York, NY: Oxford University Press.

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature, 401,* 788–791. http://dx.doi.org/10.1038/44565

Lee, D. D., & Seung, H. S. (2001). Algorithms for nonnegative matrix factorization. In T. L. Keen, T. K. Dietterich, & V. Tresp (Eds.), *Advances in neural information processing systems* (Vol. 13, pp. 556–562). Cambridge, MA: MIT Press.

Lee, M. D., & Pope, K. J. (2003). Avoiding the dangers of averaging across subjects when using multidimensional scaling. *Journal of Mathematical Psychology, 47,* 32–46. http://dx.doi.org/10.1016/S0022-2496(02)00019-6

Lee, S.-I., Ganapathi, V., & Koller, D. (2007). Efficient structure learning of Markov networks usingl_1-regularization. In B. Schölkopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems* (Vol. 19, pp. 817–824). Cambridge, MA: MIT Press.

Lee, S.-I., Lee, H., Abbeel, P., & Ng, A. Y. (2007). Efficient L_1 regularized logistic regression. In A. Cohn (Ed.), *Proceedings of the 21st National Conference on Artificial Intelligence and the 18th Innovative Applications of Artificial Intelligence Conference* (Vol. 1, pp. 401–408). Menlo Park, CA: AAAI Press.

Lima, F. W. S. (2015). Tax evasion dynamics via non-equilibrium model on complex networks. *Theoretical Economics Letters, 5,* 775–783. http://dx.doi.org/10.4236/tel.2015.56089

Liu, S., Ying, L., & Shakkottai, S. (2010). *Influence maximization in social networks: An Ising-model-based approach*. In P. Viswanath & S. Meyn (Eds.), Proceedings of the 48th Annual Allerton Conference on Communication, Control, and Computing (pp. 570–576). Piscataway, NJ: IEEE.

Loh, P.-L., & Wainwright, M. J. (2013). Structure estimation for discrete graphical models: Generalized covariance matrices and their inverses. *Annals of Statistics, 41,* 3022–3049. http://dx.doi.org/10.1214/13-AOS1162

Lorrain, F., & White, H. C. (1971). Structural equivalence of individuals in social networks. *The Journal of Mathematical Sociology, 1,* 49–80. http://dx.doi.org/10.1080/0022250X.1971.9989788

McCormick, W. T., Jr., Schweitzer, P. J., & White, T. W. (1972). Problem decomposition and data reorganization by a clustering technique. *Operations Research, 20,* 993–1009. http://dx.doi.org/10.1287/opre.20.5.993

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

Meinshausen, N., & Bühlmann, P. (2006). High dimensional graphs and variable selection with the lasso. *Annals of Statistics, 34,* 1436–1462. http://dx.doi.org/10.1214/009053606000000281

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika, 54,* 557–585. http://dx.doi.org/10.1007/BF02296397

Nowicki, K., & Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association, 96,* 1077–1087. http://dx.doi.org/10.1198/016214501753208735

Ochiai, A. (1957). Zoogeographic studies on the soleoid fishes found in Japan and its neighboring regions. *Nippon Suisan Gakkaishi, 22,* 526–530. http://dx.doi.org/10.2331/suisan.22.526

Oh, W., & Jeon, S. (2007). Membership herding and network stability in the open source community: The Ising perspective. *Management Science, 53,* 1086–1101. http://dx.doi.org/10.1287/mnsc.1060.0623

Pattison, P. E. (1993). *Algebraic models for social networks*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511571220

Pattison, P. E., & Bartlett, W. K. (1982). A factorization procedure for finite algebras. *Journal of Mathematical Psychology, 25,* 51–81. http://dx.doi.org/10.1016/0022-2496(82)90046-3

Pattison, P. E., & Breiger, R. L. (2002). Lattices and dimensional representations: Matrix decompositions and ordering structures. *Social Networks, 24,* 423–444. http://dx.doi.org/10.1016/S0378-8733(02)00015-1

Ravikumar, P., Wainwright, M. J., & Lafferty, J. D. (2010). High-dimensional Ising model selection using l_1-regularized logistic regres-

sion. *Annals of Statistics, 38,* 1287–1319. http://dx.doi.org/10.1214/09-AOS691

Rhemtulla, M., Fried, E. I., Aggen, S. H., Tuerlinckx, F., Kendler, K. S., & Borsboom, D. (2016). Network analysis of substance abuse and dependence symptoms. *Drug and Alcohol Dependence, 161,* 230–237. http://dx.doi.org/10.1016/j.drugalcdep.2016.02.005

Santhanam, N. P., & Wainwright, M. J. (2012). Information-theoretic limits of selecting binary graphical models in high dimensions. *IEEE Transactions on Information Theory, 58,* 4117–4134. http://dx.doi.org/10.1109/TIT.2012.2191659

Sarstedt, M., & Ringle, C. M. (2010). Treating unobserved heterogeneity in PLS path modeling: A comparison of FIMIX-PLS with different data analysis strategies. *Journal of Applied Statistics, 37,* 1299–1318. http://dx.doi.org/10.1080/02664760903030213

Schmittmann, V. D., Cramer, A. O., Waldorp, L. J., Epskamp, S., Kievit, R. A., & Borsboom, D. (2013). Deconstructing the construct: A network perspective on psychological phenomena. *New Ideas in Psychology, 31,* 43–53. http://dx.doi.org/10.1016/j.newideapsych.2011.02.007

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General, 116,* 250–264. http://dx.doi.org/10.1037/0096-3445.116.3.250

Sneath, P. H. A., & Sokal, R. R. (1973). *Numerical taxonomy: the principles and practice numerical classification*. San Francisco, CA: Freeman.

Sokal, R. R., & Sneath, P. H. (1963). *Principles of numerical taxonomy*. San Francisco, CA: Freeman.

Späth, H. (1979). Algorithm 39: Clusterwise linear regression. *Computing, 22,* 367–373. http://dx.doi.org/10.1007/BF02265317

Steinley, D., Brusco, M. J., & Wasserman, S. (2011). Clusterwise p^* models for social network analysis. *Statistical Analysis and Data Mining, 4,* 487–496. http://dx.doi.org/10.1002/sam.10139

Steinley, D., Hoffman, M., Brusco, M. J., & Sher, K. J. (2017). A method for making inferences in network analysis: Comment on Forbes, Wright, Markon, and Krueger (2017). *Journal of Abnormal Psychology, 126,* 1000–1010. http://dx.doi.org/10.1037/abn0000308

Steinley, D., & McDonald, R. P. (2007). Examining factor score distributions to determine the nature of latent spaces. *Multivariate Behavioral Research, 42,* 133–156. http://dx.doi.org/10.1080/00273170701341217

Sussman, D. L., Tang, M., Fishkind, D. E., & Priebe, C. E. (2012). A consistent adjacency spectral embedding for stochastic blockmodel graphs. *Journal of the American Statistical Association, 107,* 1119–1128. http://dx.doi.org/10.1080/01621459.2012.699795

Suwan, S., Lee, D. S., Tang, R., Sussman, D. L., Tang, M., & Priebe, C. E. (2016). Empirical Bayes estimation for stochastic blockmodel. *Electronic Journal of Statistics, 10,* 761–782. http://dx.doi.org/10.1214/16-EJS1115

Todeschini, R., Consonni, V., Xiang, H., Holliday, J., Buscema, M., & Willett, P. (2012). Similarity coefficients for binary chemoinformatics data: Overview and extended comparison using simulated and real data sets. *Journal of Chemical Information and Modeling, 52,* 2884–2901. http://dx.doi.org/10.1021/ci300261r

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. http://dx.doi.org/10.1038/srep05918

van Borkulo, C., Boschloo, L., Borsboom, D., Penninx, B. W., Waldorp, L. J., & Schoevers, R. A. (2015). Association of symptom network structure with the course of depression. *Journal of the American Medical Association Psychiatry, 72,* 1219–1226. http://dx.doi.org/10.1001/jamapsychiatry.2015.2079

Wainwright, M. J., Ravikumar, P., & Lafferty, J. D. (2007). High-dimensional graphical model selection usingl_1-regularized logistic regression. In B. Schölkopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems* (Vol. 19, pp. 1465–1472). Cambridge, MA: MIT Press.

Wasserman, S. (1978). Models for binary directed graphs and their applications. *Advances in Applied Probability, 10,* 803–818. http://dx.doi.org/10.2307/1426660

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511815478

Xue, L., Zou, H., & Cai, T. (2012). Nonconcave penalized composite conditional likelihood estimation of sparse Ising models. *Annals of Statistics, 40,* 1403–1429. http://dx.doi.org/10.1214/12-AOS1017

Yule, G. U. (1900). On the association of attributes in statistics. *Philosophical Transactions of the Royal Society A, 194,* 257–319. http://dx.doi.org/10.1098/rsta.1900.0019

(*Appendices follow*)

Appendix A

Formulation of the Binary Pairwise Markov Random Field (Ising) Model

To present the model formally, we use notation that draws heavily from several published works on this topic (Höfling & Tibshirani, 2009; Ravikumar et al., 2010; Santhanam & Wainwright, 2012; Wainwright et al., 2007). The model assumes a set of p vertices, $V = \{1, \ldots, p\}$, as well as an edge set (E) consisting of unordered vertex pairs $\{s, t\}$: $\{s, t\} \in E$, where $1 \leq s \neq t \leq p$. The underlying graph, $G = (V, E)$, is assumed to be *undirected*. The neighborhood set of vertex s, denoted $\eta(s) = \{t \in V: \{s, t\} \in E\}$, is the set of vertices that share an edge with s (for each $s \in V$). The degree of vertex s is $d(s) = |\eta(s)|$, which is the number of vertices that share an edge with vertex s.

Denoting $\mathbf{x} = (x_1, \ldots, x_p)$ as a binary row vector in $\mathbf{X}$, $\mathbf{\Theta}$ as a $p \times p$ matrix consisting of interaction parameters (θ_{st}) on the off-diagonal and node parameters θ_s on the main diagonal. The binary pairwise Markov random field (Ising model) generates data under the model,

$$\boldsymbol{p}(\mathbf{x}, \mathbf{\Theta}) = \exp\Big\{\sum_{s \in V} \theta_s x_s + \sum_{\{s,t\} \in E} \theta_{st} x_s x_t - \Psi(\Theta)\Big\}, \qquad \text{(A.1)}$$

where $\Psi(\mathbf{\Theta})$ is the partition function to assure that the distribution sums to one. Höfling and Tibshirani (2009) note that, under the assumption that $\{s, t\} \notin E \Rightarrow \theta_{st} = 0$, the log-likelihood function for Equation (A.1) can be written as

$$l(\mathbf{x}, \mathbf{\Theta}) = \log \mathrm{p}(\mathbf{x}, \mathbf{\Theta}) = \Big\{\sum_{s \geq t \geq 1}^{p} \theta_{st} x_s x_t - \Psi(\Theta)\Big\}, \qquad \text{(A.2)}$$

where the node parameter $\theta_s = \theta_{ss}$, for all $1 \leq s \leq p$. The off-diagonal elements of $\mathbf{\Theta} = [\theta_{st}]$ are interaction parameters, such that θ_{st} can be interpreted as a measure of conditional dependence between random variables s and t, subject to fixed values of all other vertices, $V \backslash \{s, t\}$. Although the θ_{st} values are free to vary in $\Re$ for some solution methods, symmetry is generally enforced via some type of process. If it is assumed that symmetry is enforced via consideration of only the lower triangle of $\mathbf{\Theta}$, then Höfling and Tibshirani indicate that the penalized log-likelihood function over the full sample of n individuals is

$$\Big\{\sum_{s \geq t \geq 1}^{p} (\mathbf{X}^{\mathbf{T}}\mathbf{X})\theta_{st} - n\Psi(\Theta) - n \bullet \mathbf{R}\Theta\Big\}, \qquad \text{(A.3)}$$

where the $p \times p$ matrix $\mathbf{R}$ is a lower triangular matrix of penalty parameters, and '$\bullet$' indicates an element-wise matrix product.

Appendix B

Information-Theoretic Sufficient Conditions for Sample Size

Santhanam and Wainwright (2012) derived sufficient conditions for the case in which the edge weights were assumed known, and also for situations in which they are unknown. For the case of known edge weights, the required sample size is

$$n \geq \frac{3[3\exp(2\omega) + 1]}{\sinh^2\left(\frac{\lambda}{2}\right)} d\left(3\log p + \log(2d) + \log\left(\frac{1}{\delta}\right)\right), \qquad \text{(B.1)}$$

and for the case of unknown edge weights (in which case d is unknown), the requirement is

$$n \geq \left[\frac{\omega(3\exp(2\omega) + 1)}{\sinh^2\left(\frac{\lambda}{4}\right)}\right]^2 \left(16\log p + 4\log\left(\frac{2}{\delta}\right)\right). \qquad \text{(B.2)}$$

Received January 25, 2018
Revision received October 14, 2018
Accepted November 27, 2018 ■