Towards a Grand Unified Theory of Network Models: Reply to Brusco, Steinley, Hoffman, Davis-Stober, & Wasserman

Marsman, M.[1], Waldorp, L. J.[1], & Borsboom, D.[1]
1 University of Amsterdam

Correspondence concerning this article should be addressed to:
Maarten Marsman
University of Amsterdam, Psychological Methods
Nieuwe Achtergracht 129B
PO Box 15906, 1001 NK Amsterdam, The Netherlands
E-mail may be sent to m.marsman@uva.nl.

MM was supported by a Veni grant (451-17-017) from the Netherlands Organization for Scientific Research (NWO).

Abstract

Network models like the Ising model are increasingly used in psychological research. In a recent article published in this journal, Brusco, Steinley, Hoffman, Davis-Stober, and Wasserman (in press) provide a critical assessment of the conditions that underlie the Ising model and the eLasso method that is commonly used to estimate it. In this commentary, we show that their main criticisms are unfounded. First, where Brusco et al. suggest that Ising models have little to do with classical network models such as random graphs, we show that they can be fruitfully connected. Second, if one makes this connection it is immediately evident that Brusco et al.'s second criticism – that the Ising model requires complete population homogeneity and does not allow for individual differences in network structure – is incorrect. In particular, we establish that if every individual has their own network, and these individual differences instantiate a random graph model, the Ising model will hold in the population. Hence, population homogeneity is sufficient for the Ising model, but it is not necessary, as Brusco et al. suggest. Third, we address Brusco et al.'s criticism regarding the sparsity assumption that is made in common uses of the Ising model. We show that this criticism is misdirected, as it targets a particular estimation algorithm for the Ising model rather than the model itself. We also describe various established and validated approaches for estimating the Ising model for networks that violate the sparsity assumption. Finally, we outline important avenues for future research.

Keywords: Ising model, random-cluster model, graphical model, random graph model, network model, network psychometrics, homogeneity

Introduction

In a recent article that was published in this journal, Brusco, Steinley, Hoffman, Davis-Stober, and Wasserman (in press, henceforth BSHDW) provided a critical assessment of the conditions that underlie the Ising model and the eLasso method that is commonly used in psychology to estimate the Ising model. Their reason for this critical assessment was that "*little concern has been reflected in the psychopathology literature with respect to the data assumptions and conditions that underlie the Ising (1925) model and/or the methods designed for its estimation*" (p. 13). BSHDW identified two critical assumptions of the Ising model and/or the eLasso method that they believe are unjustified in applications of psychopathology, or psychological science more generally. One of these key assumptions is that the population is homogeneous with respect to the model's characteristics, i.e., that persons are i.i.d. replications of the same associative structure. The other key assumption is the assumption of the eLasso method that the underlying associative structure is sparse in nature, i.e., that there exist relatively few connections between observed variables. The authors then conclude that there are "*serious concerns regarding the implementation of eLasso in psychopathological research*" (p. 1) and propose three distinct approaches that may help mitigate these concerns.

An apparently innocuous critique permeates the authors' assessment: "*Although some researchers will maintain that the debate over the terms networks vs. dependence graph is merely one of semantics, we believe there is a salient distinction*" (p. 2). Whereas we believe that network models comprise both graphical- and random graph models, BSHDW seem to reserve the term network model exclusively for random graph models. This suggests that it would be wrong to refer to graphical models such as the Ising model as a network model, and by extension it would also be wrong to call our area of research *network psychometrics.* We respectfully disagree with this narrow view.[1] We do agree, however, that there is a clear, formal distinction between the class of random graph models and the class of graphical models. But even though it may appear that these model families have little in common, an idea that underlies BSHDWs' critiques and proposals, we will use the theory of Fortuin and Kasteleyn (1972) to argue that these models are not only closely connected, but that further study of this connection in fact defines an absolutely crucial avenue for research in network psychometrics.

The theory of Fortuin and Kasteleyn unifies the disparate fields of graphical modeling and random graph modeling, which turns out to be quite useful in understanding BSHDWs' assessment and proposed solutions. As an example, we will consider the authors' assessment regarding homogeneity. We agree with BSHDW that homogeneity of the population with respect to the Ising model's characteristics will guarantee the model's fit to cross-sectional data. In this commentary, however, we wish to address the mistaken belief that it is the only way to ensure the model's fit. In fact, as we will show here, homogeneity is not necessary, as the authors argue, but merely sufficient. Other sufficient conditions exist however, and we will argue that studying these can reveal important connections between random graph models and graphical models. We will illustrate a particular alternative that does not require that the population is homogeneous with respect to the model's characteristics. Using a recent idiographic network characterization of the Ising model (Marsman, 2019; Savi, Marsman, van der Maas, & Maris, in press) that builds on the theory of Fortuin and Kasteleyn (1972) we demonstrate how network structures that are unique to the individual can still generate an Ising model cross-sectionally. The theory of Fortuin and Kasteleyn furthermore reveals interesting connections between two of BSHDWs' proposed solutions, the (stochastic) block model and a low-rank approximation to the gram matrix, and existing psychometric modeling approaches.

The remainder of this commentary is organized as follows. First, we introduce the two types of network models, the random graph model and the graphical model, and discuss the theory of Fortuin and Kasteleyn that consolidates the relation between these two statistical frameworks. Second, we use the consolidated theory to address the mistaken belief that a homogeneous population is required for the cross-sectional application of Ising models. Third, we address some omissions, and discuss several results and models that BSHDW have missed in their discussion centred around the sparsity assumption.

[1]Observe that this not only chastises network psychometrics for using the network name, but also other fields that centre around graphical models, such as neural networks (Hopfield, 1982) and Bayesian networks (Pearl, 1988).

A Unified Theory of Network Models

We first introduce some notation. Let the vector $\mathbf{x} = [x_1, x_2, \dots, x_p]^\mathsf{T}$ denote the states of p nodes in an p-variable graph or network. The vector $\mathbf{w} = [w_{11}, w_{12}, \dots w_{23}, \dots w_{(p-1)p}]^\mathsf{T}$ is used to denote the $\binom{p}{2}$ network relations, with $w_{ij} \in \{0, 1\}$, where $w_{ij} = 1$ implies that variables i and j are directly connected and $w_{ij} = 0$ implies that these variables are not directly connected.

A random graph model is characterized by a particular probability distribution over topological structures, i.e., it is a model for the different configurations of relations between nodes of the graph or network: $p(\mathbf{W} = \mathbf{w})$. One example could be the modeling of the connections between neurons in a person's brain, where the neurons are represented as the nodes of a network $\mathbf{x}$ and the dendrite connections then represent the edges of the network $\mathbf{w}$. A well-known statistical model that falls in this class of models is the model of Erdős and Rényi (1960)[2]

$$p(\mathbf{W} = \mathbf{w}) = \prod_{i=1}^{p-1} \prod_{j=i+1}^{p} \theta_{ij}^{w_{ij}} \, (1 - \theta_{ij})^{1-w_{ij}},$$

where the w_{ij} are independent Bernoulli(θ_{ij}) variables. It is important to see that random graph models describe the relational patterns that may exist between the network's nodes but that it does not model the network's node states $\mathbf{x}$. Thus, the random graph model describes the way the edges in a graph are distributed over pairs of nodes, without saying anything about the probability distribution of node states.

In contrast, graphical models do describe the patterns in the network's node states, such as the activity patterns of the neurons in our brain network example: $p(\mathbf{X} = \mathbf{x})$. Observe that the graphical model does not model the connections between nodes but treats them as fixed entities. A prime example of such a graphical model is the Ising model,

$$p(\mathbf{X} = \mathbf{x}) = \frac{\exp\left(\sum_{i=1}^{p} x_i \mu_i \, \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} x_i x_j \sigma_{ij}\right)}{\sum_{\mathbf{x}} \exp\left(\sum_{i=1}^{p} x_i \mu_i \, \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} x_i x_j \sigma_{ij}\right)} \tag{1}$$

where $x_i \in \{-1, +1\}$, for $i = 1, \dots, p$, and the sum in the denominator is with respect to all 2^p possible states of the vector $\mathbf{x}$. The Ising model comprises two parameters (see Marsman et al., 2018, for details), the vector of main effects or activation thresholds $\boldsymbol{\mu}$ and the matrix $\boldsymbol{\Sigma} = [\sigma_{ij}]$ of pairwise interactions. Thus, graphical models such as the Ising model describe the probability distribution of the node states as a function of the network structure that is encoded in the edge weights σ_{ij}.

The random graph model and the graph model are thus complementary to each other, as they model two distinct aspects of the network. In the view of BSHDW, however, interpreting the Ising model as a network "*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*" (p. 2). Although we agree with the authors that we should always scrutinize new methods, we also believe that their focus is primarily on how the two statistical modeling frameworks differ. We would advocate a more inclusive view towards

[2]This model is often referred to as *the* random graph model. Here, however, we will refer to it as the Erdös–Rényi model.

the two frameworks and focus on what relations there may exist between the two approaches and what we could learn from that relation.

This is what two physicists, Cees Fortuin and Piet Kasteleyn, did almost fifty years ago. In physics, graphical models such as the Ising model are used to model the behavior of particles (e.g., magnetism), and random graph models such as the Erdös–Rényi model are used to model the way that particles may trickle through a porous object (e.g., percolation; Broadbent & Hammersley, 1957). Fortuin and Kasteleyn observed several unexplained relations between the phenomena that were studied in these two disparate areas of research and set out to discover if this was more than pure coincidence (see the appendix of Grimmet, 2006). In a series of papers, Fortuin and Kasteleyn worked out how the two worlds relate (Fortuin, 1972a, 1972b; Fortuin & Kasteleyn, 1972), and revealed the following latent topology characterization of the Ising model in Eq. (1)

$$p(\mathbf{X}=\mathbf{x}) = \sum_{\mathbf{w}} \underbrace{\prod_{c=1}^{\kappa(\mathbf{w})} \frac{\exp\left(\sum_{i\in V_c} x_i\mu_i\right)}{2\cosh\left(\sum_{i\in V_c}\mu_i\right)}}_{p(\mathbf{X}=\mathbf{x}\mid\mathbf{W}=\mathbf{w})} \underbrace{\frac{\prod_{i=1}^{p-1}\prod_{j=i+1}^{p}\theta_{ij}^{w_{ij}}(1-\theta_{ij})^{1-w_{ij}}\prod_{c=1}^{\kappa(\mathbf{w})}\beta_c}{\sum_{\mathbf{w}}\prod_{i=1}^{p-1}\prod_{j=i+1}^{p}\theta_{ij}^{w_{ij}}(1-\theta_{ij})^{1-w_{ij}}\prod_{c=1}^{\kappa(\mathbf{w})}\beta_c}}_{p(\mathbf{W}=\mathbf{w})} \tag{2}$$

where the sum is with respect to all $2^{\binom{p}{2}}$ possible topological states $\mathbf{w}$, $\kappa(\mathbf{w})$ denotes the number of connected components or open clusters[3] in the topology $\mathbf{w}$, and Cioletti and Vila (2016) showed that the Ising model in Eq. (1) is obtained when the cluster weights are equal to[4]

$$\beta_c = 2\cosh\left(\sum_{i\in V_c}\mu_i\right) \geq 2,$$

where V_c denotes the set of indices that are part of the c-th cluster. The random graph model $p(\mathbf{w})$ in Eq. (2) is a generalization of the Erdös–Rényi model that is known as the random-cluster model, and the version that is used here favors network structures with more open clusters over structures that have fewer open clusters. The latent topology model $p(\mathbf{x} \mid \mathbf{w})$ is used to assign values to the nodes $\mathbf{x}$ of the topology $\mathbf{w}$ —it "colors" the graph— which assigns its values per cluster. A proof that the representation in Eq. (2) indeed leads to the Ising model in (1) can be found in the appendix and the references above.

The seminal work of Fortuin and Kasteleyn consolidates graph- and random graph models and thus offers us a glimpse of what a unified theory of network models might look like. However, there are two other approaches that consolidate the connection between the two frameworks; the divide and color model of Häggström (2001) and the random triangle model of Jonasson (1997). The divide and color model is of the form of Eq. (2) but uses the Erdös–Rényi model instead of the random-cluster model. Even though the associated graphical model cannot be expressed in closed form, and is thus not an Ising model, several of its properties are known (Bálint, 2010; Häggström, 2001), some of which are closely related to that of other latent variable characterizations of the Ising model (e.g., Fortuin, Kasteleyn, & Ginibre, 1971; Holland & Rosenbaum, 1986). The random-triangle model of

[3]An open cluster is a group of nodes that lie on an open path together, so that one can travel along all of the nodes in the group using the edges in $\mathbf{w}$.

[4]Fortuin and Kasteleyn have only treated the special case of an Ising model that excludes main effects. The cluster weights are then equal to 2.

Jonasson (1997) offers yet another approach, in which the focus on clusters in the latent topology representation in Eq. (2) is entirely abandoned in favor of triangles instead. This representation can, in principle, also characterize an Ising model (Häggström & Jonasson, 1999). In sum, there exist many interesting connections between random graph models and graphical models, which confirms that their distinction may be less salient than BSHDW make us believe.

Stochastic Block Models Generate Low-Rank Ising and Bi-Factor IRT Models.

As a testament to the above sentiment, we consider the stochastic block model of Holland, Laskey, and Leinhardt (1983) —a random-graph model with community structure (Karrer & Newman, 2011)— that was proposed by BSHDW to block or factor the network's nodes. Here, we follow Savi et al. (in press) and impose the community structure on the random-cluster model. In its simplest form, we differentiate between the probability θ_W of laying an edge between nodes within the same community, and the probability θ_B of laying an edge between nodes of different communities:

$$p(\mathbf{W} = \mathbf{w}) = \frac{\prod_{i,j \in E_W} \theta_W^{w_{ij}} (1-\theta_W)^{1-w_{ij}} \prod_{i,j \in E_B} \theta_B^{w_{ij}} (1-\theta_B)^{1-w_{ij}} \prod_{c=1}^{\kappa(\mathbf{w})} \beta_c}{\sum_{\mathbf{w}} \prod_{i,j \in E_W} \theta_W^{w_{ij}} (1-\theta_W)^{1-w_{ij}} \prod_{i,j \in E_B} \theta_B^{w_{ij}} (1-\theta_B)^{1-w_{ij}} \prod_{c=1}^{\kappa(\mathbf{w})} \beta_c},$$

where E_W comprises the node pairs that are in the same community and E_B the node pairs that are in different communities. In Savi et al. (in press) we have shown that this random-cluster model generates an Ising model with a community structure in its association matrix

$$\mathbf{\Sigma} = [\sigma_{ij}] = \sigma_B \mathbf{1}_n + (\sigma_W - \sigma_B)\mathbf{C},$$

where $\mathbf{C}$ is a symmetric $p \times p$ matrix with elements c_{ij} that are equal to one whenever i and j are in the same community and zero otherwise. This correspondence follows from the (monotonic) relation between the Ising model's associations and the edge probabilities in the random-cluster model (Fortuin & Kasteleyn, 1972; Grimmet, 2006)

$$\sigma_{ij} = -\frac{1}{2} \log(1 - \theta_{ij}). \tag{3}$$

If $\theta_B < \theta_W$, which implies that there are more connections between members of the same community than between members of different communities, the associations in the related Ising model have the same ordering, $\sigma_B < \sigma_W$, so that all elements of the association matrix $\mathbf{\Sigma}$ are positive and the related Ising model is consistent with a bi-factor item response theory (IRT) model (Marsman, Maris, Bechger, & Glas, 2015; Savi et al., in press, p. 1056–1057). This shows that the proposal of BSHDW to use a block modeling approach to factor nodes is strongly related to existing low-rank approaches, and their suggestion to block both nodes and subjects will generate mixtures of low-rank Ising models, each with a potentially different bi-factor structure. In principle, such an idea might work if one can efficiently regularize its parameters, otherwise it can become rather unwieldy and difficult to estimate, especially if many communities will be detected. In any case, the unified theory of Fortuin and Kasteleyn is useful in that it allows us to compare random graph based methods from the perspective of the graphical model, or vice versa.

An Idiographic Interpretation of Cross-Sectional Ising Models

One of the primary critiques of BSHDW against the Ising model concerns the assumption that individuals are exchangeable, i.e., that they are i.i.d. replications of a single associative structure. But in their words, "*there is little compelling evidence that the Ising model is well-suited for identifying good graphical models of psychopathological data*" (p. 8). One reason for this is that "*psychopathology data may not consist of n i.i.d. realizations from the Ising model*" (p. 8). We would argue that this assumed heterogeneity in the population need not be an issue for the empirical application of the Ising model to cross-sectional data, and illustrate an alternative interpretation that allows for heterogeneity with respect to the model's statistical relations.

It is well-known that statistical relations that are found at the group level may be fundamentally different from the relations that are found at the participant level (Kievit, Frankenhuis, Waldorp, & Borsboom, 2013) —a phenomenon that is commonly referred to as Simpson's paradox (Simpson, 1951) or the ecological fallacy (Robinson, 1950)— which is one of the factors that has lead to the division of psychological science in idiographic and cross-sectional approaches (Cronbach, 1957; Molenaar, 2004). This division can also be observed in the network psychometric literature (e.g., E. Bos & Wanders, 2016; F. M. Bos, Snippe, Bruggeman, Wichers, & van der Krieke, in press; Fisher, 2015; Fisher, Medaglia, & Jeronimus, 2018). But if statistical relations can vary at the individual level, how can idiographic networks then be related to Ising models that are estimated from cross-sectional data?

The answer is given by Marsman (2019), who used the characterization in Eq. (2) to draw two broad conclusions. First, the characterization in Eq. (2) using latent topologies illustrates how one can retrieve an Ising model when the network's topology is a random effect that may vary at an individual level. That is, idiographic networks that embody a unique set of statistical relations between variables at the participant level may very well be consistent with an Ising model at the population level. This underscores our point that homogeneity is not a necessary condition for retrieving an Ising model from cross-sectional data. Second, there exists a concrete, formal relation between the cross-sectional associations and idiographic relations. In the theory of Fortuin and Kasteleyn, the associations σ_{ij} that are estimated from cross-sectional data are monotonically related to the proportion θ_{ij} of idiographic topological structures that have a connection between nodes i and j —i.e., Eq. 3. Such a definite connection between cross-sectional models and idiographic models is rather unique, in particular in the network psychometric literature, and shows how group-level phenomena can be generated from idiographic models (e.g., Savi et al., in press). In sum, the theory of Fortuin and Kasteleyn not only consolidates the random graph- and graphical modeling approaches, it also bridges the idiographic and cross-sectional approaches in network psychometrics.

Low-Rank Models Offer a Workable Alternative to Sparsity

A second concern that BSHDW voiced about the use of the Ising model in psychopathology, or psychology more generally, is that "*psychopathological networks may not be sparse*" (p. 8). To be clear, the assumption that the network is sparse is not related to the Ising model, but to the eLasso method that is used to estimate the Ising model. The eLasso however is an algorithm, not a model, and hence BSHDWs' argument is misdirected. Waldorp, Marsman, and Maris (2019), for example, show how the assumption of having a sparse connected network underlies the eLasso method, and what the effect of violating this assumption is on both parameter estimation and the prediction of node states. Specifically, it was shown that the estimation error increases when the underlying network is no longer sparse, even though prediction accuracy actually improves. Thus for recovering the true underlying graph structure using the eLasso method it is critical to have a truly sparse graph or have sufficient data to overcome this prior assumption (Epskamp, Kruis, & Marsman, 2017).

As we have shown here, one of BSHDW's three proposed solutions, the (stochastic) block model, was strongly related to a particular low-rank Ising model. The low-rank Ising model was proposed by Marsman et al. (2015) for situations where one expects the underlying graph to be densely connected (see also Marsman, Waldorp, & Maris, 2017). This method applies the Eckart and Young (1936) Theorem to formulate a low-rank approximation to the Ising model's association matrix,

$$\mathbf{\Sigma} \approx \sum_{r=1}^{R} \lambda_r \, \mathbf{q}\mathbf{q}^{\mathsf{T}},$$

in which all but the largest R eigenvalues λ_r are equated to zero (Marsman et al., 2018, 2015). We were therefore pleasantly surprised that BSHDW proposed "*…a singular value decomposition of* $\mathbf{X}$*, whereby a low-rank approximation based on eigenvectors from* $\mathbf{XX}'$ *or* $\mathbf{X}'\mathbf{X}$ *is obtained (Eckart & Young, 1936…*" (p. 15), although our surprise would have been even more pleasant had the authors bothered to include a reference to the original proposal. We believe that there is much merit to this approach, although we would argue the restriction should be imposed on the association matrix, and not the raw data, for reasons of statistical cleanliness. Observe, however, that the low-rank approach is related to a form of regularization that is known as the nuclear- or trace norm, which crucially depends on the assumption that the underlying network is densely connected. As an alternative, one may also consider approaches that bridge between a sparse and a dense network, such as the fused latent and graphical IRT model of Chen, Li, Liu, and Ying (2018). In sum, BSHDW's arguments are a) misdirected, because they mistake an estimation algorithm for a model, b) insufficiently informed by the extant literature, which already includes validated estimation algorithms for dense network structures, and c) poorly referenced, because BSHDW fail to properly acknowledge the original authors who proposed these algorithms.

Concluding Comments

In this commentary we have used the seminal work of Fortuin and Kasteleyn to refute BSHDWs' theses that homogeneity is a necessary condition for cross-sectional applications of the Ising model, and that graphical models have nothing to do with networks. Fortuin and Kasteleyn's unified approach towards random graph models and graphical models allowed us to reveal how BSHDWs' critiques and proposed solutions, which are heavily influenced by ideas from social network analysis, translate to the realm of network psychometrics. We hope that someday we will be able to construct a grand unified theory of these two statistical realms, and we believe that the work of Fortuin and Kasteleyn is an important first step in that direction.

References

Bálint, A. (2010). Gibbsianness and non-Gibsianness in divide and color models. *The Annals of Statistics*, *38*(4), 1609–1638. doi: 10.1214/09-AOP518

Bos, E., & Wanders, R. (2016). Group-level symptom networks in depression. *JAMA Psychiatry*, *73*(4), 411. doi: 1

Bos, F. M., Snippe, E., Bruggeman, R., Wichers, M., & van der Krieke, L. (in press). Insights of patients and clinicians on the promise of the experience sampling method for psychiatric care. *Psychiatric Services*. doi: 10.1176/appi.ps.201900050

Broadbent, S., & Hammersley, J. (1957). Percolation processes I. Crystals and mazes. *Mathematical Proceedings of the Cambridge Philosophical Society*, *53*(3), 629–641. doi: 10.1017/S0305004100032680

Brusco, M., Steinley, D., Hoffman, M., Davis-Stober, C., & Wasserman, S. (in press). On Ising models and algorithms for the construction of symptom networks in psychopathological research. *Psychological Methods*, 1–19. doi: 10.1037/met0000207

Chen, Y., Li, X., Liu, J., & Ying, Z. (2018). Robust measurement via a fused latent and graphical item response theory model. *Psychometrika*, *83*(3), 538–562. doi: 10.1007/s11336-018-9610-4

Cioletti, L., & Vila, R. (2016). Graphical representations for Ising and Potts models in general external fields. *Journal of Statistical Physics*, *162*(1), 81–122. doi: 10.1007/s10955-015-1396-5

Cronbach, L. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*(11), 671–684. doi: 10.1037/h0043943

Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, *1*(3), 211–218. doi: 10.1007/BF02288367

Epskamp, S., Kruis, J., & Marsman, M. (2017). Estimating psychopathological networks: Be careful what you wish for. *PLoS One*, *12*(e0179891). doi: 10.1371/journal.pone.0179891

Erdős, P., & Rényi, A. (1960). On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, *5*(1), 17–60.

Fisher, A. J. (2015). Toward a dynamic model of psychological assessment: Implications for personalized care. *Journal of Consulting and Clinical Psychology*, *83*(4), 825–836. doi: 10.1037/ccp0000026

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences*, *115*(27), E6106–E6115. doi: 10.1073/pnas.1711978115

Fortuin, C. (1972a). On the Random-Cluster model: III. The simple Random-Cluster model. *Physica*, *59*(4), 545-570. doi: 10.1016/0031-8914(72)90087-0

Fortuin, C. (1972b). On the Random-Cluster model II. The percolation model. *Physica*, *58*(3), 393–418. doi: 10.1016/0031-8914(72)90161-9

Fortuin, C., & Kasteleyn, P. (1972). On the Random-Cluster model: I. Introduction and relation to other models. *Physica*, *57*(4), 536–564. doi: 10.1016/0031-8914(72)90045-6

Fortuin, C., Kasteleyn, P., & Ginibre, J. (1971). Correlation inequalities on some partially ordered sets. *Communications in Mathematical Physics*, *22*(2), 89–103. Retrieved from `https://projecteuclid.org/euclid.cmp/1103857443`

Grimmet, G. (2006). *The random-cluster model*. Heidelberg, Germany: Springer-Verlag.

Häggström, O. (2001). Coloring percolation clusters at random. *Stochastic Processes and their Applications*, *96*(2), 213–242. doi: 10.1016/S0304-4149(01)00115-6

Häggström, O., & Jonasson, J. (1999). Phase transition in the random triangle model. *Journal of Applied Probability*, *36*(4), 1101–1115. Retrieved from `https://www.jstor.org/stable/3215581`

Holland, P., Laskey, K., & Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks*, *5*(2), 109–137. doi: 10.1016/0378-8733(83)90021-7

Holland, P., & Rosenbaum, P. (1986). Conditional association and unidimensionality in monotone latent variable models. *The Annals of Statistics*, *14*(4), 1523–1543. Retrieved from https://www.jstor.org/stable/2241486

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, *79*(8), 2554–2558. doi: 10.1073/pnas.79.8.2554

Ising, E. (1925). Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik*, *31*(1), 253–258. doi: 10.1007/BF02980577

Jonasson, J. (1997). The random triangle model. *Journal of Applied Probability*, *36*(3), 852–867. Retrieved from https://www.jstor.org/stable/3215446

Karrer, B., & Newman, M. (2011). Stochastic blockmodels and community structure in networks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, *83*(1), 016107. doi: 10.1103/PhysRevE.83.016107

Kievit, R. A., Frankenhuis, W. E., Waldorp, L. J., & Borsboom, D. (2013). Simpson's paradox in psychological science: A practical guide. *Frontiers in Psychology*, *4*(513), 1–14. doi: 10.3389/fpsyg.2013.00513

Marsman, M. (2019). The idiographic Ising model. *Retrieved from psyarxiv.com/h3ka5*.

Marsman, M., Borsboom, D., Kruis, J., Epskamp, S., van Bork, R., Waldorp, L., ... Maris, G. (2018). An introduction to network psychometrics: Relating Ising network models to item response theory models. *Multivariate Behavioral Research*, *53*(1), 15–35. doi: 10.1080/00273171.2017.1379379

Marsman, M., Maris, G., Bechger, T., & Glas, C. (2015). Bayesian inference for low-rank Ising networks. *Scientific Reports*, *5*(9050). doi: 10.1038/srep09050

Marsman, M., Waldorp, L., & Maris, G. (2017). A note on large-scale logistic prediction: Using an approximate graphical model to deal with collinearity and missing data. *Behaviormetrika*, *44*(2), 513–534. doi: 10.1007/s41237-017-0024-x

Molenaar, P. (2004). A manifesto on psychology as idiographic science: Brining the person back into scientific psychology, this time forever. *Measurement*, *2*(4), 201–218. doi: 10.1207/s15366359mea0204_1

Pearl, J. (1988, April). *Bayesian networks: A model of self-activated memory for evidential reasoning* (Tech. Rep. No. CSD-850017). Irvine, CA: University of California.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*(3), 351–357.

Savi, A., Marsman, M., van der Maas, H., & Maris, G. (in press). The wiring of intelligence. *Perspectives on Psychological Science*. doi: 10.31234/osf.io/32wr8

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, *13*(2), 238–241.

Waldorp, L., Marsman, M., & Maris, G. (2019). Logistic regression and Ising networks: prediction and estimation when violating lasso assumptions. *Behaviprmetrika*, *46*(1), 49–72. doi: 10.1007/s41237-018-0061-0

Appendix

The relation between Eqs. (1) and (2)

The original formulation of Eq. (2) by Fortuin and Kasteleyn is equal to

$$\sum_{\mathbf{w}} \frac{\prod_{i=1}^{p-1}\prod_{j=i+1}^{p}\left\{\theta_{ij}\,\delta_{(w_{ij},1)}\,\delta_{(x_i,\,x_j)} + (1-\theta_{ij})\,\delta_{(w_{ij},0)}\right\}\exp\left(\sum_{i=1}^{p} x_i\mu_i\right)}{\sum_{\mathbf{x}}\sum_{\mathbf{w}}\prod_{i=1}^{p-1}\prod_{j=i+1}^{p}\left\{\theta_{ij}\,\delta_{(w_{ij},1)}\,\delta_{(x_i,\,x_j)} + (1-\theta_{ij})\,\delta_{(w_{ij},0)}\right\}\exp\left(\sum_{i=1}^{p} x_i\mu_i\right)}, \tag{4}$$

where $\delta_{(a,\,b)}$ is an indicator function that is equal to one whenever $a = b$ and zero otherwise. We first illustrate that Eqs. (2) and (4) are, in fact, the same, and characterize the marginal

$p(\mathbf{w})$ and conditional $p(\mathbf{x} \mid \mathbf{w})$. Then, we illustrate the relation between Eqs. (1) and (4).

We start with formulating the marginal distribution $p(\mathbf{w})$. Following Grimmet (2006), we observe that the product can be reformulated as

$$\prod_{i=1}^{p-1}\prod_{j=i+1}^{p}\left\{\theta_{ij}\,\delta_{(w_{ij},1)}\,\delta_{(x_i,\,x_j)}+(1-\theta_{ij})\,\delta_{(w_{ij},0)}\right\}=\prod_{i=1}^{p-1}\prod_{j=i+1}^{p}\theta_{ij}^{w_{ij}}(1-\theta_{ij})^{1-w_{ij}}\,\mathbf{1}_F(\mathbf{w},\,\mathbf{x}),$$

where $\mathbf{1}_F(\mathbf{w},\,\mathbf{x})$ is the indicator function that $\delta_{(x_i,\,x_j)}=1$ whenever $\delta_{(w_{ij},1)}=1$, i.e., an indicator function for the event

$$F=\{(\mathbf{x},\,\mathbf{w}):\delta_{(x_i,\,x_j)}=1\text{ whenever }\delta_{(w_{ij},\,1)}=1\}.$$

Observe that $\mathbf{1}_F(\mathbf{w},\,\mathbf{x})=1$ only when $\mathbf{x}$ is constant on every cluster of $\mathbf{w}$. But in that case

$$\prod_{i=1}^{p}\exp\left(x_i\mu_i\right)\mathbf{1}_F(\mathbf{w},\,\mathbf{x})=\prod_{c=1}^{\kappa(\mathbf{w})}\exp\left(\sum_{i\in V_c}x_i\mu_i\right)\mathbf{1}_F(\mathbf{w},\,\mathbf{x})\Rightarrow\prod_{c=1}^{\kappa(\mathbf{w})}\exp\left(x_c\sum_{i\in V_c}\mu_i\right),$$

where $\kappa(\mathbf{w})$ denotes the number of open clusters, V_c is the set of nodes in the c-th cluster, and x_c is the state of the nodes in the c-th cluster. We can now express the marginal model $p(\mathbf{w})$ as

$$\begin{aligned}p(\mathbf{W}=\mathbf{w})&=\frac{\prod_{i=1}^{p-1}\prod_{j=i+1}^{p}\theta_{ij}^{w_{ij}}(1-\theta_{ij})^{1-w_{ij}}\,\sum_{\mathbf{x}}\prod_{c=1}^{\kappa(\mathbf{w})}\exp\left(\sum_{i\in V_c}x_i\mu_i\right)\mathbf{1}_F(\mathbf{w},\,\mathbf{x})}{\sum_{\mathbf{w}}\prod_{i=1}^{p-1}\prod_{j=i+1}^{p}\theta_{ij}^{w_{ij}}(1-\theta_{ij})^{1-w_{ij}}\,\sum_{\mathbf{x}}\prod_{c=1}^{\kappa(\mathbf{w})}\exp\left(\sum_{i\in V_c}x_i\mu_i\right)\mathbf{1}_F(\mathbf{w},\,\mathbf{x})}\\&=\frac{\prod_{i=1}^{p-1}\prod_{j=i+1}^{p}\theta_{ij}^{w_{ij}}(1-\theta_{ij})^{1-w_{ij}}\,\prod_{c=1}^{\kappa(\mathbf{w})}\beta_c}{\sum_{\mathbf{w}}\prod_{i=1}^{p-1}\prod_{j=i+1}^{p}\theta_{ij}^{w_{ij}}(1-\theta_{ij})^{1-w_{ij}}\,\prod_{c=1}^{\kappa(\mathbf{w})}\beta_c},\end{aligned}$$

where we recognize the cluster-weight $\beta_c=2\cosh\left(\sum_{i\in V_c}\mu_i\right)$ of Cioletti and Vila (2016). Given the expression for the marginal distribution, it is now a trivial exercise to find the conditional distribution

$$\begin{aligned}p(\mathbf{x}\mid\mathbf{w})&=\frac{\dfrac{\prod_{i=1}^{p-1}\prod_{j=i+1}^{p}\theta_{ij}^{w_{ij}}(1-\theta_{ij})^{1-w_{ij}}\,\prod_{c=1}^{\kappa(\mathbf{w})}\exp\left(\sum_{i\in V_c}x_i\mu_i\right)}{\sum_{\mathbf{w}}\prod_{i=1}^{p-1}\prod_{j=i+1}^{p}\theta_{ij}^{w_{ij}}(1-\theta_{ij})^{1-w_{ij}}\,\prod_{c=1}^{\kappa(\mathbf{w})}\beta_c}}{\dfrac{\prod_{i=1}^{p-1}\prod_{j=i+1}^{p}\theta_{ij}^{w_{ij}}(1-\theta_{ij})^{1-w_{ij}}\,\prod_{c=1}^{\kappa(\mathbf{w})}\beta_c}{\sum_{\mathbf{w}}\prod_{i=1}^{p-1}\prod_{j=i+1}^{p}\theta_{ij}^{w_{ij}}(1-\theta_{ij})^{1-w_{ij}}\,\prod_{c=1}^{\kappa(\mathbf{w})}\beta_c}}\\&=\frac{\prod_{i=1}^{p-1}\prod_{j=i+1}^{p}\theta_{ij}^{w_{ij}}(1-\theta_{ij})^{1-w_{ij}}\,\prod_{c=1}^{\kappa(\mathbf{w})}\exp\left(\sum_{i\in V_c}x_i\mu_i\right)}{\prod_{i=1}^{p-1}\prod_{j=i+1}^{p}\theta_{ij}^{w_{ij}}(1-\theta_{ij})^{1-w_{ij}}\,\prod_{c=1}^{\kappa(\mathbf{w})}\beta_c}\\&=\prod_{c=1}^{\kappa(\mathbf{w})}\frac{\exp\left(\sum_{i\in V_c}x_i\mu_i\right)}{\beta_c}\\&=\prod_{c=1}^{\kappa(\mathbf{w})}\frac{\exp\left(\sum_{i\in V_c}x_i\mu_i\right)}{\exp\left(\sum_{i\in V_c}\mu_i\right)+\exp\left(-\sum_{i\in V_c}\mu_i\right)},\end{aligned}$$

This confirms that Eqs. (2) and (4) are indeed equivalent, and we can now use the characterization (4) to work out the relation to the Ising model in Eq. (1).

We wish to show that the marginal distribution $p(\mathbf{x})$ of Eq. (4) gives the Ising model. To this aim, we start with summing over the edge configurations $\mathbf{w}$, and readily find

$$p(\mathbf{X}=\mathbf{x}) = \frac{\prod_{i=1}^{p-1}\prod_{j=i+1}^{p}\left\{\theta_{ij}\delta_{(x_i,x_j)}+1-\theta_{ij}\right\}\exp\left(\sum_{i=1}^{p}\mu_i x_i\right)}{\sum_{\mathbf{x}}\prod_{i=1}^{p-1}\prod_{j=i+1}^{p}\left\{\theta_{ij}\delta_{(x_i,x_j)}+1-\theta_{ij}\right\}\exp\left(\sum_{i=1}^{p}\mu_i x_i\right)}.$$

Next, we plug in the relation $\theta_{ij}=1-\exp(-2\sigma_{ij})$ from Eq. (3) to obtain

$$\begin{aligned}
p(\mathbf{X}=\mathbf{x}) &= \frac{\prod_{i=1}^{p-1}\prod_{j=i+1}^{p}\left\{(1-\exp(-2\sigma_{ij}))\delta_{(x_i,x_j)}+\exp(-2\sigma_{ij})\right\}\exp\left(\sum_{i=1}^{p}\mu_i x_i\right)}{\sum_{\mathbf{x}}\prod_{i=1}^{p-1}\prod_{j=i+1}^{p}\left\{(1-\exp(-2\sigma_{ij}))\delta_{(x_i,x_j)}+\exp(-2\sigma_{ij})\right\}\exp\left(\sum_{i=1}^{p}\mu_i x_i\right)}\\
&= \frac{\prod_{i=1}^{p-1}\prod_{j=i+1}^{p}\exp\left(2\sigma_{ij}\delta_{(x_i,\,x_j)}\right)\exp\left(\sum_{i=1}^{p}\mu_i x_i\right)}{\sum_{\mathbf{x}}\prod_{i=1}^{p-1}\prod_{j=i+1}^{p}\exp\left(2\sigma_{ij}\delta_{(x_i,\,x_j)}\right)\exp\left(\sum_{i=1}^{p}\mu_i x_i\right)},
\end{aligned}$$

and since $\delta_{(x_i,\,x_j)}=\frac{1}{2}(1+x_i x_j)$, we have, in fact, obtained the Ising model

$$\begin{aligned}
p(\mathbf{X}=\mathbf{x}) &= \frac{\exp\left(\sum_{i=1}^{p}\mu_i x_i+\sum_{i=1}^{p-1}\sum_{j=i+1}^{p}\sigma_{ij}x_i x_j+\sum_{i=1}^{p-1}\sum_{j=i+1}^{p}\sigma_{ij}\right)}{\sum_{\mathbf{x}}\exp\left(\sum_{i=1}^{p}\mu_i x_i+\sum_{i=1}^{p-1}\sum_{j=i+1}^{p}\sigma_{ij}x_i x_j+\sum_{i=1}^{p-1}\sum_{j=i+1}^{p}\sigma_{ij}\right)}\\
&= \frac{\exp\left(\sum_{i=1}^{p}\mu_i x_i+\sum_{i=1}^{p-1}\sum_{j=i+1}^{p}\sigma_{ij}x_i x_j\right)}{\sum_{\mathbf{x}}\exp\left(\sum_{i=1}^{p}\mu_i x_i+\sum_{i=1}^{p-1}\sum_{j=i+1}^{p}\sigma_{ij}x_i x_j\right)}.
\end{aligned}$$