OPEN

Bayesian inference for low-rank Ising networks

SUBJECT AREAS:

STATISTICS

COMPUTATIONAL SCIENCE

Maarten Marsman¹, Gunter Maris^{1,2}, Timo Bechger¹ & Cees Glas³

Received

20 August 2014

Accepted

12 February 2015

Published

12 March 2015

¹Psychometric Research Center, Cito, the Netherlands, ²Department of Psychology, University of Amsterdam, the Netherlands, ³Department of Research Methodology, Measurement and Data Analysis, University of Twente, the Netherlands.

Estimating the structure of Ising networks is a notoriously difficult problem. We demonstrate that using a latent variable representation of the Ising network, we can employ a full-data-information approach to uncover the network structure. Thereby, only ignoring information encoded in the prior distribution (of the latent variables). The full-data-information approach avoids having to compute the partition function and is thus computationally feasible, even for networks with many nodes. We illustrate the full-data-information approach with the estimation of dense networks.

Correspondence and requests for materials should be addressed to M.M. (m.marsman@uva.nl)

Modelling the joint distribution of binary variables is of importance in many fields of science, ranging from the study of phase transitions in statistical mechanics¹ and the study of spatial statistics in biology² to the study of comorbidity of mental disorder symptoms in psychiatry³. The Ising model⁴ is an appropriate model for such distributions, as it captures all main effects and pairwise interactions⁵.

Applications of the Ising model come in two distinct flavours, which are represented schematically in Figure 1 using the GNU-R package qgraph⁶. On the left hand side, variables interact only with their nearest neighbours. This is the situation for which physicists originally developed the Ising model. On the right hand side, every variable is (positively) correlated with nearly all other variables. This situation is typical for the social sciences^{7–9}.

A key distinction between social science and physical science applications of the Ising model is in the rank of the connectivity matrix, which contains the strength of the pairwise interactions. Figure 2 presents eigenvalue spectra for each of the two networks in Figure 1. A near linear spectrum is found in the nearest neighbour network, whereas the eigenvalues in the dense network rapidly decrease in magnitude. As shown in the Methods section, the eigenvalues dictate the contribution of dimensions in the predictive distributions of variables in the network, and we have that the smaller eigenvalues in the dense network hold low predictive value in comparison with the larger eigenvalues. This observation suggests that the connectivity matrix of a dense network is well approximated by a low-rank matrix.

A vital property of low-rank approximations follows from a theorem published by Eckart and Young in the first volume of the journal *Psychometrika*¹⁰. The Eckart-Young Theorem shows that in a least-squares sense, the best approximation of rank r to a matrix consists of the eigenvalue decomposition in which all but the largest r eigenvalues are equated to zero. This theorem allows us to find a low-rank approximation to the full connectivity matrix, with the key property that the first eigenvalues and their corresponding eigenvectors can be recovered even if the remaining eigenvalues and eigenvectors are ignored. To demonstrate the power of this result, Figure 3 shows the rank four approximation to the two networks in Figure 1. As expected, the nearest neighbour network is not recovered in the rank four approximation, yet the dense network clearly is. This shows that this parsimonious low-rank approximation is useful to uncover the structure of the connectivity matrices of dense networks. In both cases, the Eckart-Young Theorem implies that the estimated eigenvectors in the low-rank approximation are the first few eigenvectors of the true connectivity matrix.

The connectivity matrix encoding the network structure is usually unknown and needs to be estimated from independent realisations of the network state. Estimating the network structure is difficult, however, because the likelihood is intractable and the number of unknown parameters is usually very large. As a solution, the pseudo-likelihood method¹¹ has been developed. This method optimises each of the full-conditional (predictive) distributions of variables in the network, often in combination with regularisation constraints, either on the interactions in sparse networks and/or on the eigenvalues in dense networks. Here, we introduce an alternative that uses all available information from the data and not just the information in the full-conditional distributions. Although not the topic of this paper, this full-data-information method can be used in conjunction with regularisations of the types described above through a judicious choice of prior distributions.

Figure 1 | Two distinct flavours in applications of the Ising model.

The full-data-information approximation uses a latent variable representation of the Ising model that was developed by Kac¹², further developed by Emch and Knops¹³ and independently discovered in many places in the statistical literature^{2,14–18}. Specifically, every eigenvector for a connectivity matrix gives rise to a latent variable, such that all variables are independent conditionally on the full set of latent variables:

$$\exists \Theta : \perp \mathbf{X} | \Theta. \quad (1)$$

That is, there exist latent variables (Θ) that explain all the pairwise interactions in a statistical sense. The distribution of the variables conditionally on the latent variables is known as a multidimensional Item Response Theory (IRT) model in the field of psychometrics¹⁹, see the Methods section. The insight of Kac, Emch and Knops is schematically represented in Figure 4 using four latent variables. Ignoring some of the latent variables by equating the smallest eigen-

values to zero amounts to ignoring residual pairwise interactions but leaves the recovered eigenvalues and corresponding eigenvectors unaffected.

Results

Full-data-information estimation. The Ising model is mathematically elegant, yet notoriously difficult to compute. The main problem is the normalising constant Z in equation (5), called the partition function, which involves a sum over all 2^n possible states of an n variable network. As the partition function depends on all of the model parameters, likelihood based statistical inference is impossible except for very small or severely constrained Ising models.

The computational problem becomes more tractable when we use the latent variable representation of the Ising model. The conditional distribution of the observed variables conditionally on the latent variables does not depend on the partition function and is available

Figure 2 | Key distinctions in eigenvalue spectra.

Figure 3 | Rank four approximation to the networks in Figure 1.

in an easily computed closed form. The partition function only figures in the distribution of the latent variables themselves. The posterior distribution of the Ising model parameters $\mathbf{W}$ (i.e., the connectivity matrix and main effects) and the latent variables θ given the data σ is proportional to:

$$f(\mathbf{W}, \theta | \sigma) \propto p(\sigma | \theta, \mathbf{W}) f(\theta | \mathbf{W}) f(\mathbf{W}), \quad (2)$$

where $p(\sigma | \theta, \mathbf{W})$ and $f(\theta | \mathbf{W})$ are the multidimensional IRT model and latent variable distribution, as derived in the Methods section, and $f(\mathbf{W})$ is a prior distribution for the model parameters. The whole computational complexity of this posterior distribution resides in the distribution of the latent variables, which depends on the model parameters and, in particular, on the partition function. Considering a Gibbs sampler²⁰ for simulating from this posterior

distribution, we find that the full conditional distribution $f(\theta | \mathbf{W}, \sigma)$ of the latent variables is highly tractable and does not involve the partition function, whereas the full conditional distribution $f(\mathbf{W} | \theta, \sigma)$ of the Ising model parameters is intractable because it involves the partition function.

When the latent variable distribution $f(\theta | \mathbf{W})$ in the Ising model is replaced by a prior distribution $g(\theta)$ that does not depend on the model parameters, we have a regular multidimensional IRT problem¹⁹:

$$g(\mathbf{W}, \theta | \sigma) \propto p(\sigma | \theta, \mathbf{W}) g(\theta) f(\mathbf{W}), \quad (3)$$

for which the full-conditionals $g(\theta | \mathbf{W}, \sigma)$ and $g(\mathbf{W} | \theta, \sigma)$ are easily sampled from. Multidimensional IRT models of this form are commonly applied in educational²¹ and psychological assessment^{22,23}.

Figure 4 | Rank four latent variable approximation to the dense network.

Figure 5 | The scaled densities $f(\theta|W)$ in a Curie-Weiss model for $a = \frac{1}{4}$ given as a solid line, $a = \frac{1}{1.9}$ as a dashed line and $a = 1$ as a dotted line.

Typically, the prior distribution $g(\theta)$ is a multivariate normal, and multidimensional IRT models with this choice of prior can be estimated using the MIRT package [www.utwente.nl/gw/omd/medewerkers/temp_test/mirt-manual.pdf].

It is clear that whenever $g(\theta)$ closely approximates $f(\theta|W)$, the much simpler problem in equation (3) can be used to infer about the problem in equation (2). Whether $g(\theta)$ closely approximates $f(\theta|W)$ is an empirical question. Figure 5 shows the (scaled) density $f(\theta|W)$ in the simplest nontrivial case, a fully connected network with all pairwise interactions equal to a , known as the Curie-Weiss model, for which $p(\sigma|\theta, W)$ is a Rasch model²⁴. It is clear that the typically used normal approximation would work well for $a = 1/4$; yet, when a increases, $f(\theta|W)$ first becomes skewed ($a = 1/1.9$) and ultimately becomes bimodal ($a = 1$) with modes drifting further apart as a increases.

In general, we cannot assume that the multivariate normal prior $g(\theta)$ closely approximates the Ising models latent variable distribution $f(\theta|W)$; thus, a different approach is needed. As we illustrate here and show in the Methods section, we can disregard the distribution of the latent variables $f(\theta|W)$ when we simulate from the conditional distributions of the Ising models parameters. In this way, all conditional distributions become tractable and, at the same time, all direct information on them provided by the observed variables is retained. The only information we ignore is that which is encoded in the prior distribution for the latent variables. That is, we combine the full-conditional of the latent variables from equation (2) with the full-conditional of the model parameters from equation (3). We call this approach full-data-information estimation.

We illustrate this using a simple Curie-Weiss model involving only pairwise interactions. This simplified model only involves one parameter, the unknown interaction strength a . In Figure 6, we see that the approximate posterior distribution nicely covers the true parameter value and becomes more concentrated around this value as the sample size increases.

Data example. It is the combination of the Eckart-Young Theorem and the latent variable representation with full-data-information estimation that allows us to estimate low-rank Ising networks as we see them in the social sciences. To illustrate the approach, we consider a large educational measurement application. The Cito Eindtoets (www.cito.com) is a test consisting of 200 questions

Figure 6 | Partial conditional distribution of the interaction strength a in a Curie-Weiss model for $N = 100$ replications (solid line) and $N = 1,000$ replications (dashed line). The grey vertical line represents the true value.

related to 12 theoretically distinct primary school subjects in arithmetic, language, and general study skills. The test is administered yearly to some 130,000 children at the end of Dutch primary education. We present here the results from a rank three approximation to data from the 2012 Eindtoets.

Figure 7 displays both the rank three approximation and the individual rank one components as a heatmap. As argued above, even though the true connectivity matrix might be of a much higher rank than three, the three estimated components correspond to the three eigenvectors of the true connectivity matrix with the highest eigenvalues. The first component corresponds to a network in which all nodes are connected to one another, and (almost) all interactions are positive. The second and third component are such that particular sets of questions get higher positive interactions amongst themselves, whereas interactions between questions from different sets are negative. The second component is a contrast between the different language subjects (writing (W), spelling (S), and reading comprehension and vocabulary (RV)) and the subjects of mathematics (M) and study skills (SK). The third component is a contrast between the spelling subject and the other language subjects combined with the study skills subject. Note that, the positive interactions in one component can cancel the negative interactions in another component. For instance, mathematics and study skills have a positive interaction in the first and second component, whereas in the third component they have a negative interaction.

From the network replications (pupil responses) σ_p , $p = 1, \dots, N$, we have that the matrix

$$S = [S_{ij}] = \sum_{p=1}^N \sigma_{pi} \sigma_{pj}, \quad (4)$$

is sufficient for the connectivity matrix. We now study how the rank of the approximation impacts the prediction of S . To this aim, we perform a posterior predictive check and use the latent variable model $p(\sigma|W_r^*, \theta_r^*)$ to generate new data, given draws W_r^* and θ_r^* from the (partial-)posterior distributions (see the Methods section) of a rank r approximation; we then construct a new matrix S_r^* from the newly generated data. We show the residuals $S - S_1^*$ in Figure 8(a) as a heatmap, and similarly show the residuals $S - S_2^*$ in Figure 8(b) and $S - S_3^*$ in Figure 8(c). In Figure 8(a), (b) and (c), we see that

Figure 7 | Heatmap of the connectivity matrix estimated in a rank three approximation to the 2012 Eindtoets data. The connectivity matrix is the sum of three rank-one matrices. Negative interactions are in red, positive interactions in blue, and small or absent interactions in grey.

increasing the rank results in predicting more of the underlying structure in S . For instance, from Figure 7 we know that the second eigenvector captures the relation between mathematics and language items, and this structure in S is visible in the residuals $S - S_1^*$ but not in the residuals $S - S_2^*$.

Although higher rank approximations predict more of the structure in S , the rank one approximation already captures most of the variation in S . To see this, we plot the lower triangular elements of the matrix S against the lower triangular elements of S_1^* in Figure 8(d), against S_2^* in Figure 8(e) and against S_3^* in Figure 8(f). The lower triangular elements of S_1^* are highly correlated with the lower triangular elements from S (the correlation equals 0.995) with only minor improvements in the correlation for higher rank approximations. The correlation equals 0.997 in the rank two approximation and 0.998 in the rank three approximation.

Discussion

We have shown how the Ising model could be estimated using full-data-information, in which we ignore prior structure on the para-

meters that resides in the latent variable model, and thereby effectively eliminate the need for computing the partition function from the statistical inference whilst retaining full-data-information. This approximate estimation technique opens the door for the estimation of other models, such as the Potts model²⁵ or a mix of models for discrete and continuous random variables.

As depicted in Figure 2(b), typical eigenvalue spectra found in the social sciences have a sharp drop in magnitude for the first few eigenvalues, after which the eigenvalues slowly decay. These plots resemble a mountain cliff with broken rock fragments at the base and are therefore called scree plots, where scree refers to the set of slowly decaying eigenvalues after the elbow. Scree plots are used to determine the relative importance of the eigenvalues, in which values after the elbow are often assumed ignorable due to, for instance, sampling error. However, the eigenvalues after the elbow in Figure 2(b) have a near linear spectrum, which resembles the spectrum found for the nearest neighbour network in Figure 2(a). This suggests that social science applications call for a mix of a dense network approximation, to use for the first few eigenvectors, and the nearest neighbour

Figure 8 | A heatmap of the matrix of residuals $S - S_1^*$ is shown in (a), that of $S - S_2^*$ is shown in (b) and that of $S - S_3^*$ is shown in (c). Negative residuals are in red, positive residuals in blue, and small or absent residuals in grey. Also shown are the scatterplots of lower triangle elements of S against S_1^* in (d), against S_2^* in (e) and against S_3^* in (f). The green line is the first bisection.

approximation of van Borkulo et al.²⁶, to explain the residual structure.

The latent variable representation opens the way to new areas of research. For instance, the IRT model is closed under marginalisation, which shows that the IRT problem in equation (3) can be used as a starting point to approximate incomplete networks.

Methods

From Ising to Item Response Theory. The Ising model is characterised by the following distribution:

$$p(\sigma|\mathbf{A}, \mathbf{b}) = \frac{1}{Z} \exp\{\sigma^T \mathbf{A} \sigma + \sigma^T \mathbf{b}\}, \quad (5)$$

where the partition function Z serves to make the distribution sum to one and is a function of the main effects $\mathbf{b}$ and the connectivity matrix $\mathbf{A}$ containing the pairwise interactions. It is readily observed that all parameters are identifiable from the data, except for entries on the diagonal of the connectivity matrix.

Choosing diagonal values for the connectivity matrix such that all eigenvalues are non-negative, we obtain:

$$\mathbf{A} + \mathbf{cI} = \mathbf{Q}(\mathbf{A} + \mathbf{cI})\mathbf{Q}^T = \mathbf{E}^T \mathbf{E} \quad (6)$$

where $\mathbf{A}$ is a diagonal matrix with the eigenvalues of the original matrix $\mathbf{A}$. In this expression, we conserve the off-diagonal entries in the connectivity matrix and, at the same time, ensure that the matrix $\mathbf{E}^T \mathbf{E}$ is positive (semi-)definite. This allows us to use the well known Gaussian identity to represent the Ising model equivalently in the following form:

$$p(\sigma|\mathbf{E}, \mathbf{b}) = \int_{\mathbb{R}^n} \frac{1}{Z\sqrt{\pi}^n} \exp\{\sigma^T \mathbf{b} + 2\sigma^T \mathbf{E} \theta - \theta^T \theta\} d\theta. \quad (7)$$

In this expression, the quadratic form is linearised, allowing for an explicit factorisation:

$$p(\sigma|\mathbf{E}, \mathbf{b}) = \int_{\mathbb{R}^n} \prod_i p(\sigma_i|\theta, e_i, b_i) f(\theta|\mathbf{E}, \mathbf{b}) d\theta. \quad (8)$$

We can now recognise a multidimensional IRT model:

$$p(\sigma_i|\theta, e_i, b_i) = \frac{\exp\{\sigma_i[b_i + 2e_i\theta]\}}{\exp\{[b_i + 2e_i\theta]\} + \exp\{-[b_i + 2e_i\theta]\}} \quad (9)$$

with a particular distribution for the latent variables:

$$f(\theta|\mathbf{E}, \mathbf{b}) = \frac{\prod_i (\exp\{[b_i + 2e_i\theta]\} + \exp\{-[b_i + 2e_i\theta]\})}{Z\sqrt{\pi}^n} \exp\{-\theta^T \theta\}, \quad (10)$$

where $\mathbf{e}_i$ is the i -th row-vector of $\mathbf{E}$. Note that, in this representation, the partition function figures as a normalising constant of the latent variable distribution.

The full-conditional distribution $p(\sigma_i|\sigma_{\setminus i})$. The full-conditional distribution of a variable σ_i given the other variables $\sigma_{\setminus i}$ equals:

$$p(\sigma_i|\sigma_{\setminus i}, \mathbf{A}, \mathbf{b}) = \frac{\exp(\sigma_i [b_i + 2 \sum_{r=1}^n \lambda_r q_{ir} \sum_{j \neq i} \sigma_j q_{jr}])}{2 \cosh(b_i + 2 \sum_{r=1}^n \lambda_r q_{ir} \sum_{j \neq i} \sigma_j q_{jr})}, \quad (11)$$

where $\sum_{j \neq i} \sigma_j q_{jr}$ is recognised as the r -th principal component rest-score (i.e., the r -th principal component score minus $\sigma_i q_{ir}$). That is, $p(\sigma_i|\sigma_{\setminus i}, \mathbf{A}, \mathbf{b})$ is a logistic regression model with intercept b_i , principal component rest-scores as predictor variables and eigenvalues (times the i -th entry in the eigenvector) as regression coefficients.

Full-data-information estimation. Upon choosing a proper prior distribution for the model parameters $\mathbf{E}$ and $\mathbf{b}$, we obtain a posterior distribution for both the model parameters and the latent variables. It is not simple to simulate directly from this joint posterior distribution, thus we use the Gibbs sampler²⁰. The full-conditional distribution of the latent variables corresponding to observation p , $p = 1, \dots, N$, is a multivariate normal distribution:

$$f(\theta_p|\mathbf{E}, \mathbf{b}, \sigma_p) \propto \exp\{2\sigma_p^T \mathbf{E} \theta_p - \theta_p^T \theta_p\}. \quad (12)$$

For the model parameters, we find more complicated full-conditional distributions that depend on the partition function. Because the partition function only figures as the normalising constant of the latent variable distribution, our proposal is to simulate the model parameters from the *partial* conditional distribution:

$$f_{\text{partial}}(\mathbf{E}, \mathbf{b}|\theta, \sigma) \propto \prod_p p(\sigma_p|\theta_p, e_i, b_i) f(\mathbf{E}, \mathbf{b}). \quad (13)$$

This partial conditional distribution ignores the information encoded in the marginal distribution of the latent variables but retains all the information about the model

parameters that is contained in the data. This is why we refer to our approach as full-data-information estimation.

We alternately sample from the full-conditional distribution of the latent variables and the partial conditionals of the Ising model parameters. This does not amount to a proper Gibbs sampler in the sense that, after discarding the latent variables, the posterior distribution of the model parameters is not the invariant distribution. One way of looking at this scheme is that a second posterior distribution is set up for the latent variables and model parameters, one in which we have a proper prior distribution for the latent variables that does not depend on the model parameters. In this second posterior distribution, the partial full conditional distributions of the model parameters would be the correct full conditional distributions in a Gibbs sampler and it would be the multivariate normal full conditional distribution for the latent variables that would no longer be correct. When the Bayesian Central Limit Theorem would be the same for both posteriors, our scheme that mixes the full-conditional for the one posterior with another from the other posterior would still converge to the same Central Limit Theorem. In that sense, our full-data-information estimation admits the same large sample properties as does the full Bayesian estimation but without ever having to compute the partition function.

Simulating from the partial conditionals. Simulating from the partial-conditional distribution of the parameters may seem to be a difficult problem in its own right, but this problem has been resolved in many different places^{27–31}. We consider here a new method that is both computationally simple and highly efficient.

Our proposal is to use a Metropolis-Hastings algorithm^{32,33} for simulating from the partial-conditional distribution of the model parameters, which differs from earlier such approaches^{27,28} in the particular choice of the proposal distribution. We consider an independence Metropolis-Hastings algorithm³⁴ in which the proposal distribution is easy to simulate from and the approximation of its target improves as the amount of data increases. This combination makes the algorithm ideally suited for large data sets.

Consider a set of random variables $\mathbf{Z}$, such that $\mathbb{I}_p Z_p$ and $Z_p \sim F_p$. Define a matrix of binary indicator variables (coded as zero/one) with entries $x_{pj} = (Z_p < z_j)$, such that for column j of the matrix we obtain:

$$f_j(z_j|x_j) \propto \prod_{p \neq j} F_p(z_j)^{x_{pj}} (1 - F_p(z_j))^{1-x_{pj}} f_j(z_j) = f(\mathbf{x}_j, z_j). \quad (14)$$

This distribution closely resembles the partial-conditionals. For any parameter W , the partial conditional is of the form:

$$f_{\text{partial}}(w|\mathbf{y}) \propto \prod_{p=1}^N F_p(w)^{y_p} (1 - F_p(w))^{1-y_p} f(w), \quad (15)$$

where $y_p = (\sigma_p + 1)/2$, F_p is a (logistic) distribution function and $f(w)$ the prior density of W . Thus, $f_j(z_j|x_j)$ will be used as a proposal density.

To illustrate how the algorithm works, consider a simple case with $N = 2$ and a (target) partial conditional:

$$f_{\text{partial}}(w|\mathbf{y}) \propto F_1(w)(1 - F_2(w))f(w), \quad (16)$$

where $y_+ = \sum_p y_p = 1$. We now generate the vector $\mathbf{z}$ and choose j (in equation (14)) such that $x_+ = \sum_{p \neq j} x_{pj} = y_+$ and find (for instance):

$$f(z_j|x_j) \propto F_2(z_j)(1 - F(z_j))f_1(z_j). \quad (17)$$

That is, z_j is a draw from a posterior based on N observations and a total score x_+ , which differs from the target distribution w.r.t. the distribution of one of the observations and the prior density. In fact, the prior density and the distribution of the first observation have switched places.

That this method is feasible in practice is apparent from the Eindtoets data example. The Eindtoets data that we have used consisted of 200 variables (items) with 133,768 replications (pupils), and we considered a rank three approximation. A simple GNU-R implementation on a MacBook Pro with a 2.6 GHz Intel Core i5 processor (single core) took approximately 7 seconds to update the 200 unknown values in $\mathbf{b}$ and approximately 35 seconds to update the 600 unknown values in $\mathbf{E}$. The average acceptance rate in 1,000 iterations of the Gibbs sampler was equal to 0.99 for the elements in $\mathbf{b}$ and 0.42 for the elements in $\mathbf{E}$.

- Lee, T. & Yang, C. Statistical theory of equations of state and phase transitions II. Lattice gas and Ising model. *Phys. Rev.* **87**, 410–419 (1952).
- Besag, J. Spatial interaction and the statistical analysis of lattice systems. *J. R. Stat. Soc. Series B Stat. Methodol.* **36**, 192–236 (1974).
- Cramer, A., Waldorp, L., van der Maas, H. & Borsboom, D. Comorbidity: A network perspective. *Behav. Brain Sci.* **33**, 137–150 (2010).
- Ising, E. Beitrag zur theorie des ferromagnetismus. *Zeit. Phys.* **31**, 253–258 (1925).
- Jaynes, E. Information theory and statistical mechanics. *Phys. Rev.* **106**, 620–630 (1957).
- Epskamp, S., Cramer, A., Waldorp, L., Schmittmann, V. & Borsboom, D. qgraph: Network visualizations of relationships in psychometric data. *J. Stat. Softw.* **48**, 1–18 (2012).
- Jensen, A. *The g factor: The science of mental ability* (Praeger, Westport, 1998).

8. Deary, I. *Looking Down on Human Intelligence: From Psychometrics to the Brain*. (Oxford University Press, Oxford, 2000).
9. van der Maas, H. *et al.* A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychol. Rev.* **113**, 842–861 (2006).
10. Eckart, C. & Young, G. The approximation of one matrix by another of lower rank. *Psychometrika* **1**, 211–218 (1936).
11. Besag, J. Statistical analysis of non-lattice data. *Statistician* **24**, 179–195 (1975).
12. Kac, M. *Statistical Physics: Phase Transitions and Superfluidity*, Vol. 1, *Brandeis University Summer Institute in Theoretical Physics*. [Chretien, M., Gross, E. & Deser, S. (eds.)] [241–305] (Gordon and Breach Science Publishers, New York, 1968).
13. Emch, G. & Knops, H. Pure thermodynamical phases as extremal KMS states. *J. Math. Phys.* **11**, 3008–3018 (1970).
14. Olkin, I. & Tate, R. Multivariate correlation models with mixed discrete and continuous variables. *Ann. Math. Stat.* **32**, 448–465 (1961).
15. Lauritzen, S. & Wermuth, N. Graphical models for associations between variables, some of which are qualitative and some quantitative. *Ann. Stat.* **17**, 31–57 (1989).
16. McCullagh, P. Exponential mixtures and quadratic exponential families. *Biometrika* **81**, 721–729 (1994).
17. Anderson, C. & Yu, H. Log-multiplicative association models as item response models. *Psychometrika* **72**, 5–23 (2007).
18. Anderson, C., Li, Z. & Vermunt, J. Estimation of models in a Rasch family of polytomous items and multiple latent variables. *J. Stat. Softw.* **20** (2007).
19. Reckase, M. *Multidimensional item response theory* (Springer, 2009).
20. Geman, S. & Geman, D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 721–741 (1984).
21. Ackerman, T. Developments in multidimensional item response theory. *Appl. Psychol. Meas.* **20**, 309–310 (1996).
22. Makransky, G., Mortensen, E. & Glas, C. Improving personality facet scores with multidimensional computerized adaptive testing: An illustration with the NEO PI-R. *Assessment* **20**, 3–13 (2013).
23. Nikolaus, S. *et al.* Items and dimensions for the construction of a multidimensional computerized adaptive test to measure fatigue in patients with rheumatoid arthritis. *J. Clin. Epidemiol.* **66**, 1175–1183 (2013).
24. Rasch, G. *Probabilistic Models for some Intelligence and Attainment Tests* (The Danish Institute of Educational Research, Copenhagen, 1960).
25. Potts, R. Some generalized order-disorder transformations. *Math. Proc. Cambridge Philos. Soc.* **48**, 106–109 (1952).
26. van Borkulo, C. *et al.* A new method for constructing networks from binary data. *Sci. Rep.* **4** (2014).
27. Patz, R. & Junker, B. A straightforward approach to Markov chain Monte Carlo methods for item response models. *J. Educ. Behav. Stat.* **24**, 146–178 (1999).
28. Patz, R. & Junker, B. Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses. *J. Educ. Behav. Stat.* **24**, 342–366 (1999).
29. Albert, J. Bayesian estimation of Normal Ogive item response curves using Gibbs sampling. *J. Educ. Stat.* **17**, 251–269 (1992).
30. Béguin, A. & Glas, C. MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika* **66**, 541–562 (2001).
31. Maris, G. & Maris, E. A MCMC-method for models with continuous latent responses. *Psychometrika* **67**, 335–350 (2002).
32. Metropolis, N., Rosenbluth, A., Rosenbluth, M. & Teller, A. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092 (1953).
33. Hastings, W. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109 (1970).
34. Tierney, L. Markov chains for exploring posterior distributions. *Ann. Stat.* **22**, 1701–1762 (1994).

Author contributions

M.M. and G.M. wrote the main manuscript, T.B. and C.G. contributed to manuscript revisions, and M.M. prepared the figures. All authors reviewed the manuscript.

Additional information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Marsman, M., Maris, G., Bechger, T. & Glas, C. Bayesian inference for low-rank Ising networks. *Sci. Rep.* **5**, 9050; DOI:10.1038/srep09050 (2015).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>