Joint Modeling of Multiple Network Views

Isabella Gollini & Thomas Brendan Murphy

Joint Modeling of Multiple Network Views

Isabella GOLLINI and Thomas Brendan MURPHY

Latent space models (LSM) for network data rely on the basic assumption that each node of the network has an unknown position in a D-dimensional Euclidean latent space: generally the smaller the distance between two nodes in the latent space, the greater their probability of being connected. In this article, we propose a variational inference approach to estimate the intractable posterior of the LSM. In many cases, different network views on the same set of nodes are available. It can therefore be useful to build a model able to jointly summarize the information given by all the network views. For this purpose, we introduce the latent space joint model (LSJM) that merges the information given by multiple network views assuming that the probability of a node being connected with other nodes in each network view is explained by a unique latent variable. This model is demonstrated on the analysis of two datasets: an excerpt of 50 girls from "Teenage Friends and Lifestyle Study" data at three time points and the *Saccharomyces cerevisiae* genetic and physical protein–protein interactions. Supplementary materials for this article are available online.

Key Words: Latent space model; Latent variable; Social network analysis; Variational methods.

1. INTRODUCTION

Network data consist of a set of nodes and a list of edges between the nodes. Recently, there has been a growing interest in the modeling of network data. A number of models have been proposed for network data including exponential random graph models (ERGMs; Holland and Leinhardt 1981), stochastic blockmodels (Holland, Laskey, and Leinhardt 1983; Airoldi et al. 2008), and latent space models (Hoff, Raftery, and Handcock 2002; Handcock, Raftery, and Tantrum 2007). Recent reviews of various network modeling approaches include Goldenberg et al. (2010) and Salter-Townshend et al. (2012).

Latent space models (LSM) are a well-known family of latent variable models for network data introduced by Hoff, Raftery, and Handcock (2002) under the basic assumption that each node has an unknown position in a D-dimensional Euclidean latent space: generally the smaller the distance between two nodes in the latent space, the greater the

Isabella Gollini is Lecturer in Statistics, Department of Economics, Mathematics and Statistics, Birkbeck, University of London, England (E-mail: *i.gollini@bbk.ac.uk*). Thomas Brendan Murphy is Professor, School of Mathematical Sciences, Complex & Adaptive Systems Laboratory and Insight Research Centre, University College Dublin, Ireland (E-mail: *brendan.murphy@ucd.ie*).

probability of them being connected. Unfortunately, the posterior distribution of the LSM cannot be computed analytically. For this reason, we propose a variational inferential approach that proves to be less computationally intensive than the Markov chain Monte Carlo (MCMC) procedure proposed in Hoff, Raftery, and Handcock (2002) and can therefore easily handle large networks.

In many cases, multiple network link relations on the same set of nodes are available. Multiple network views, also known as multiplex networks (Mucha et al. 2010), can be intended either as multiple link relations among the nodes of the network or a single link relation observed over different conditions, such as one network evolving over time (longitudinal networks). To deal with multiplex networks, we present a latent space joint model (LSJM) that merges the information given by the multiple network views by assuming that the probability of a node being connected with other nodes in each view is explained by a unique latent variable. To estimate this model, we propose an expectation-maximization (EM) algorithm: the parameter estimates obtained from fitting an LSM for each network view independently are used to approximate the joint posterior distribution of the LSJM; then these results are used to update the parameter estimates of every LSM. This process is iterated until convergence. This model has a wide range of applications. For example, in computer science it is of interest to summarize the different relations (e.g., friend, fan, follower, or like) that we observe in social media sites like Facebook, Twitter, and YouTube (Tang, Wang, and Liu 2011). Another important application is in systems biology where the joint modeling of physical and genetic protein–protein interactions is of wide interest (Bandyopadhyay et al. 2008). Other contexts in which this model can be useful include social sciences and business marketing (Ansari, Koenigsberg, and Stahl 2011). The LSJM is demonstrated on the analysis of an excerpt of 50 girls from "Teenage Friends and Lifestyle Study" data at three time points (Pearson and Michell 2000; Pearson and West 2003), and two *Saccharomyces cerevisiae* networks (Stark et al. 2006). All the methods of this article are implemented in the `lvm4net` package for R (R Core Team 2014).

The article is organized as follows. Section 2 provides an introduction to latent space models for network data with a particular focus on the variational inference approach to fit the latent space model. In Section 3, we introduce the latent space joint model for multiple network view data. In Section 4, we show how missing link data can be managed using the LSJM. In Section 5, we illustrate the capabilities of the LSJM and we analyze its performance in the presence of missing edges by using cross-validation; the model is illustrated on the two example datasets (Sections 5.2 and 5.3). We conclude, in Section 6, with a discussion of the model.

2. LATENT SPACE MODEL

Latent space models for network data have been introduced by Hoff, Raftery, and Handcock (2002) under the basic assumption that each node i has an unknown position $\mathbf{z}_i$ in a D-dimensional Euclidean latent space. The *distance model* is an easy-to-interpret LSM, which is based on the distance between the nodes in the latent space. Generally, the smaller the distance between two nodes in the latent space, the greater the probability that they connect. This model supposes the network to be intrinsically symmetric since the
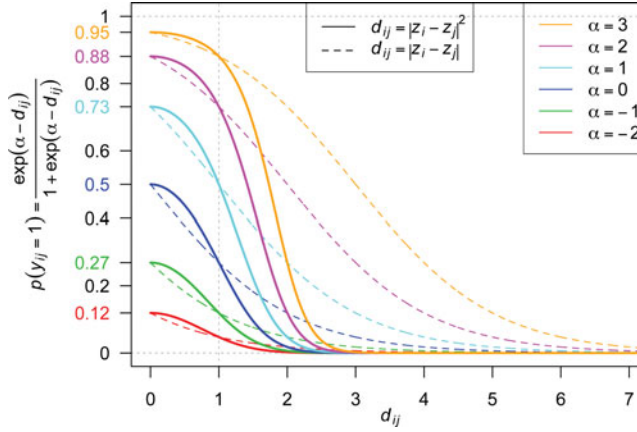
Figure 1.   Probability of $y_{ij} = 1$ as a function of the distance metrics. For $d_{ij} < 1$, the solid lines representing the probability of a link based on the squared Euclidean metric are higher than the dotted lines representing the probability of a link based on the Euclidean distance. For $d_{ij} > 1$, the solid lines decrease more rapidly than the dotted lines.

distance between nodes in the latent space is symmetric and thus it has the feature of being reciprocal: if $y_{ij} = 1$, then the probability of $y_{ji} = 1$ is large, where y_{ij} is the observed variable that is 1 if we observe a link from node i to node j, and 0 otherwise. For this reason, the distance model is particularly suitable for undirected networks or directed networks that exhibit strong reciprocity.

Let N be the number of observed nodes and $\mathbf{Y}$ be the $N \times N$ adjacency matrix containing the network information, with entries y_{ij} (where $y_{ij} = 0$ or 1), and null diagonal. Let $\mathbf{Z}$ be an $N \times D$ matrix of latent positions where each row is composed by $\mathbf{z}_i = (z_{i1}, \ldots, z_{iD})$ the D-dimensional vector indicating the position of observation i in the D-dimensional latent space. The latent space model can be written as

$$p(\mathbf{Y}|\mathbf{Z}, \alpha) = \prod_{i \neq j}^{N} p(y_{ij}|\mathbf{z}_i, \mathbf{z}_j, \alpha) = \prod_{i \neq j}^{N} \frac{\exp(\alpha - |\mathbf{z}_i - \mathbf{z}_j|^2)^{y_{ij}}}{1 + \exp(\alpha - |\mathbf{z}_i - \mathbf{z}_j|^2)},$$

where for ease of notation, $\prod_{i \neq j}^{N}$ is equivalent to $\prod_{i=1}^{N} \prod_{j=1, j \neq i}^{N}$.

We assume the following distributions for the model unknowns, where $p(\alpha) = \mathcal{N}(\xi, \psi^2)$, $p(\mathbf{z}_i) \overset{\text{iid}}{=} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I_D})$ and σ^2, ξ, ψ^2 are fixed parameters, and the squared Euclidean distance between observations i and j is $|\mathbf{z}_i - \mathbf{z}_j|^2 = \sum_{d=1}^{D} (z_{id} - z_{jd})^2$. The squared Euclidean distance measure is employed instead of the Euclidean distance used in Hoff, Raftery, and Handcock (2002). This choice has been made for two main reasons: first, it allows one to visualize the data more clearly, giving a higher probability of a link between two close nodes in the latent space and lower probabilities to two nodes lying far away from each other (see Figure 1); second, it requires fewer approximation steps to be made in the estimation procedure. An empirical comparison of the results obtained fitting an LSM using the Euclidean and the squared Euclidean distance is given in the supplementary material.

The posterior probability is of the unknown $(\mathbf{Z}, \alpha)$ is of the form

$$p(\mathbf{Z}, \alpha|\mathbf{Y}) = C p(\mathbf{Y}|\mathbf{Z}, \alpha) p(\alpha) \prod_{i=1}^{N} p(\mathbf{z}_i),$$

where C is the unknown normalizing constant.

2.1 VARIATIONAL INFERENCE APPROACH

Since the posterior distribution cannot be calculated analytically, we make use of a variational inference approach to estimate the model. To do this, we aim at maximizing a lower bound of the likelihood function. This approach has been proposed for several latent variable models (Attias 1999; Jordan et al. 1999) and we refer to Beal (2003) for an extensive introduction to the variational methods. In the statistical network models context, Airoldi et al. (2008) proposed the use of the variational method to fit mixed-membership stochastic blockmodels and Salter-Townshend and Murphy (2013) applied variational methods to fit the latent position cluster model (Handcock, Raftery, and Tantrum 2007); the latent position cluster model is an extension of the original LSM in which the latent positions are assumed to come from a Gaussian mixture model.

We define the variational posterior $q(\mathbf{Z}, \alpha|\mathbf{Y})$ introducing the variational parameters $\Theta = (\tilde{\xi}, \tilde{\psi}^2), \tilde{\mathbf{z}}_i$, and $\tilde{\Sigma}$:

$$q(\mathbf{Z}, \alpha|\mathbf{Y}) = q(\alpha) \prod_{i=1}^{N} q(\mathbf{z}_i),$$

where $q(\alpha) = \mathcal{N}(\tilde{\xi}, \tilde{\psi}^2)$ and $q(\mathbf{z}_i) = \mathcal{N}(\tilde{z}_i, \tilde{\Sigma})$.

The basic idea behind the variational approach is to find a lower bound of the log marginal likelihood $\log p(\mathbf{Y})$ by introducing the variational posterior distribution $q(\mathbf{Z}, \alpha|\mathbf{Y})$. This approach leads to minimize the Kullback–Leibler divergence between the variational posterior $q(\mathbf{Z}, \alpha|\mathbf{Y})$ and the true posterior $p(\mathbf{Z}, \alpha|\mathbf{Y})$:

$$
\begin{aligned}
\mathrm{KL}[q(\mathbf{Z}, \alpha|\mathbf{Y})||p(\mathbf{Z}, \alpha|\mathbf{Y})] &= -\int q(\mathbf{Z}, \alpha|\mathbf{Y}) \log \frac{p(\mathbf{Z}, \alpha|\mathbf{Y})}{q(\mathbf{Z}, \alpha|\mathbf{Y})} \, d(\mathbf{Z}, \alpha) \\
&= \int q(\mathbf{Z}, \alpha|\mathbf{Y}) \log \frac{p(\mathbf{Y}, \mathbf{Z}, \alpha)}{p(\mathbf{Y}) q(\mathbf{Z}, \alpha|\mathbf{Y})} \, d(\mathbf{Z}, \alpha) \\
&= \int q(\mathbf{Z}, \alpha|\mathbf{Y}) \log \frac{p(\mathbf{Y}, \mathbf{Z}, \alpha)}{q(\mathbf{Z}, \alpha|\mathbf{Y})} \, d(\mathbf{Z}, \alpha) \\
&\quad - \int q(\mathbf{Z}, \alpha|\mathbf{Y}) \log p(\mathbf{Y}) \, d(\mathbf{Z}, \alpha) \\
&= \int q(\mathbf{Z}, \alpha|\mathbf{Y}) \log \frac{p(\mathbf{Y}, \mathbf{Z}, \alpha)}{q(\mathbf{Z}, \alpha|\mathbf{Y})} \, d(\mathbf{Z}, \alpha) - \log p(\mathbf{Y}).
\end{aligned}
$$

The last line follows as $\log p(\mathbf{Y})$ is neither a function of $\mathbf{Z}$ and α. From this equation, it is evident that minimizing $\mathrm{KL}[q(\mathbf{Z}, \alpha|\mathbf{Y})||p(\mathbf{Z}, \alpha|\mathbf{Y})]$ corresponds to maximizing the following lower bound:

$$\log p(\mathbf{Y}) \geq \int q(\mathbf{Z}, \alpha|\mathbf{Y}) \log \frac{p(\mathbf{Y}, \mathbf{Z}, \alpha)}{q(\mathbf{Z}, \alpha|\mathbf{Y})} \, d(\mathbf{Z}, \alpha).$$

The Kullback–Leibler divergence between the variational posterior and the true posterior for the LSM can be written as

$$
\mathrm{KL}[q(\mathbf{Z}, \alpha | \mathbf{Y}) || p(\mathbf{Z}, \alpha | \mathbf{Y})] = \mathrm{KL}[q(\alpha) || p(\alpha)] + \sum_{i=1}^{N} \mathrm{KL}[q(\mathbf{z}_i) || p(\mathbf{z}_i)] - \mathbb{E}_{q(\mathbf{Z}, \alpha | \mathbf{Y})}[\log(p(\mathbf{Y} | \mathbf{Z}, \alpha))
$$

$$
= \frac{1}{2} \left(\frac{\tilde{\psi}^2}{\psi^2} - \log \frac{\tilde{\psi}^2}{\psi^2} + \frac{(\tilde{\xi} - \xi)^2}{\psi^2} + ND \log(\sigma^2) - N \log(\det(\tilde{\mathbf{\Sigma}})) \right)
$$

$$
+ \frac{N}{2\sigma^2} \mathrm{tr}(\tilde{\mathbf{\Sigma}}) + \frac{\sum_{i=1}^{N} \tilde{\mathbf{z}}_i^T \tilde{\mathbf{z}}_i}{2\sigma^2} - \mathbb{E}_{q(\mathbf{Z}, \alpha | \mathbf{Y})}[\log(p(\mathbf{Y} | \mathbf{Z}, \alpha))] - \frac{1 + ND}{2},
$$

where the expected log-likelihood $\mathbb{E}_{q(\mathbf{Z}, \alpha | \mathbf{Y})}[\log(p(\mathbf{Y} | \mathbf{Z}, \alpha))]$ is approximated using the Jensen's inequality:

$$
\mathbb{E}_{q(\mathbf{Z}, \alpha | \mathbf{Y})}[\log(p(\mathbf{Y} | \mathbf{Z}, \alpha))] = \sum_{i \neq j}^{N} y_{ij} \mathbb{E}_{q(\mathbf{Z}, \alpha | \mathbf{Y})}[\alpha - |\mathbf{z}_i - \mathbf{z}_j|^2] - \mathbb{E}_{q(\mathbf{Z}, \alpha | \mathbf{Y})}
$$

$$
[\log(1 + \exp(\alpha - |\mathbf{z}_i - \mathbf{z}_j|^2))]
$$

$$
\leq \sum_{i \neq j}^{N} y_{ij} (\mathbb{E}_{q(\mathbf{Z}, \alpha | \mathbf{Y})}[\alpha - |\mathbf{z}_i - \mathbf{z}_j|^2]) - \log(1 + \mathbb{E}_{q(\mathbf{Z}, \alpha | \mathbf{Y})}
$$

$$
[\exp(\alpha - |\mathbf{z}_i - \mathbf{z}_j|^2)])
$$

$$
= \sum_{i \neq j}^{N} y_{ij} (\tilde{\xi} - 2\mathrm{tr}(\tilde{\mathbf{\Sigma}}) - |\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j|^2)
$$

$$
- \log \left(1 + \frac{\exp\left(\tilde{\xi} + \frac{1}{2}\tilde{\psi}^2\right)}{\det(\mathbf{I} + 4\tilde{\mathbf{\Sigma}})^{\frac{1}{2}}} \exp\left(-(\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j)^T \right.
$$

$$
\left. (\mathbf{I} + 4\tilde{\mathbf{\Sigma}})^{-1}(\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j)) \right). \tag{1}
$$

From this equation, it is possible to understand the computational advantage of the squared Euclidean distance model with respect to the Euclidean distance model. In fact, the expected log-likelihood has been approximated using the Jensen's inequality whereas Salter-Townshend and Murphy (2013) needed to use three first-order Taylor-expansions to fit the model with the Euclidean distance. An alternative approach to approximate the expected log-likelihood is given by the Jaakola and Jordan bound (Jaakkola and Jordan 2000), but it would require further approximations, and it would be more difficult to compute.

To estimate the model, an EM algorithm (Dempster, Laird, and Rubin 1977) can be applied. The EM algorithm consists of two main steps: the first step, called the E-step, aims to estimate the parameters $\tilde{\mathbf{z}}_i$, $\tilde{\mathbf{\Sigma}}$ of the posterior distribution of the latent space positions by maximizing the complete data log-likelihood given all the other parameters Θ. The second step is the M-step where Θ is updated maximizing the complete data log-likelihood given $\tilde{\mathbf{z}}_i$ and $\tilde{\mathbf{\Sigma}}$. As observed above, in this context maximizing the log-likelihood corresponds to minimize the Kullback–Leibler divergence between the variational posterior and the true posterior. Therefore, the EM algorithm can be written as a function of the

Kullback–Leibler divergence, this approach is commonly known as Variational EM algorithm (Jordan et al. 1999). This method scales as $\mathcal{O}(N^2)$, but it converges in just a few iterations, and the calculations performed in the estimation procedure are pretty simple (see the supplementary material for a comparison of CPU times with other methods and models).

The analytical form of the parameter estimates will be found introducing the first- and second-order Taylor series expansion approximation of the following function:

$$f(\tilde{\mathbf{z}}_i, \tilde{\boldsymbol{\Sigma}}, \tilde{\xi}, \tilde{\psi}^2) = \log\left(1 + \frac{\exp\left(\tilde{\xi} + \frac{1}{2}\tilde{\psi}^2\right)}{\det(\mathbf{I} + 4\tilde{\boldsymbol{\Sigma}})^{\frac{1}{2}}} \exp\left(-(\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j)^T(\mathbf{I} + 4\tilde{\boldsymbol{\Sigma}})^{-1}(\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j)\right)\right) \quad (2)$$

calculated around the estimates calculated at the previous step of the algorithm (see supplementary material).

Here, we outline the Variational EM algorithm on the $(i+1)$th iteration:

E-Step: Estimate the parameters of the latent posterior distributions $\tilde{\mathbf{z}}_i^{(i+1)}$ and $\tilde{\boldsymbol{\Sigma}}^{(i+1)}$ evaluating:

$$\mathcal{Q}(\Theta; \Theta^{(i)}) = -\text{KL}(q(\mathbf{Z}, \alpha|\mathbf{Y})||p(\mathbf{Z}, \alpha|\mathbf{Y}))$$
$$= \int q(\mathbf{Z}, \alpha|\mathbf{Y}) \log \frac{p(\mathbf{Z}, \alpha|\mathbf{Y})}{q(\mathbf{Z}, \alpha|\mathbf{Y})} \, d(\mathbf{Z}, \alpha),$$

where $\Theta = (\tilde{\xi}, \tilde{\psi}^2)$. This gives

$$\tilde{\boldsymbol{\Sigma}}^{(i+1)} = \frac{N}{2}\left[\left(\frac{N}{2\sigma^2} + 2\sum_{i=1}^{N}\sum_{j\neq i} y_{ij}\right)\mathbf{I} + J(\tilde{\boldsymbol{\Sigma}}^{(i)})\right]^{-1},$$

where J is the Jacobian matrix of $f(\tilde{\mathbf{z}}_i^{(i)}, \tilde{\boldsymbol{\Sigma}}^{(i)}, \tilde{\xi}^{(i)}, \tilde{\psi}^{2(i)})$ (Equation (2)) evaluated at $\tilde{\boldsymbol{\Sigma}} = \tilde{\boldsymbol{\Sigma}}^{(i)}$. And

$$\tilde{\mathbf{z}}_i^{(i+1)} = \left[\left(\frac{1}{2\sigma^2} + \sum_{j\neq i}(y_{ji} + y_{ij})\right)\mathbf{I} + H\left(\tilde{\mathbf{z}}_i^{(i)}\right)\right]^{-1}$$
$$\left[\sum_{j\neq i}(y_{ji} + y_{ij})\tilde{\mathbf{z}}_j - G(\tilde{\mathbf{z}}_i^{(i)}) + H\left(\tilde{\mathbf{z}}_i^{(i)}\right)\tilde{\mathbf{z}}_i^{(i)}\right], \quad (3)$$

where G is the gradient and H is the Hessian matrix of $f(\tilde{\mathbf{z}}_i^{(i)}, \tilde{\boldsymbol{\Sigma}}^{(i+1)}, \tilde{\xi}^{(i)}, \tilde{\psi}^{2(i)})$ evaluated at $\tilde{\mathbf{z}}_i = \tilde{\mathbf{z}}_i^{(i)}$.

M-Step: Estimate the parameters of the posterior distribution of α evaluating:

$$\Theta^{(i+1)} = \text{argmax } \mathcal{Q}(\Theta; \Theta^{(i)}).$$

This gives

$$\tilde{\xi}^{(i+1)} = \frac{\xi + \psi^2\left(\sum_{i=1}^{N}\sum_{j\neq i} y_{ij} - f'\left(\tilde{\xi}^{(i)}\right) + \tilde{\xi}^{(i)} f''\left(\tilde{\xi}^{(i)}\right)\right)}{1 + \psi^2 f''(\tilde{\xi}^{(i)})},$$
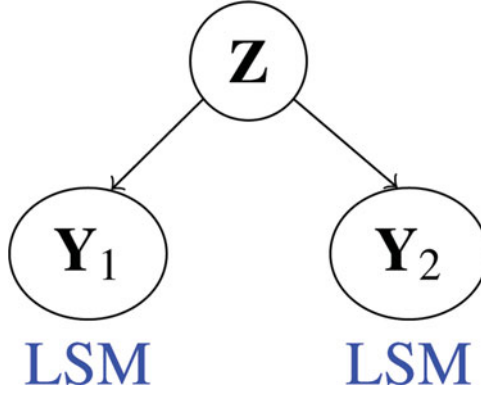
Figure 2.   Latent space joint model for two network views.

where f' and f'' are the first and the second derivatives of $f(\tilde{\mathbf{z}}_i^{(i+1)}, \tilde{\boldsymbol{\Sigma}}^{(i+1)}, \tilde{\xi}^{(i)}, \tilde{\psi}^{2(i)})$ evaluated at $\tilde{\xi} = \tilde{\xi}^{(i)}$, and

$$\tilde{\psi}^{2(i+1)} = \left(\frac{1}{\psi^2} + 2f'(\tilde{\psi}^{2(i)}) \right)^{-1},$$

where f' is the first derivative of $f(\tilde{\mathbf{z}}_i^{(i+1)}, \tilde{\boldsymbol{\Sigma}}^{(i+1)}, \tilde{\xi}^{(i+1)}, \tilde{\psi}^{2(i)})$ evaluated at $\tilde{\psi}^2 = \tilde{\psi}^{(i)2}$.

3. LATENT SPACE JOINT MODEL

Let us suppose to have K network views on the same set of N nodes. We introduce a model that assumes that a continuous latent variable $\mathbf{z}_i \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_D)$ is able to summarize the information given by all the network views $\mathbf{Y}_1, \ldots, \mathbf{Y}_K$ identifying the position of node i in a D-dimensional latent space. In this case, the network data assume conditionally independence given the latent variable. Our purpose is to model each network view by using an LSM (see Figure 2).

This yields to the following joint model:

$$p(\mathbf{Y}_1, \ldots, \mathbf{Y}_K | \mathbf{Z}, \alpha_1, \ldots, \alpha_K) = \prod_{k=1}^{K} p(\mathbf{Y}_k | \mathbf{Z}; \alpha_k)$$

$$= \prod_{k=1}^{K} \prod_{i \neq j}^{N} \frac{\exp(\alpha_k - |\mathbf{z}_i - \mathbf{z}_j|^2)^{y_{ijk}}}{1 + \exp(\alpha_k - |\mathbf{z}_i - \mathbf{z}_j|^2)},$$

where $p(\alpha_k) = \mathcal{N}(\xi_k, \psi_k^2)$, $p(\mathbf{z}_i) \stackrel{\text{iid}}{=} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_D)$ with $\sigma^2, \xi_k, \psi_k^2$ set to be fixed parameters, and the dyad y_{ijk} takes value 1 if there is a link between node i and node j in network k, and 0 otherwise.

The following identity allows one to find the model parameters $\alpha_1, \ldots, \alpha_K$ and the posterior distribution of the latent variable $\mathbf{z}_i$ given the K models:

$$p(\mathbf{z}_i | \mathbf{Y}_1, \ldots, \mathbf{Y}_K; \alpha_1, \ldots, \alpha_K) \propto p(\mathbf{z}_i) \prod_{k=1}^{K} p(\mathbf{Y}_k | \mathbf{z}_i; \alpha_k)$$

$$\propto p(\mathbf{z}_i) \prod_{k=1}^{K} \frac{p(\mathbf{z}_i | \mathbf{Y}_k; \alpha_k)}{p(\mathbf{z}_i)}$$

$$= \frac{\prod_{k=1}^{K} p(\mathbf{z}_i | \mathbf{Y}_k; \alpha_k)}{p(\mathbf{z}_i)^{K-1}}. \tag{4}$$

Applying the variational inference approach presented in Section 2.1 introducing the variational parameters $\Theta_k = (\tilde{\xi}_k, \tilde{\psi}_k^2)$, $\tilde{\mathbf{z}}_{ik}$, and $\tilde{\boldsymbol{\Sigma}}_k$, it is possible to approximate $p(\mathbf{z}_i | \mathbf{Y}_k; \alpha_k)$ with $q(\mathbf{z}_i) \sim \mathcal{N}(\tilde{\mathbf{z}}_{ik}, \tilde{\boldsymbol{\Sigma}}_k)$. Recalling that $\mathbf{z}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_D)$ and Equation (3), we obtain that the posterior distribution of the latent variables given all the network views can be written as

$$p(\mathbf{z}_i | \mathbf{Y}_1, \ldots, \mathbf{Y}_K; \Theta_1, \ldots, \Theta_K) \propto \mathcal{N}(\bar{\mathbf{z}}_i, \bar{\boldsymbol{\Sigma}}),$$

where the parameters are

$$\bar{\boldsymbol{\Sigma}} = \left[\sum_{k=1}^{K} \tilde{\boldsymbol{\Sigma}}_k^{-1} - \frac{K-1}{\sigma^2} \mathbf{I}_D \right]^{-1} \qquad \text{and} \qquad \bar{\mathbf{z}}_i = \bar{\boldsymbol{\Sigma}} \left[\sum_{k=1}^{K} \tilde{\boldsymbol{\Sigma}}_k^{-1} \tilde{\mathbf{z}}_{ik} \right]. \tag{5}$$

By fitting LSJM, we get information on $p(\mathbf{z}_i | \mathbf{Y}_1, \ldots, \mathbf{Y}_K; \Theta_1, \ldots, \Theta_K)$ and $p(\mathbf{z}_i | \mathbf{Y}_k; \Theta_k)$ for $k = 1, \ldots, K$. This way it is possible to have estimates for both the overall positions $\bar{\mathbf{z}}_i$ and the position given one particular network view $\tilde{\mathbf{z}}_{ik}$.

The estimates of $\tilde{\mathbf{z}}_{ik}$ and $\tilde{\boldsymbol{\Sigma}}_k$ are updated from $\bar{\mathbf{z}}_i$ and $\bar{\boldsymbol{\Sigma}}$, so we can locate the unconnected nodes or subgraphs in the latent space depending on their position conditional on the other network views, avoiding the usual tendency of pushing away the unconnected nodes to maximize the likelihood when fitting the classical LSM. This approach also allows one to plot the positions given each network view in the same latent space, and to look at how the nodes in each network change the positions.

Let $(\Theta_1^{(i)}, \ldots, \Theta_K^{(i)})$ be the current estimates of $(\Theta_1, \ldots, \Theta_K)$ and initialize $(\Theta_1^{(0)}, \ldots, \Theta_K^{(0)})$. The variational EM algorithm at the iteration $(i + 1)$ can be summarized as follows:

E-Step: Estimate the parameters $\bar{\boldsymbol{\Sigma}}^{(i+1)}$ and $\bar{\mathbf{z}}_i^{(i+1)}$ of the posterior distribution of the latent variables given all the network views evaluating:

$$\mathcal{Q}(\Theta_1, \ldots, \Theta_K; \Theta_1^{(i)}, \ldots, \Theta_K^{(i)})$$

$$= \mathbb{E}_{p(\mathbf{Z}|\mathbf{Y}_1, \ldots, \mathbf{Y}_K; \Theta_1^{(i)}, \ldots, \Theta_K^{(i)})}[\log(p(\mathbf{Y}_1, \ldots, \mathbf{Y}_K, \mathbf{Z}|\Theta_1, \ldots, \Theta_K))]$$

$$= \sum_{k=1}^{K} \mathbb{E}_{p(\mathbf{Z}|\mathbf{Y}_1, \ldots, \mathbf{Y}_K; \Theta_1^{(i)}, \ldots, \Theta_K^{(i)})}[\log(p(\mathbf{Y}_k, \mathbf{Z}|\Theta_k))]$$

$$- (K-1)\mathbb{E}_{p(\mathbf{Z}|\mathbf{Y}_1, \ldots, \mathbf{Y}_K; \Theta_1^{(i)}, \ldots, \Theta_K^{(i)})}[\log(p(\mathbf{Z}))]. \tag{6}$$

Thus, we can estimate the parameters of the posterior distribution $p(\mathbf{z}_i | \mathbf{Y}_k; \Theta_k)$ given each network separately:

$$\tilde{\boldsymbol{\Sigma}}_k^{(i+1)} = \frac{N}{2} \left[\left(\frac{N}{2\sigma^2} + 2 \sum_{i=1}^{N} \sum_{j \neq i} y_{ijk} \right) \mathbf{I} + J_k(\bar{\boldsymbol{\Sigma}}^{(i)}) \right]^{-1},$$

where J_k is the Jacobian matrix of $f(\bar{\mathbf{z}}_i^{(i)}, \bar{\boldsymbol{\Sigma}}^{(i)}, \tilde{\xi}_k^{(i)}, \tilde{\psi}_k^{2(i)})$ (Equation (2)) evaluated at $\tilde{\boldsymbol{\Sigma}} = \bar{\boldsymbol{\Sigma}}^{(i)}$, and,

$$\tilde{\mathbf{z}}_{ik}^{(i+1)} = \left[\left(\frac{1}{2\sigma^2} + \sum_{j \neq i} (y_{jik} + y_{ijk}) \right) \mathbf{I} + H_k \left(\bar{\mathbf{z}}_i^{(i)} \right) \right]^{-1}$$

$$\left[\sum_{j \neq i} (y_{jik} + y_{ijk}) \bar{\mathbf{z}}_j^{(i)} - G_k \left(\bar{\mathbf{z}}_i^{(i)} \right) + H_k \left(\bar{\mathbf{z}}_i^{(i)} \right) \bar{\mathbf{z}}_i^{(i)} \right], \qquad (7)$$

where G_k and H_k are, respectively, the gradient and the Hessian matrices of $f(\bar{\mathbf{z}}_i^{(i)}, \bar{\boldsymbol{\Sigma}}^{(i+1)}, \tilde{\xi}_k^{(i)}, \tilde{\psi}_k^{2(i)})$ evaluated at $\tilde{\mathbf{z}}_i = \bar{\mathbf{z}}_i^{(i)}$.

The posterior distribution of the latent positions given all the network views is estimated merging the estimates of the single models:

$$p(\mathbf{z}_i | \mathbf{Y}_1, \ldots, \mathbf{Y}_K; \Theta_1, \ldots, \Theta_K) \propto \mathcal{N}(\bar{\mathbf{z}}_i, \bar{\boldsymbol{\Sigma}}),$$

where

$$\bar{\boldsymbol{\Sigma}}^{(i+1)} = \left[\sum_{k=1}^{K} \left[\tilde{\boldsymbol{\Sigma}}_k^{(i+1)} \right]^{-1} - \frac{K-1}{\sigma^2} \mathbf{I}_D \right]^{-1}, \text{ and } \bar{\mathbf{z}}_i^{(i+1)} = \bar{\boldsymbol{\Sigma}}^{(i+1)} \left[\sum_{k=1}^{K} \left[\tilde{\boldsymbol{\Sigma}}_k^{(i+1)} \right]^{-1} \tilde{\mathbf{z}}_{ik}^{(i+1)} \right].$$

M-Step: Update the model parameters evaluating

$$\left(\Theta_1^{(i+1)}, \ldots, \Theta_K^{(i+1)} \right) = \text{argmax } \mathcal{Q} \left(\Theta_1, \ldots, \Theta_K; \Theta_1^{(i)}, \ldots, \Theta_K^{(i)} \right).$$

This gives

$$\tilde{\xi}_k^{(i+1)} = \frac{\xi_k + \psi_k^2 \left(\sum_{i=1}^{N} \sum_{j \neq i} y_{ijk} - f_k' \left(\tilde{\xi}_k^{(i)} \right) + \tilde{\xi}_k^{(i)} f_k'' \left(\tilde{\xi}_k^{(i)} \right) \right)}{1 + \psi_k^2 f_k'' \left(\tilde{\xi}_k^{(i)} \right)},$$

where f_k' and f_k'' are the first and the second derivatives of $f(\bar{\mathbf{z}}_i^{(i+1)}, \bar{\boldsymbol{\Sigma}}^{(i+1)}, \tilde{\xi}_k^{(i)}, \tilde{\psi}_k^{2(i)})$ evaluated at $\tilde{\xi} = \tilde{\xi}_k^{(i)}$, and

$$\tilde{\psi}_k^{2(i+1)} = \left(\frac{1}{\psi_k^2} + 2 f_k'(\tilde{\psi}_k^{(i)2}) \right)^{-1},$$

where f_k' is the first derivative of $f(\bar{\mathbf{z}}_i^{(i+1)}, \bar{\boldsymbol{\Sigma}}^{(i+1)}, \tilde{\xi}_k^{(i+1)}, \tilde{\psi}_k^{2(i)})$ evaluated at $\tilde{\psi}^2 = \tilde{\psi}_k^{(i)2}$.

4. MISSING LINK DATA

Recent Bayesian approaches to predict missing links in network data have been introduced by Hoff (2009) who proposed to use multiplicative latent factor models, and Koskinen, Robins, and Pattison (2010) and Koskinen et al. (2013) in the context of Bayesian exponential random graph models. However, there is lack of methods able to take into account the information given by the presence of multiple network views.

Missing (unobserved) links can be easily managed by the LSJM using the information given by all the network views. To estimate the probability of the presence ($y_{ijk} = 1$) or

absence ($y_{ijk} = 0$) of an edge, we employ the posterior mean of the α_k and of the latent positions so that we get the following equation:

$$y_{ijk}^* = p(y_{ijk} = 1|\bar{\mathbf{z}}_i, \bar{\mathbf{z}}_j, \tilde{\xi}_k) = \frac{\exp\left(\tilde{\xi}_k - |\bar{\mathbf{z}}_i - \bar{\mathbf{z}}_j|^2\right)}{1 + \exp\left(\tilde{\xi}_k - |\bar{\mathbf{z}}_i - \bar{\mathbf{z}}_j|^2\right)}.$$

If we want to infer whether to assign $y_{ijk} = 1$ or $y_{ijk} = 0$, we need to introduce a threshold τ_k, and let $y_{ijk} = 1$ if $p(y_{ijk} = 1|\tilde{\mathbf{z}}_{ik}, \tilde{\mathbf{z}}_{jk}, \tilde{\xi}_k) > \tau_k$, and $y_{ijk} = 0$ otherwise. We set the threshold to be equal to the median probability of a link for the subset of the actual observed links in network k.

To evaluate link prediction in presence of missing links, we use a 10-fold cross-validation procedure consisting of randomly splitting the set of all the possible dyads in each network view into 10 subsets. We can predict the links of each subset given the others fitting an LSM to each network independently and then fitting the LSJM. Finally, we can compare the link prediction performance given by these two methods.

The LSJM allows one to locate in the latent space a missing node (no information about links sent and received by the node) in one network by employing the information provided by the other network views. We evaluate the link prediction approach for the missing nodes applying a 10-fold cross-validation on the nodes, randomly dividing the set of nodes in each network into 10 subsets, and then predicting the links using the LSJM. We do not use a single LSM to locate missing nodes in the latent space and to estimate their probabilities of links since the only information that the model would use estimating of the latent positions would be the prior distribution of the nodes $p(\mathbf{z}_i)$.

To facilitate the interpretation of the results, we matched the rotation of the latent positions in the single LSM with the ones obtained from the LSJM.

5. APPLICATIONS

5.1 COMPUTATIONAL ASPECTS

The LSM and the LSJM have been fitted assuming that $p(\alpha) = \mathcal{N}(0, 2)$ and $p(\mathbf{z}_i) \stackrel{\text{iid}}{=} \mathcal{N}(\mathbf{0}, \mathbf{I}_2)$, initializing the variational parameters $\tilde{\xi}_k = 0$ and $\tilde{\psi}_k^2 = 2$, and latent positions $\bar{\mathbf{z}}_i$ by random generated numbers from $\mathcal{N}(\mathbf{0}, \mathbf{I}_2)$ and setting $\bar{\boldsymbol{\Sigma}} = \mathbf{I}_2$.

We have set the latent space to be bi-dimensional to be able to visualize and easily interpret the results. Ten random starts of the algorithm were used and the solution returning the maximum expected likelihood value was selected. The latent positions are identifiable up to a rotation of the latent space. For this reason, to speed up the convergence of the algorithm we matched the estimates of $\tilde{\mathbf{z}}_{ik} \ \forall k > 1$ with $\tilde{\mathbf{z}}_{i1}$ via singular value decomposition in the first 10 iterations of the EM algorithm.

The EM algorithm was stopped after at least 10 iterations when

$$\|\mathbb{E}_q[\log(p(\mathbf{Y}|\mathbf{Z}, \alpha))]^{(i+1)} - \mathbb{E}_q[\log(p(\mathbf{Y}|\mathbf{Z}, \alpha))]^{(i)}\| < \text{tol},$$

where $\mathbb{E}_q[\log(p(\mathbf{Y}|\mathbf{Z}, \alpha))]$ is given by Equation (1) if we fit the LSM or by Equation (6) if we fit the LSJM, i indicates the iteration, tol is a desired tolerance value (which, in this case, was set tol $= 10^{-2}$).
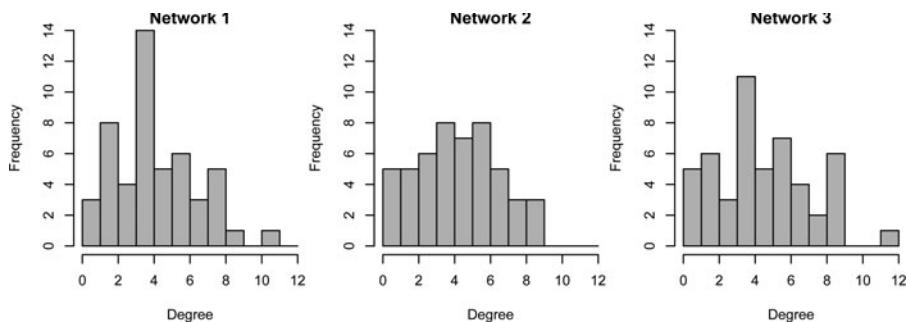
Figure 3.   Degree distribution for the girls networks.

To assess the fit of the model, we evaluated the in-sample predictions producing the receiver operating characteristic (ROC) curve of the estimated link probabilities and calculating the area under the curve (AUC) and the boxplots of the estimated link probabilities for both the true positive and the true negative links. We estimated the probability of a link under each network view by calculating:

$$p(y_{ijk} = 1 | \tilde{\mathbf{z}}_{ik}, \tilde{\mathbf{z}}_{jk}, \tilde{\xi}_k) = \frac{\exp\left(\tilde{\xi}_k - |\tilde{\mathbf{z}}_{ik} - \tilde{\mathbf{z}}_{jk}|^2\right)}{1 + \exp\left(\tilde{\xi}_k - |\tilde{\mathbf{z}}_{ik} - \tilde{\mathbf{z}}_{jk}|^2\right)}.$$

All the calculations have been done using the R package lvm4net.

5.2   EXCERPT OF 50 GIRLS FROM "TEENAGE FRIENDS AND LIFESTYLE STUDY"

Pearson and Michell (2000) and Pearson and West (2003) collected data for a "Teenage Friends and Lifestyle Study." The dataset contains three directed networks about friendship relations between students in a school in Glasgow, Scotland. Each student was asked to name up to six best friends in the cohort. The data come from three yearly waves, from 1995 to 1997. An extended description of all the data in the study can be found in Pearson and Michell (2000). In this article, we will focus on an excerpt of 50 girls that were present at all three measurement points. This dataset is available at the SIENA software website (*http://www.stats.ox.ac.uk/snijders/siena/*, Ripley, Snijders, and Preciado 2011). The networks are formed of 113, 116, and 122 links, respectively, and have density of 0.046, 0.047, 0.049. Their degree distributions are shown in Figure 3.

5.2.1 LSM.   In this section, we applied the LSM to the excerpt of 50 girls from the "Teenage Friends and Lifestyle Study" data.

We fitted an LSM to each network separately assuming that $p(\alpha) = \mathcal{N}(0, 2)$ and $p(\mathbf{z}_i) \overset{\text{iid}}{=} \mathcal{N}(\mathbf{0}, \mathbf{I}_2)$. The estimated latent posterior positions $\tilde{\mathbf{z}}_i$ (defined in Equation (3)) are shown in Figure 4. A matching rotation of the final estimates was applied to facilitate the interpretation of the results. The posterior distributions for the α parameters are quite similar over time $q(\alpha_1) = \mathcal{N}(-0.63, 0.01)$, $q(\alpha_2) = \mathcal{N}(-0.66, 0.01)$, and $q(\alpha_3) = \mathcal{N}(-0.48, 0.01)$; this means that there are not big changes in terms of network density over time.
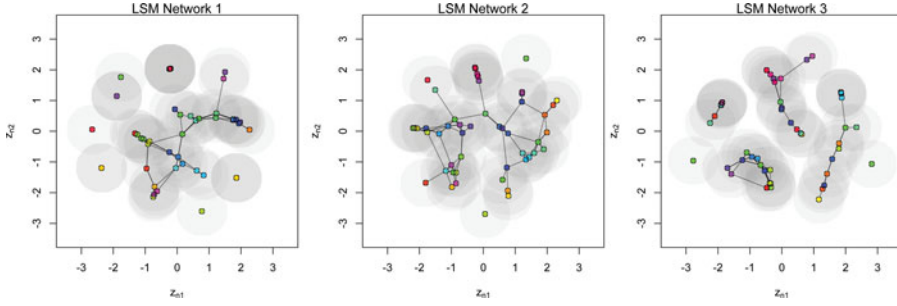
Figure 4. Latent positions $\tilde{\mathbf{z}}_i$ for the girls' networks fitting the LSM. The gray ellipses represent the 95% approximate credible intervals. Overlapping approximate credible intervals make darker shades of gray.

From the analysis of the ROC curve (Figure 5 left) and AUC of the in-sample estimated link probabilities, it seems clear that the proposed LSM fits quite well the data separately. The boxplots in Figure 5 (right) show that the estimated probabilities distinguish quite well the true negatives with low probability of forming a link from the true positives with high probability of forming a link.

5.2.2 LSJM. In this section, we fitted the LSJM to the Girls' dataset. The overall positions $\bar{\mathbf{z}}_i$ (defined in Equation (5)) in the latent space are displayed in Figure 6. In Figure 6 (right), the dots represent the overall positions $\bar{\mathbf{z}}_i$ and the arrows connect the estimated positions $\tilde{\mathbf{z}}_{ik}$ under each model k (defined in Equation (7)) from network $k = 1$ to $k = 3$, so that it is possible to see their variation over time.

Figure 7 shows the estimated positions $\tilde{\mathbf{z}}_{ik}$ under each model in the latent space. These plots allow a direct comparison between the positions in the latent space $\tilde{\mathbf{z}}_{ik}$ given by the LSJM and the ones obtained by fitting a single LSM for each network view (Figure 4). It is interesting to observe that in the LSJM context, the proximity between the disconnected components in network 3 depends on their previous relations. The posterior probabilities of the model parameters $q(\alpha_1) = \mathcal{N}(-0.42, 0.01)$, $q(\alpha_2) = \mathcal{N}(-0.39, 0.01)$, and $q(\alpha_3) = \mathcal{N}(-0.32, 0.01)$ are lower than the single LSM approaches (Section 5.2), implying that a given distance d between two nodes in the latent space corresponds to a higher probability of a link.
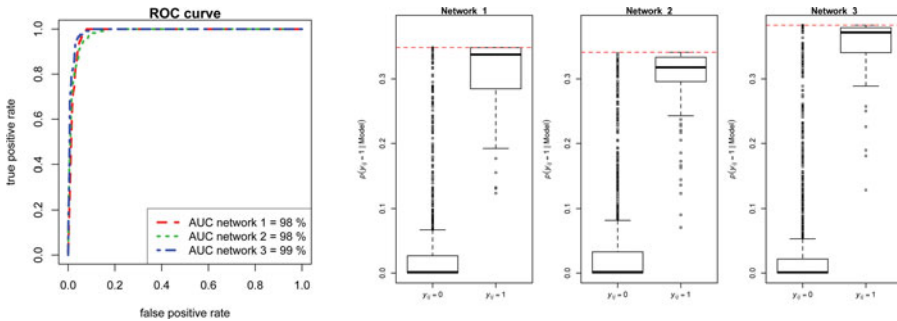

Figure 5. ROC curves (left) and boxplots (right) of the estimated probabilities of a link for the true negatives and true positives obtained by fitting the LSM to the three girls' networks.
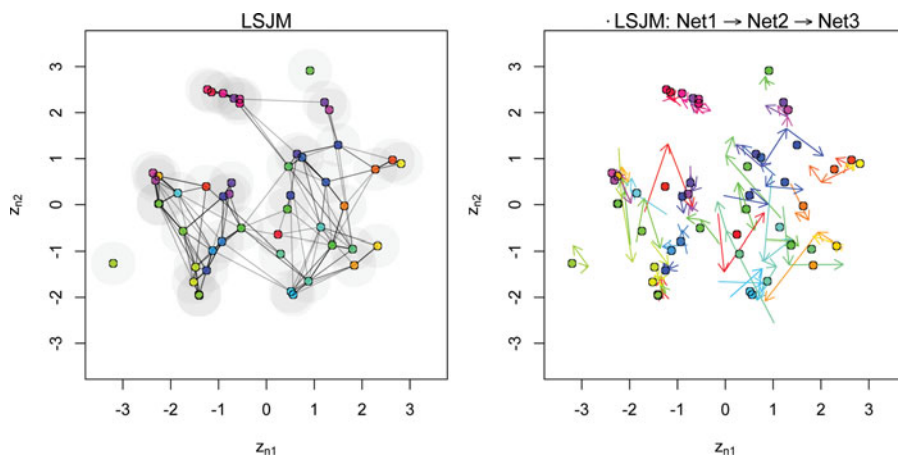
Figure 6. On the left are shown the latent positions $\bar{\mathbf{z}}_i$ for the girls' networks fitting the LSJM. In the plot on the right the dots represent the overall positions $\bar{\mathbf{z}}_i$ and the arrows connect the estimated position under each model $\tilde{\mathbf{z}}_{i1}, \tilde{\mathbf{z}}_{i2}, \tilde{\mathbf{z}}_{i3}$.

From the ROC curves, AUC, and the boxplots (Figure 8) of the estimated probabilities of a link, it appears clear that the LSJM fits the data quite well.

To evaluate the link prediction, we applied a 10-fold cross-validation setting 10% of the links to be missing at each time point. The areas under the ROC curves are 97%, 96%, 99% fitting the LSJM and 89%, 97%, 98% fitting the LSM, it shows that the estimates of the links are quite good, especially when applying the LSJM.

Setting $p(y_{ijk} = 1 | \tilde{\mathbf{z}}_{ik}, \tilde{\mathbf{z}}_{jk}, \tilde{\xi}_k) > \tau_k$, where τ_k is equal to the median probability of a link for the subgroup of the actual observed links in network k and applying the LSJM, we obtained a misclassification rate of 4% for every network, whereas applying three single LSM we obtain a misclassification rate of 4% for network 1, and 5% for networks 2 and 3.

The estimated networks are formed of 109, 115, and 118 links, respectively, and have density of 0.044, 0.047, 0.048. Their degree distributions are shown in Figure 9. Comparing these results with the true network statistics (Section 5.2 and Figure 3), it is possible to observe that the distributions are quite similar.
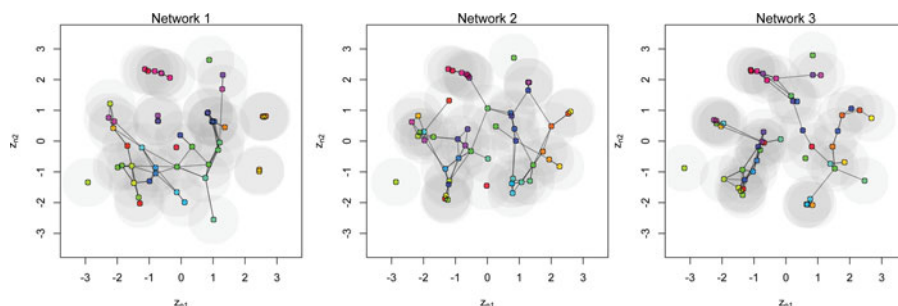

Figure 7. Estimated latent positions under each model $\tilde{\mathbf{z}}_{i1}, \tilde{\mathbf{z}}_{i2}, \tilde{\mathbf{z}}_{i3}$ for the girls' networks fitting the LSJM. The gray ellipses represent the 95% approximate credible intervals.
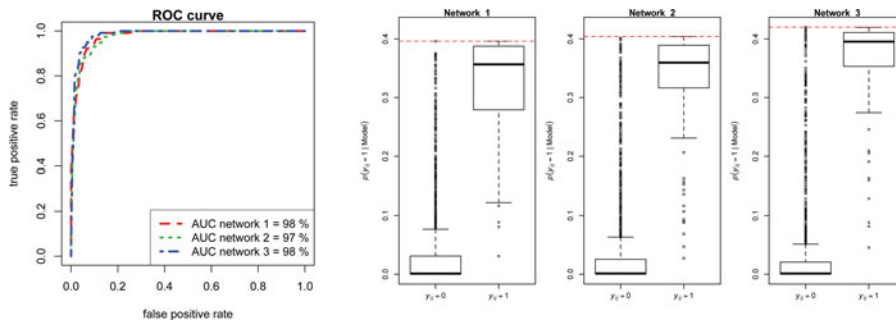
Figure 8.   ROC curves and boxplots of the estimated probabilities of a link for the true negatives and true positives obtained by fitting the LSJM to the girls' networks.

The LSJM allows one to also manage missing nodes, to do this we applied a 10-fold cross-validation setting the 10% of the nodes in each network to be missing. We obtained a misclassification rate of 9% for all the three networks. As mentioned above in this case, the LSM approach would be useless since it would locate the nodes only relying on the prior information.

5.3   SACCHAROMYCES CEREVISIAE PROTEIN–PROTEIN INTERACTIONS

We analyze a dataset containing two undirected networks formed by genetic and physical protein–protein interactions (PPI) between 67 *Saccharomyces cerevisiae* proteins. The genetic interactions network is formed of 294 links, and its density is 0.066, while the physical interactions network is formed of 190 links, and its density is 0.043. Their degree distributions are shown in Figure 10. The complex relational structure of this dataset has led to implementation of models aiming at describing the functional relationships between the observations (Troyanskaya et al. 2003; Bandyopadhyay et al. 2008). A list of proteins included in this dataset is displayed in the supplementary material. The dataset is available in the lvm4net package, and was downloaded from the Biological General Repository for Interaction Datasets (BioGRID) database (Stark et al. 2006, *http://thebiogrid.org/*). We refer to Stark et al. (2006, 2011) for a description of BioGRID, and for details regarding how the data were collected.
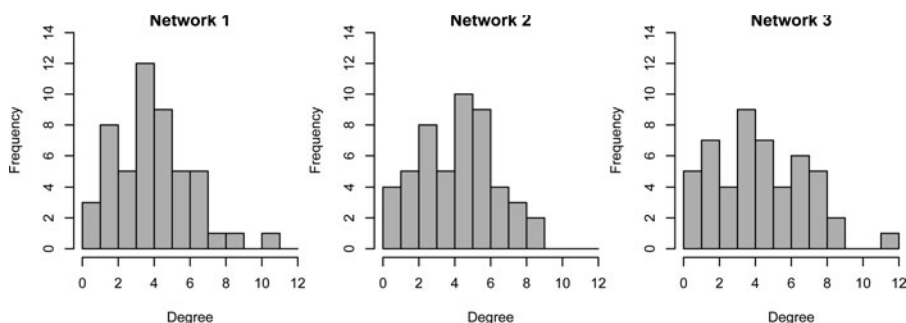

Figure 9.   Estimated degree distribution for the girls' networks fitting the LSJM.
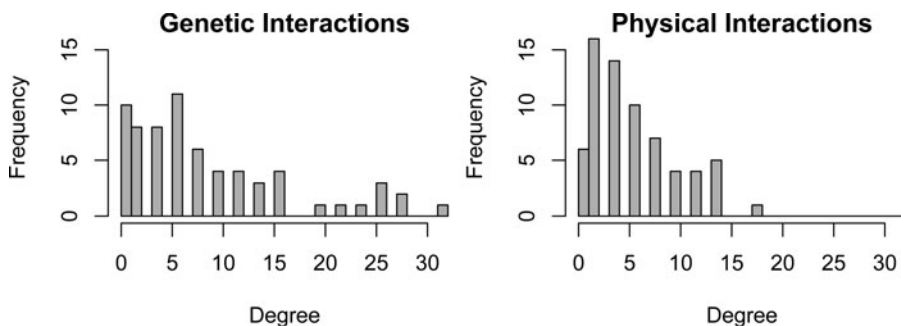
Figure 10. Degree distribution for the PPI dataset.

5.3.1 LSM. We fit the LSM to the PPI data working with the genetic and physical interaction networks separately. We assume that $p(\alpha) = \mathcal{N}(0, 2)$ and $p(\mathbf{z}_i) \overset{\text{iid}}{=} \mathcal{N}(\mathbf{0}, \mathbf{I}_2)$. The estimated latent positions $\tilde{\mathbf{z}}_i$ (defined in Equation (3)) are shown in Figure 11. The posterior distributions of the α's $q(\alpha_1) = \mathcal{N}(-0.332, 0.003)$ and $q(\alpha_2) = \mathcal{N}(-1.001, 0.005)$ indicate that network 1 regarding the genetic interactions is much more dense than network 2, which refers to the physical interactions. The ROC curves, AUC, and the boxplots (Figure 12) show that the proposed LSM fit the data quite well.

5.3.2 LSJM. We apply the LSJM to the PPI dataset. Figure 13 shows the estimated overall latent positions $\bar{\mathbf{z}}_i$ (defined in Equation (5)) and in the plot on the right each arrow starts from the positions $\tilde{\mathbf{z}}_{i1}$ for the genetic interaction dataset and points to the latent positions $\tilde{\mathbf{z}}_{i2}$ for the physical interaction dataset (defined in Equation (7)). Figure 14 shows the estimated position under each model $\tilde{\mathbf{z}}_{ik}$ in the latent space. It is possible to compare these results with Figure 11 in which we have fitted two single LSM to the data.

The ROC curve, AUC, and the boxplots in Figure 15 show that the LSJM fits the data quite well. The posterior distributions for the α's are $q(\alpha_1) = \mathcal{N}(-0.410, 0.003)$ $q(\alpha_2) = \mathcal{N}(-0.940, 0.005)$.
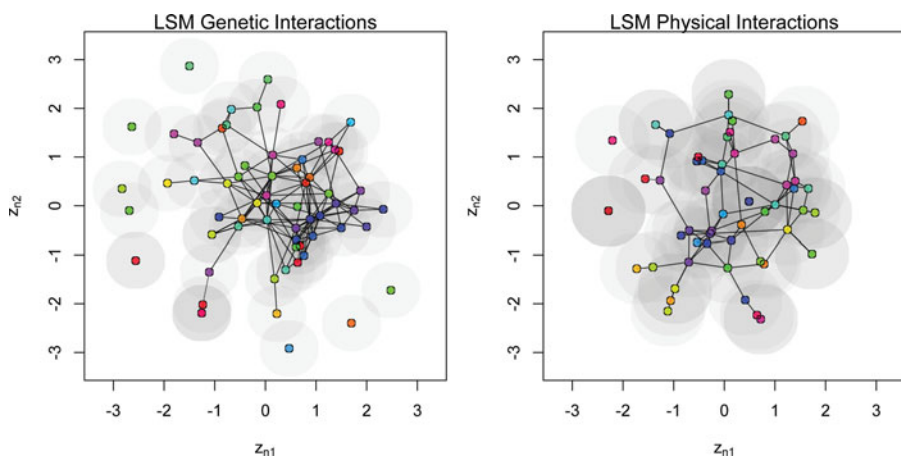
Figure 11. Latent positions $\tilde{\mathbf{z}}_i$ for the genetic (left) and physical (right) interaction data fitting the LSM.
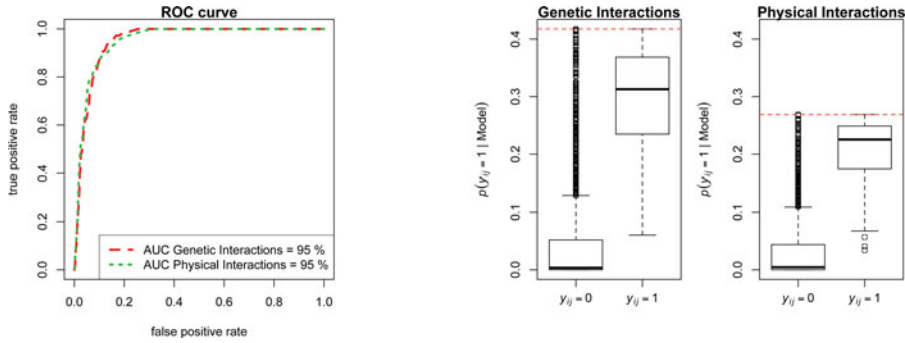
Figure 12. ROC curves and boxplots of the estimated probabilities of a link for the true negatives and true positives obtained by fitting the LSM to the PPI data.

We applied a 10-fold cross-validation to evaluate the prediction of missing links. The area under the ROC curve is 94% for both the genetic and the physical interaction networks fitting the LSJM, fitting the LSM the AUC values are 66% for the genetic interaction network and 97% for the physical interaction network. The results of the LSJM show a much better fit in terms of estimates of missing links for the genetic interaction network compared to the ones obtained by the single LSM approach. Setting the threshold to be equal to the median probability of a link for the subgroup of the actual observed links, we obtained similar misclassification rates for the missing links in both LSJM and single LSM approach: 9% for the genetic interaction network, and 6% for the physical interaction network.

The estimated genetic interactions network is formed of 297 links, and its density is 0.067, while the estimated physical interactions network is formed of 185 links, and its density is 0.042. Their degree distributions are shown in Figure 16. Comparing these results
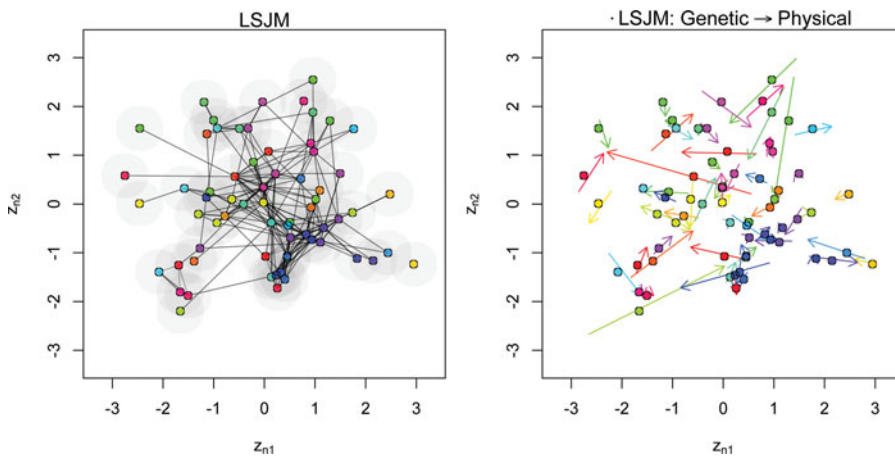

Figure 13. On the left are shown the latent positions $\bar{z}_i$ for the PPI networks fitting the LSJM. In the plot on the right, the dots represent the overall positions $\bar{z}_i$ and the arrows connect the estimated position under each model $\tilde{z}_{i1}, \tilde{z}_{i2}$.
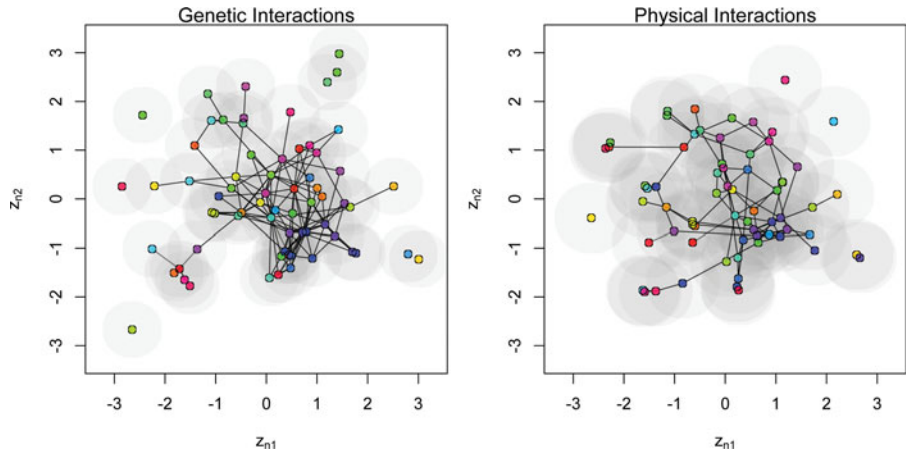
Figure 14.   Estimated latent positions under each model $\tilde{\mathbf{z}}_{i1}$, $\tilde{\mathbf{z}}_{i2}$ for the PPI networks fitting the LSJM.
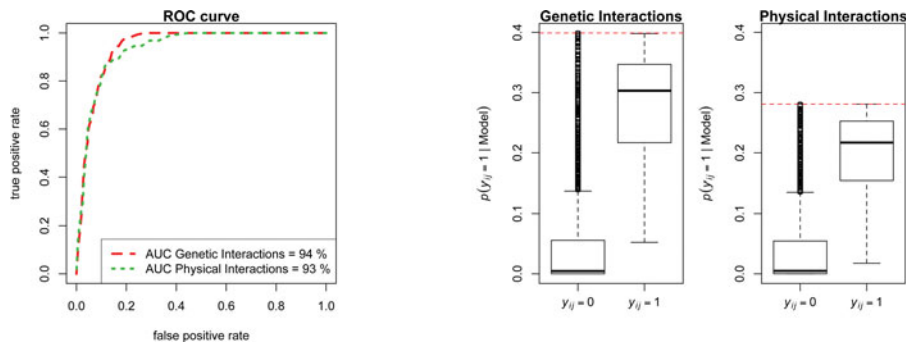

Figure 15.   ROC curves of the estimated probabilities of a link for the true negatives and true positives obtained by fitting the LSJM to the PPI Networks.
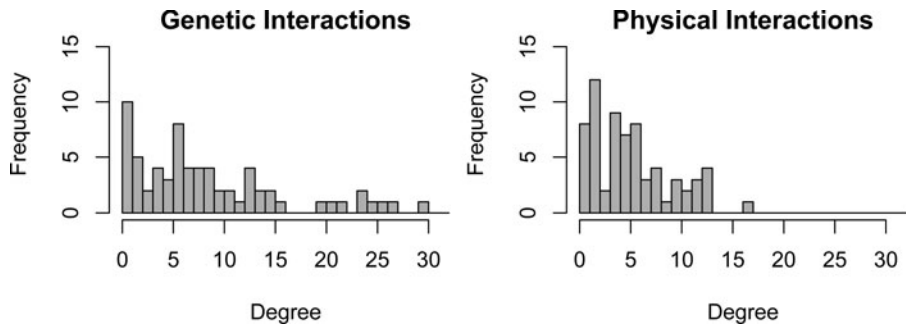

Figure 16.   Estimated degree distribution for the PPI networks fitting the LSJM.

with the true network statistics (Section 5.3), it is possible to observe that the simulated degree distributions (Figure 16) are smoother than the original one (Figure 10), but overall they are quite similar.

Applying a 10-fold cross-validation for missing nodes using the LSJM, we obtained a misclassification rate of 24% for the genetic interactions dataset and 20% for the physical interaction network.

6. CONCLUSIONS

A lot of network data require the introduction of novel models able to describe their complex connectivity structure. On the other hand, new inferential methods are needed to carry out estimation efficiently. In this article, we proposed a latent variable model (LSJM) for multiple network views, which extends the latent space model proposed by Hoff, Raftery, and Handcock (2002), allowing the information given by different relations on the same nodes to be summarized in the same latent space. The use of the variational approach to compute the model parameters allows us to apply the latent space joint model to larger networks (of the order of thousands of nodes). A comparison between the variational method and MCMC for the single view network latent space model is given in Salter-Townshend and Murphy (2013). An alternative variational algorithm that can be used in this context could be derived from the methods outlined in Opper and Archambeau (2009). This model allows the position of each observation in a latent space to be found based on all the available information in the datasets. Further information like the latent positions in each network view separately is obtained updating the single-network estimates given the overall positions. This information allows effective visualization and prediction of the data. The examples presented show how the LSJM facilitates the interpretation of the different positions of the network nodes in the latent space according to longitudinal measurements in the excerpt of 50 girls from "Teenage Friends and Lifestyle Study" example and multiple relations in the *Saccharomyces cerevisiae* genetic and physical protein–protein interactions dataset. All the methods presented in this article are included in the `lvm4net` package for R (Gollini 2014).

Future work may lead to an extension of the model allowing cluster formation by assuming that the latent positions come from a Gaussian mixture model fitting each network using a Latent Position Cluster Model (Handcock, Raftery, and Tantrum 2007). Or a novel modeling approach that takes explicitly into account the sequential feature as in the dynamic network analysis (Sarkar and Moore 2005; Hoff 2011; Westveld and Hoff 2011).

SUPPLEMENTARY MATERIALS

In Section 1 we perform an empirical analysis in order to compare the performance and prediction properties of the squared Euclidean distance model to the nonsquared Euclidean model. We use the following R packages: `latentnet` (Krivitsky and Handcock 2008), `VBLPCM` (Salter-Townshend and Murphy 2013) and `lvm4net` (Gollini 2014).

In Section 2 there is a description of the PPI data analyzed in Section 5.3 of the main paper. It also includes the list of the proteins.

Section 3 provides details on the estimation procedures for the LSM parameters.

ACKNOWLEDGMENTS

REFERENCES

Airoldi, E. M., Blei, D. M., Fienberg, S. E., and Xing, E. P. (2008), "Mixed-Membership Stochastic Blockmodels," *Journal of Machine Learning Research*, 9, 1981–2014. [246,249]

Ansari, A., Koenigsberg, O., and Stahl, F. (2011), "Modeling Multiple Relationships in Social Networks," *Journal of Marketing Research*, 48, 713–728. [247]

Attias, H. (1999), "Inferring Parameters and Structure of Latent Variable Models by Variational Bayes," in *Proceedings of the Fifteenth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-99)*, San Francisco, CA: Morgan Kaufmann, pp. 21–30. [249]

Bandyopadhyay, S., Kelley, R., Krogan, N. J., and Ideker, T. (2008), "Functional Maps of Protein Complexes From Quantitative Genetic Interaction Data," *PLoS Computational Biology*, 4, 1–8. [247,259]

Beal, M. J. (2003), "Variational Algorithms for Approximate Bayesian Inference," Ph.D. dissertation, Gatsby Computational Neuroscience Unit, University College London. [249]

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood for Incomplete Data via the EM Algorithm" (with discussion), *Journal of the Royal Statistical Society,* Series B, 39, 1–38. [250]

Goldenberg, A., Zheng, A. X., Fienberg, S. E., and Airoldi, E. M. (2010), "A Survey of Statistical Network Models," *Foundations and Trends in Machine Learning*, 2, 129–233. [246]

Gollini, I. (2014), *lvm4net: Latent Variable Models for Networks*, R package Version 0.1. Available at *http://cran.r-project.org/web/packages/lvm4net/*. [263]

Handcock, M., Raftery, A., and Tantrum, J. (2007), "Model-Based Clustering for Social Networks," (with discussion), *Journal of the Royal Statistical Society,* Series A, 170, 1–22. [246,249,263]

Hoff, P. (2009), "Multiplicative Latent Factor Models for Description and Prediction of Social Networks," *Computational and Mathematical Organization Theory*, 15, 261–272. [254]

——— (2011), "Hierarchical Multilinear Models for Multiway Data," *Computational Statistics & Data Analysis*, 55, 530–543. [263]

Hoff, P., Raftery, A., and Handcock, M. (2002), "Latent Space Approaches to Social Network Analysis," *Journal of the American Statistical Association*, 97, 1090–1098. [246,247,248,263]

Holland, P. W., Laskey, K., and Leinhardt, S. (1983), "Stochastic Blockmodels: First Steps," *Social Networks*, 5, 109–137. [246]

Holland, P. W., and Leinhardt, S. (1981), "An Exponential Family of Probability Distributions for Directed Graphs," *Journal of the American Statistical Association*, 76, 33–50. [246]

Jaakkola, T. S., and Jordan, M. I. (2000), "Bayesian Parameter Estimation Via Variational Methods," *Statistics and Computing*, 10, 25–37. [250]

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., and Saul, L. K. (1999), "An Introduction to Variational Methods for Graphical Models," *Machine Learning*, 37, 183–233. [249,251]

Koskinen, J. H., Robins, G. L., and Pattison, P. E. (2010), "Analysing Exponential Random Graph (p-star) Models With Missing Data Using Bayesian Data Augmentation," *Statistical Methodology*, 7, 366–384. [254]

Koskinen, J. H., Robins, G. L., Wang, P., and Pattison, P. E. (2013), "Bayesian Analysis for Partially Observed Network Data, Missing Ties, Attributes and Actors," *Social Networks*, 35, 514–527. [254]

Mucha, P. J., Richardson, T., Macon, K., Porter, M. A., and Onnela, J.-P. (2010), "Community Structure in Time-Dependent, Multiscale, and Multiplex Networks," *Science*, 328, 876–878. [247]

Opper, M., and Archambeau, C. (2009), "The Variational Gaussian Approximation Revisited," *Neural Computation*, 21, 786–792. [263]

Pearson, M., and Michell, L. (2000), "Smoke Rings: Social Network Analysis of Friendship Groups, Smoking and Drug-Taking," *Drugs: Education, Prevention and Policy*, 7, 21–37. [247,256]

Pearson, M., and West, P. (2003), "Drifting Smoke Rings: Social Network Analysis and Markov Processes in a Longitudinal Study of Friendship Groups and Risk-Taking," *Connections*, 25, 59–76. [247,256]

R Core Team (2014), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing. [247]

Ripley, R. M., Snijders, T. A., and Preciado, P. (2011), *Manual for SIENA Version 4.0*, Oxford: University of Oxford, Department of Statistics, Nuffield College. [256]

Salter-Townshend, M., and Murphy, T. B. (2013), "Variational Bayesian Inference for the Latent Position Cluster Model for Network Data," *Computational Statistics & Data Analysis*, 57, 661–671. [249,250,263]

Salter-Townshend, M., White, A., Gollini, I., and Murphy, T. B. (2012), "Review of Statistical Network Analysis: Models, Algorithms, and Software," *Statistical Analysis and Data Mining*, 5, 243–264. [246]

Sarkar, P., and Moore, A. W. (2005), "Dynamic Social Network Analysis Using Latent Space Models," *SIGKDD Explorations*, 7, 31–40. [263]

Stark, C., Breitkreutz, B. J., Chatr-Aryamontri, A., Boucher, L., Oughtred, R., Livstone, M. S., Nixon, J., Van Auken, K., Wang, X., Shi, X., Reguly, T., Rust, J. M., Winter, A., Dolinski, K., and Tyers, M. (2011), "The BioGRID Interaction Database: 2011 Update," *Nucleic Acids Research*, 39, D698–D704. [259]

Stark, C., Breitkreutz, B. J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. (2006), "BioGRID: A General Repository for Interaction Datasets," *Nucleic Acids Research*, 34, D535–D539. [247,259]

Tang, L., Wang, X., and Liu, H. (2012), "Community Detection via Heterogeneous Interaction Analysis," *Data Mining and Knowledge Discovery*, 25, 1–33. [247]

Troyanskaya, O. G., Dolinski, K., Owen, A. B., Altman, R. B., and Botstein, D. (2003), "A Bayesian Framework for Combining Heterogeneous Data Sources for Gene Function Prediction (in Saccharomyces Cerevisiae)," *Proceedings of the National Academy of Sciences of the United States of America*, 100, 8348–8353. [259]

Westveld, A., and Hoff, P. (2011), "A Mixed Effects Model for Longitudinal Relational and Network Data, With Applications to International Trade and Conflict," *Annals of Applied Statistics*, 5, 843–872. [263]