Journal of Social Structure

JoSS Article: Volume 16

Sender- and receiver-specific blockmodels

Zhi Geng and
Krzysztof Nowicki
Lund University, Sweden
Krzysztof.Nowicki@stat.lu.se

Abstract

We propose a sender-specific blockmodel for network data which utilizes both the group membership and the identities of the vertices. This is accomplished by introducing the edge probabilities $(\theta_{i,v})$ for $1 \leq i \leq c, 1 \leq v \leq n$, where i specifies the group membership of a sending vertex and v specifies the identity of the receiving vertex. In addition, group membership is consider to be random, with parameters $(p_i)_{i=1}^{c}$. We present methods based on the EM algorithm for the parameter estimations and discuss the recovery of latent group memberships. A companion model, the receiver-specific blockmodel, is also introduced in which the edge probabilities $(\psi_{u,j})$ for $1 \leq u \leq n, 1 \leq j \leq c$ depend on the membership of a vertex receiving a directed edge. We apply both models to several sets of social network data.

Key words and phrases: Directed graph, Blockmodeling, Out-nets, In-nets, Ego-nets, EM algorithm, Multinomial distribution

1 Introduction and summary

Consider a directed graph (network) on the set $V = \{1, ..., n\}$ of n vertices. The pattern of directed edges can be described using the adjacency matrix $\mathbf{y} = (y_{uv})$ for $1 \leq u, v \leq n$, where $y_{uv} = 1$ if there is a (directed) edge from vertex u to vertex v and $y_{uv} = 0$ otherwise. (We do not allow loops and thus we set $y_{uu} \equiv 0, 1 \leq u \leq n$.) $\mathbf{y}$ can also be represented as $\mathbf{y} = (\mathbf{y_1}, ..., \mathbf{y_n})^T$ where $\mathbf{y_u} = (y_{u1}, ..., y_{un})$ provides a sequence of edge indicators of an out-net at vertex u.

We assume that each vertex belongs to one of c groups (blocks) and let a categorical variable $1 \leq x_u \leq c$ denote the group to which vertex u belongs. For convenience we refer to vertices belonging to the same group as having the same color. We let $n_i = \sum_{u=1}^{n} I(x_u = i)$ denote the number of vertices which have color i.

Knowledge of group membership can be used to model network data where edges can represent contacts between actors. These contacts can then be analyzed not only for the entire network, but also within the same block and between different blocks. The simplest approach is to consider colors to be deterministic and observed, an approach often referred to as blockmodeling.

Snijders and Nowicki (1997) extended this approach to random colors by letting $(X_u)_{u=1}^{n}$ be independent, identically distributed rv's, with $P(X_u = i) = p_i, 1 \leq i \leq c$. The parameters $\mathbf{p} = (p_i)_{i=1}^{c}$ can thus be interpreted as the expected fraction of vertices in group $1, ..., c$, respectively and the following obvious condition holds: $\sum_{i=1}^{c} p_i = 1$.

Nowicki and Snijders (2001) modeled a directed graph in the presence of random vertex coloring, by firstly assuming dyads (Y_{uv}, Y_{vu}), $1 \leq u < v \leq n$ to be independent conditionally on vertex colors

$$P(\mathbf{Y} = \mathbf{y} \mid \mathbf{X} = \mathbf{x}) = \prod_{1 \leq u < v \leq n} P((Y_{uv} = y_{uv}, Y_{vu} = y_{vu}) \mid \mathbf{X} = \mathbf{x}), \quad (1)$$

and, secondly, that

$$\begin{aligned} P((Y_{uv} = y_{uv}, &Y_{vu} = y_{vu}) \mid \mathbf{X} = \mathbf{x}) \\ &= P((Y_{uv} = y_{uv}, Y_{vu} = y_{vu}) \mid X_u = x_u, X_v = x_v) \quad (2) \\ &= \eta_{x_u, x_v}(y_{uv}, y_{vu}), \end{aligned}$$

where $\sum_{0 \leq s,t \leq 1} \eta_{i,j}(s,t) = 1$ for $1 \leq i, j \leq c$ and $\eta_{i,i}(0,1) = \eta_{i,i}(1,0)$ for $i = 1, ..., c$.

In the present paper we modify the conditional step of the block model above, where dyad probabilities depend only on the blocks to which both vertices are assigned. We instead allow edge probabilities also depend on the identities of the vertices. In the most general formulation it is possible to let the conditional dyad probabilities be $\eta_{u,v,i,j}(s,t)$ and the conditional edge probabilities be $\theta_{u,v,i,j}$. We specialize, however, by considering conditional independent edges with probabilities $\theta_{u,v,i,j}$ to be further specified as either $\theta_{i,v}$ or $\psi_{u,j}$ in the following two parametrizations:

- $\Theta = (\theta_{i,v})$ for $1 \le v \le n, 1 \le i \le c$ where $\theta_{i,v} = P(Y_{uv} = 1 \mid X_u = i)$,
- $\Psi = (\psi_{u,j})$ for $1 \le u \le n, 1 \le j \le c$ where $\psi_{u,j} = P(Y_{uv} = 1 \mid X_v = j)$.

Thus an edge y_{uv} from vertex u (sender) to vertex v (receiver) has a probability that depends on both the block of the sender and the identity of the receiver, or vice versa. The proposed parametrization, as given by Θ and Ψ, can not longer explicitly model recioprocity in the network since we assume that edges are independent.

Using the above parametrization Θ we can define the conditional step (2) as

$$\begin{aligned} P(\mathbf{Y} = \mathbf{y} \mid \mathbf{X} = \mathbf{x}) &= \prod_{1 \le u \ne v \le n} P(Y_{uv} = y_{uv} \mid X_u = x_u) \\ &= \prod_{1 \le u \ne v \le n} \theta_{x_u,v}^{y_{uv}} (1 - \theta_{x_u,v})^{1-y_{uv}}, \end{aligned} \tag{3}$$

thus obtaining

$$\begin{aligned} P(\mathbf{Y} = \mathbf{y}, \mathbf{X} = \mathbf{x}) &= P(\mathbf{Y} = \mathbf{y} \mid \mathbf{X} = \mathbf{x}) P(\mathbf{X} = \mathbf{x}) \\ &= \prod_{u=1}^{n} \left(p_{x_u} \prod_{v: v \ne u} \theta_{x_u,v}^{y_{uv}} (1 - \theta_{x_u,v})^{1-y_{uv}} \right). \end{aligned} \tag{4}$$

We will refer to (4) as the sender-specific blockmodel (the SB-model) in the sequel.

The parametrization proposed by $\Psi = (\psi_{u,i})$ for $1 \le u \le n$ and $1 \le i \le c$ leads to the replacement of (2) with

$$\begin{aligned} P(\mathbf{Y} = \mathbf{y} \mid \mathbf{X} = \mathbf{x}) &= \prod_{1 \le u \ne v \le n} P(Y_{uv} = y_{uv} \mid X_v = x_v) \\ &= \prod_{1 \le u \ne v \le n} \psi_{u,x_v}^{y_{uv}} (1 - \psi_{u,x_v})^{1-y_{uv}}. \end{aligned} \tag{5}$$

We will refer to model derived using (5) instead of (3) as the receiver-specific blockmodel (the RB-model) in the sequel.

While existing block models work with conditionally independent dyads, the block/identity combinations proposed here work either with independent out-nets or independent in-nets. For review of ego-nets in social networks analysis see Crossley et al. (2015). The SB-model has independent out-nets but dependent in-nets that are mixture of conditionally independent edges. The RB-model has independent in-nets but dependent out-nets that are a mixture of conditionally independent edges.

These two models make it possible to shift the focus of analysis from the individual attractiveness of the actors in the networks studied to the individual activeness of the actors. Although the choice between the two models should be governed by the properties that practitioners want to explore, technically they are equivalent and can be handled by similar statistical methods.

The parameterization specified as $\theta(x_u, x_v, u, v) \equiv \theta_{x_u,v}$ was originally proposed in Newman and Leicht (2007). However, they based their analysis on the following probability function

$$P(\mathbf{Y} = \mathbf{y} \mid \mathbf{X} = \mathbf{x}) = \prod_{1 \leq u \neq v \leq n} P(Y_{uv} = y_{uv} \mid X_u = x_u) = \prod_{1 \leq u \neq v \leq n} \theta_{x_u,v}^{y_{uv}}, \quad (6)$$

which is unfortunately erroneous; since for example, it assigns probability 1 to the empty digraph. By correcting their model we were able to take advantage of their approach.

In the analysis of network data carried out by Snijders and Nowicki (1997), Nowicki and Snijders (2001) and Daudin et al. (2008), it was assumed that group memberships are latent (unobserved) while the data only consisted of the observed network $\mathbf{y}$. This assumption precludes obtaining a closed form solution of the ML-estimators of the vertex parameters $\mathbf{p}$ and the edge parameters Θ.

One solution proposed for dealing with latent colors is based on the EM algorithm. Snijders and Nowicki (1997) proposed using this algorithm for an undirected graph (network) $\mathbf{y} = (y_{uv})$ for $1 \leq u, v \leq n$, where $y_{uv} = 1$ if there is an edge between vertex u and vertex $v \neq u$ and $y_{uv} = 0$, otherwise, and in which assumptions (1) and (2) (suitably modified for the undirected graph) give:

$$P(\mathbf{Y} = \mathbf{y} \mid \mathbf{X} = \mathbf{x}) = \prod_{1 \leq u < v \leq n} P(Y_{uv} = y_{uv} \mid X_u = x_u, X_v = x_v). \quad (7)$$

This algorithm was used iteratively first to color vertices and then to obtain estimators of parameters.

In a similar manner, we apply the EM algorithm iteratively to obtain the estimators for the parameters in (4) when it is assumed that the group memberships are latent. The following equations are employed:

$$q_{ui} = P(X_u = i \mid \mathbf{y}, \mathbf{p}, \Theta) = \frac{p_i \prod\limits_{v:v\neq u} \theta_{i,v}^{y_{uv}} (1-\theta_{i,v})^{1-y_{uv}}}{\sum_{j=1}^{c} \left[p_j \prod\limits_{v:v\neq u} \theta_{j,v}^{y_{uv}} (1-\theta_{j,v})^{1-y_{uv}}\right]}, \tag{8}$$

where q_{ui} denotes the probabilities of the group memberships given the parameters and the observed data. This allows us to replace the log likelihood function $L(\mathbf{p}, \Theta, \mathbf{x})$ with its expectation $E(L(\mathbf{p}, \Theta, \mathbf{x}) \mid \mathbf{y})$. Together with the condition $\sum_{i=1}^{c} p_i = 1$ employed in the evaluation process of the Lagrange multipliers, the following ML estimates of the parameters can thus be obtained:

$$\hat{p}_i = \frac{1}{n} \sum_{u=1}^{n} q_{ui}, \qquad \hat{\theta}_{i,v} = \frac{\sum\limits_{u:u\neq v} y_{uv} q_{ui}}{\sum\limits_{u:u\neq v} q_{ui}}. \tag{9}$$

Equations (8) and (9) define the EM algorithm and we iterate these equations until convergence using a stopping criterion. The resulting parameter estimates can then be used to recover group memberships.

The statistical estimation of the parameters in the RB-model parallels the estimation procedures for the SB-model, thus allowing us to consider both observed and unobserved (latent) group memberships.

The paper is organized as follows. Section 2 is devoted to the estimation of parameters in the SB-model, where the latent setting is in focus. Section 3 presents the analysis of several data sets and outlines methods for interpreting parameters and uncovering group composition. In Section 4 we consider the RB-model and compare it with the SB-model for one of the data sets.

2 Parameter estimation for the sender-specific blockmodel

In this section we propose a parameter estimation method for the SB-model. Our data consist of the directed graph $\mathbf{y} = (y_{uv})$ for $1 \leq u, v \leq n$ on the set of vertices $V = \{1, ..., n\}$ and the vector $\mathbf{x} = (x_u)_{u=1}^{n}$ where $1 \leq x_u \leq c$ specifies to which of c groups vertex u belongs. The probabilistic model is defined by letting letting $(X_u)_{u=1}^n$ be independent, identically distributed rv's, with $P(X_u = i) = p_i, 1 \leq i \leq c$. Assuming the conditional independence of directed edges given vertex colors and

$$P(Y_{uv} = 1 \mid X_u = i) = \theta_{i,v},$$

we get, according to (4),

$$P(\mathbf{Y} = \mathbf{y}, \mathbf{X} = \mathbf{x}) = \prod_{v=1}^{n} \prod_{i=1}^{c} p_i^{n_i} \theta_{i,v}^{s_{iv}} (1 - \theta_{i,v})^{n_{iv} - s_{iv}}, \tag{10}$$

where $n_i = \sum_{v=1}^n I(x_v = i)$, $s_{iv} = \sum_{u=1}^n I(x_u = i) y_{uv}$ and $n_{iv} = n_i - I(x_v = i)$.

It is not straightforward to test how well the SB-model fits the data. One possible way to do so with a given set of parameters $\mathbf{p}$ and Θ is to rely on the fact that the in-degrees $y_{\cdot v} = \sum_{u=1}^n y_{uv}$ are independently distributed for $v = 1, ..., n$. Since it holds that $y_{\cdot v} \sim \text{Bin}(n - 1, \sum_{i=1}^c \theta_{i,v} p_i)$, we have that

$$Q = \sum_{v=1}^{n} \frac{\Big(y_{\cdot v} - (n-1) \sum_{i=1}^c \theta_{i,v} p_i\Big)^2}{(n-1) \sum_{i=1}^c \theta_{i,v} p_i}$$

is asymptotically $\chi^2(n)$ distributed.

In order to estimate the model parameters, we work with the log likelihood:

$$L(\mathbf{p}, \Theta) = \ln P(\mathbf{y}, \mathbf{x}; \mathbf{p}, \Theta) =$$
$$\sum_{u=1}^{n} \sum_{v: v \neq u} (y_{uv} \ln \theta_{x_u,v} + (1 - y_{uv}) \ln(1 - \theta_{x_u,v})) + \sum_{i=1}^{c} n_i \ln p_i.$$

Straightforward calculations produce the following ML-estimates of the parameters:

$$\hat{p}_i = \frac{n_i}{n}, \qquad \hat{\theta}_{i,v} = \frac{s_{iv}}{n_{iv}}. \tag{11}$$

Proposition 2.1 *The mean and variance of the parameter estimates are:*

$$E(\hat{p}_i) = p_i, \qquad Var(\hat{p}_i) = \frac{p_i(1 - p_i)}{n}, \tag{12}$$

$$E(\hat{\theta}_{i,v}) = \theta_{i,v}, \qquad Var(\hat{\theta}_{i,v}) = (\theta_{i,v} - \theta_{i,v}^2) \left(\frac{(n-2)p_i + 1}{(n-1)^2 p_i^2} + O(\frac{1}{n^3}) \right). \tag{13}$$

Proof: (12) follows since $n_i \sim \text{Bin}(n, p_i)$. For the proof of (13) see Appendix A.

2.1 Parameter estimation in a latent setting

The estimation problem becomes much more intricate when it is assumed that group membership vector $\mathbf{x}$ is unobserved (latent) and our data consist only of the observed network $\mathbf{y}$. The log likelihood function is no longer only a function of $\mathbf{p}$ and Θ but instead $L = L(\mathbf{p}, \Theta, \mathbf{x})$. Consequently, it is not possible to obtain a closed form solution for the ML-estimators.

To address this problem we employ the EM algorithm which was introduced in Dempster et al. (1977). The application of the EM algorithm to mixture models for random graphs in the parametric context has been widely discussed, see, for example, Snijders and Nowicki (1997), Nowicki and Snijders (2001) and Daudin et al. (2008). This algorithm is based on cycling two consecutive steps: Step E and Step M.

In Step E we derive the distribution of $P(\mathbf{x} \mid \mathbf{y}; \mathbf{p}^{(m)}, \Theta^{(m)})$ with $\mathbf{p}^{(m)}$, $\Theta^{(m)}$ denoting estimators of $\mathbf{p}, \Theta$ being obtained after the mth cycle. This distribution is then used to "eliminate" $\mathbf{x}$ from $L = L(\mathbf{p}, \Theta, \mathbf{x})$ by calculating the expected value for the log likelihood $\bar{L} \equiv \bar{L}(\mathbf{p}, \Theta) = E\big(L(\mathbf{p}, \Theta, \mathbf{x}) \mid \mathbf{y}; \mathbf{p}^{(m)}, \Theta^{(m)}\big)$ by averaging over all possible $\mathbf{x}'s$.

In Step M we carry out the maximization of $\bar{L}$ by choosing

$$(\mathbf{p}^{(m+1)}, \Theta^{(m+1)}) = \underset{\mathbf{p},\Theta}{\text{Argmax}}\ \bar{L}(\mathbf{p}, \Theta),$$

thus obtaining $(\mathbf{p}^{(m+1)}, \Theta^{(m+1)})$, an update of parameter estimates after the $(m+1)$th cycle. More formally, with details of equations (14) and (15) given in Appendices B and C,

$$\begin{aligned}
\bar{L} = &\sum_{x_1=1}^{c} \cdots \sum_{x_n=1}^{c} P(\mathbf{x} \mid \mathbf{y}; \mathbf{p}^{(m)}, \Theta^{(m)}) \\
&\sum_{u=1}^{n} \left[\ln p_{x_u} + \sum_{v:v\neq u} (y_{uv} \ln \theta_{x_u,v} + (1 - y_{uv}) \ln(1 - \theta_{x_u,v})) \right] \qquad (14) \\
= &\sum_{u=1}^{n} \sum_{i=1}^{c} q_{ui}^{(m)} \Big[\ln p_i + \sum_{v:v\neq u} (y_{uv} \ln \theta_{i,v} + (1 - y_{uv}) \ln(1 - \theta_{i,v})) \Big],
\end{aligned}$$

with $q_{ui}^{(m)} = P(X_u = i \mid \mathbf{y}; \mathbf{p}^{(m)}, \Theta^{(m)})$ given by:

$$q_{ui}^{(m)} = \frac{p_i^{(m)} \prod_{v:v\neq u} (\theta_{i,v}^{(m)})^{y_{uv}} (1-\theta_{i,v}^{(m)})^{1-y_{uv}}}{\sum_{j=1}^{c} \left(p_j^{(m)} \prod_{v:v\neq u} (\theta_{j,v}^{(m)})^{y_{uv}} (1-\theta_{j,v}^{(m)})^{1-y_{uv}} \right)}, \tag{15}$$

Clearly $q_{ui}^{(m)}$ satisfies the normalizing condition $\sum_{i=1}^{c} q_{ui}^{(m)} = 1$.

We then use Lagrange multipliers with the following constraint on $\mathbf{p}$: $\sum_{i=1}^{c} p_i = 1$ and the following constraints on Θ,

$$0 \leq \theta_{i,v} \leq 1 \quad \text{for} \quad i = 1, ..., c \quad \text{and} \quad v = 1, ..., n.$$

The Lagrange function is given as

$$\begin{aligned} \Gamma &\equiv \Gamma(\mathbf{p}, \Theta, \Lambda, \Delta, \phi) \\ &= \sum_{u=1}^{n} \sum_{i=1}^{c} q_{ui}^{(m)} \left[\ln p_i + \sum_{v:v\neq u} \left(y_{uv} \ln \theta_{i,v} + (1-y_{uv}) \ln(1-\theta_{i,v}) \right) \right] \\ &\quad + \phi\,(1 - p_1 - \cdots - p_c) + \sum_{i=1}^{c} \sum_{v=1}^{n} \Big[\lambda_{i,v}(1-\theta_{i,v}) + \delta_{i,v}\theta_{i,v} \Big]. \end{aligned}$$

where $\Lambda = (\lambda_{i,v})$ for $1 \leq i \leq c, 1 \leq v \leq n$, $\Delta = (\delta_{i,v})$ for $1 \leq i \leq c, 1 \leq v \leq n$ and ϕ are the Lagrange multipliers. The solutions, which maximize $\Gamma \equiv \Gamma(\mathbf{p}, \Theta, \Lambda, \Delta, \phi)$, are given by the equation system

$$\frac{\partial \Gamma}{\partial p_i} = \frac{\sum_{u=1}^{n} q_{ui}^{(m)}}{p_i} - \phi = 0,$$

$$\frac{\partial \Gamma}{\partial \theta_{i,v}} = \sum_{u:u\neq v} q_{ui}^{(m)} \left(\frac{y_{uv}}{\theta_{i,v}} - \frac{1-y_{uv}}{1-\theta_{i,v}} \right) - \lambda_{i,v} + \delta_{i,v} = 0,$$

$$\frac{\partial \Gamma}{\partial \phi} = 1 - (p_1 + \cdots + p_c) = 0,$$

$$\frac{\partial \Gamma}{\partial \phi_{i,v}} = 1 - \theta_{i,v} \geq 0,$$

$$\frac{\partial\Gamma}{\partial\delta_{i,v}} = \theta_{i,v} \geq 0,$$
$$\lambda_{i,v} \geq 0, \qquad \delta_{i,v} \geq 0,$$
$$\lambda_{i,v}(1-\theta_{i,v}) = 0, \qquad \delta_{i,v}\theta_{i,v} = 0,$$

for $i = 1, ..., c$ and $v = 1, ..., n$.

By solving the above equations the following maximum likelihood estimates of the parameters are obtained:

$$\hat{p}_i^{(m+1)} = \frac{1}{n}\sum_{u=1}^{n} q_{ui}^{(m)}, \qquad \hat{\theta}_{i,v}^{(m+1)} = \frac{\sum_{u:u\neq v} y_{uv} q_{ui}^{(m)}}{\sum_{u\neq v} q_{ui}^{(m)}}, \tag{16}$$

for $m \geq 0$; $i = 1, ..., c$ and $v = 1, ..., n$.

By choosing a starting point $(p_1^{(0)}, ..., p_c^{(0)})$ and $(\theta_{i,v}^{(0)})$ for $1 \leq i \leq c; 1 \leq v \leq n$, and iterating equations (15) and (16) to convergence with the following stopping rules (with N denoting the number of rounds required for the algorithm to converge):

$$\mid \hat{\theta}_{i,v}^{(N)} - \hat{\theta}_{i,v}^{(N-1)} \mid < 0.01, \qquad \text{for} \;\; i = 1, ..., c; \;\; v = 1, ..., n,$$
$$\mid \hat{p}_i^{(N)} - \hat{p}_i^{(N-1)} \mid < 0.01, \qquad \text{for} \;\; i = 1, ..., c,$$

We then obtain parameter estimates $(\hat{p}_1, ..., \hat{p}_c)$, and $(\hat{\theta}_{i,v})$ for $1 \leq i \leq c; 1 \leq v \leq n$.

An alternative and possible better approach is the following: instead of using arbitrary or randomly generated colors or parameter estimates, one can use as starting points the parameter estimates obtained using (11), in which group memberships are obtained using the profile likelihood method, presented in Appendix D and then applying EM-algorithm.

2.2 The choice of the number of blocks

An important question when fitting the SB-model to data is the choice of the number of blocks. We choose to apply two slightly different model selection criteria for selecting the number of blocks.

One is the Bayesian information criterion (BIC), given in its general formulation below:

$$BIC(\mathcal{M}) = -2\ln P(\mathbf{y} \mid \hat{\Theta}, \hat{\mathbf{p}}) + V_{\mathcal{M}} \ln N,$$

where

1. N is the number of observations
2. $\mathbf{y}$ denotes the observed data (i.e., the adjacency matrix)
3. $V_{\mathcal{M}}$ is the number of free parameters of model $\mathcal{M}$
4. $\hat{\Theta}, \hat{\mathbf{p}}$ are the ML estimates of the parameters in the latent case

Thus for our model the criterion can be written as:

$$BIC(\mathcal{M}_c) = -2\sum_{u=1}^{n} \ln\left(\sum_{j=1}^{c} \hat{p}_j \prod_{v:v\neq u} \hat{\theta}_{j,v}^{y_{uv}} (1-\hat{\theta}_{j,v})^{1-y_{uv}}\right) + (cn+c-1)\ln n(n-1), \tag{17}$$

where $\mathcal{M}_c$ denotes the SB-model with c blocks. The model with the lowest BIC value is to be preferred.

The regularity conditions for BIC do not hold for our model since the estimates can be on the boundary of the parameter space. However, this criterion is still recommended in practice for the SB-model; see Picard (2007) and Daudin et al. (2008) for discussion of the model selection criteria in the context of mixture models.

The other closely related model selection criterion we apply is the Akaike information criterion (AIC), which is defined as:

$$AIC(\mathcal{M}) = -2\ln P(\mathbf{y} \mid \hat{\Theta}, \hat{\mathbf{p}}) + 2V_{\mathcal{M}}.$$

For the SB-model it will be written as:

$$AIC(\mathcal{M}_c) = -2\sum_{u=1}^{n} \ln\left(\sum_{j=1}^{c} \hat{p}_j \prod_{v:v\neq u} \hat{\theta}_{j,v}^{y_{uv}} (1-\hat{\theta}_{j,v})^{1-y_{uv}}\right) + 2(cn+c-1). \tag{18}$$

The formula above is then used to select the c value corresponding to the lowest AIC value, thus giving the number of blocks.

Both BIC and AIC are partially based on the logarithm of the likelihood function, and both introduce a penalty term for the number of free parameters in the model (in the SB-model, it is $cn + c - 1$). However, the penalty term for BIC is larger than for AIC; namely, BIC penalizes the number of free parameters more strongly than AIC does. Since an increase of the number of blocks c will lead to considerable increase of the number of free parameters in the SB-model using BIC maybe a hindrance when considering models with more than 2-blocks.

3 Applications of the SB-model to social networks

In this section we apply our model to two data sets previously studied in the literature: Hansell's classroom data and Krackhardt's high-tech network data. The parametrization of the model allows us to uncover several structural characteristics which previous analyses were unable to address. The numerical calculations required by the EM-algorithm were obtained using the R programming language.

3.1 Hansell's friendship network

Hansell (1984) collected and analyzed data on friendship relations among 27 students, 13 boys and 14 girls, in a sixth-grade class in Baltimore, USA. Every student was asked to rate the strength of his/hers friendship as either "like very much","like somewhat" or "dislike". In this paper a directed friendship relation was defined as when a student likes another one "very much".

These relationships are represented by a 27×27-dimensional adjacency matrix, where "1" represents a (directed) friendship and "0" otherwise. While this matrix is observed, the friendship pattern is assumed to be related to latent group memberships, which we want to recover based on the observed friendship structures.

Both the 2-block model and the 3-block model were fitted to the data. Applying the model selection criteria proposed in Section 2.2 and inserting the parameter estimates we have

	$c = 2$	$c = 3$
$BIC(\mathcal{M}_c)$	1024	1063
$AIC(\mathcal{M}_c)$	773	685

which indicates that the 3-block model is preferred by AIC, while the 2-block model is preferred by BIC. We choose to present both models and provide a comparison of the results.

3.1.1 The 2-block model

We first explore the group structure of the classmates. Using the EM-algorithm we obtain estimates of the relative sizes of Groups 1 resp 2:

$$\hat{\mathbf{p}} = (0.5169, 0.4831).$$

The color recovery $(\hat{x}_1, ..., \hat{x}_{27})$ of the network's vertices is based on the analysis of $q_{ui}, i = 1, 2$, the conditional probabilities of the group membership given the observed network:

u	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
q_{u1}	0	.99	0	0	0	0	.99	0	.99	1	1	1	1	0	.98	0	1	1	0	0	1	0	0	.99	0	.99	1
q_{u2}	1	.01	1	1	1	1	.01	1	.01	0	0	0	0	1	.02	1	0	0	1	1	0	1	1	.01	1	.01	0

To recover the group membership we use a threshold value of 0.98 by letting

$$\hat{x}_u = 1 \quad \text{if} \quad q_{u1} \geq 0.98,$$

and

$$\hat{x}_u = 2 \quad \text{if} \quad q_{u2} \geq 0.98,$$

which gives the following partition:

Group 1: 2,7,9,10,11,12,13,15,17,18,21,24,26,27,
Group 2: 1,3,4,5,6,8,14,16,19,20,22,23,25.

The students are clearly classified since no ambiguous members are found (that is, members for which $0.02 < q_{u1}, q_{u2} < 0.98$), so the EM-algorithm reveals a clear-cut block structure. Fig. 1 visualizes the Hansell's network data with the 2-block model.

We also observe that groups are balanced with respect to gender:

No. of students	Male	Female
Group 1	7	7
Group 2	6	7

This would indicate that gender does not play an essential role in the obtained block partition.

It is interesting to compare the above group partition with that obtained when the standard stochastic blockmodel which assumes a 2-block structure is fitted to the data. Using the program STOCNET (see Boer et al. (2003)) the following partition was obtained:

Group 1: 2,9,10,11,12,13,**14**,15,17,18,21,24,**25**,26,27,
Group 2: 1,3,4,5,6,**7**,8,16,19,20,22,23.

This alternative partition coincides to a large extent with that obtained using the SB-model 2-block structure, even though three vertices, 7, 14 and 25, have changed their recovered group memberships.

Next we study the structure of the friendship pattern. We first classify this pattern with respect to the recovered group memberships by providing the following density table:

Density of directed contacts	Group 1	Group 2
Group 1	0.0816	0.0824
Group 2	0.4066	0.3077

This table reveals that Group 2 mainly consists of students seeking contacts, both within their own group and among students in Group 1. Students from Group 1 are considerably less prone to looking for friendship contacts, both within their own group and in the other group. The table above can be compared with the density table below, where the classification is based on gender:

Density of directed contacts	Male	Female
Male	0.2663	0.1374
Female	0.1044	0.3469

We observe that in the partition obtained by gender, friendship contacts exist mainly within each of the gender groups. Thus the partition obtained by friendship patterns differs considerably from that obtained by gender.

Next, the EM-algorithm gives the following estimators of the parameter vector Θ:

v	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$\hat{\theta}_{1,v}$	0.072	0	0	0.072	0.071	0.142	0.072	0	0.071	0.072	0	0	0.142	0.213
$\hat{\theta}_{2,v}$	0.460	0.460	0.230	0.153	0.461	0.385	0.153	0.153	0.384	0.460	0.307	0.230	0.308	0.462

j	15	16	17	18	19	20	21	22	23	24	25	26	27
$\hat{\theta}_{1,v}$	0.215	0.072	0.143	0.146	0.072	0	0.213	0.142	0	0	0.213	0.072	0
$\hat{\theta}_{2,v}$	0.690	0.307	0.690	0.615	0.153	0	0.539	0.308	0.383	0.690	0.539	0	0.153

The estimated parameters $(\hat{\theta}_{1,v})$ for $v = 1, ..., 27$ and $(\hat{\theta}_{2,v})$ for $v = 1, ..., 27$ can be interpreted as each student's ability to attract friendship contacts from students in Group 1 and Group 2, respectively. Using these parameters, it is found that student No.15 is the most contact-attracting of students from Group 1 since

$$\hat{\theta}_{1,15} = \max\{\{\hat{\theta}_{1,v}\}_{v=1,\ldots,27}\} = 0.215.$$

Moreover, students No. 15, 17 and 24 appear to be the most contact-attracting ones for students from Group 2 since

$$\hat{\theta}_{2,15} = \hat{\theta}_{2,17} = \hat{\theta}_{2,24} = \max\{\{\hat{\theta}_{2,v}\}_{v=1,\dots,27}\} = 0.690.$$

All three of these contact-attracting students are female, with student No.15 being the most attractive in the entire student population.

Furthermore, observing that $\hat{\theta}_{2,v} > \hat{\theta}_{1,v}, v \in \{1, 2, ..., 27\} \setminus \{20, 26\}$, we can conclude that the members of Group 2 are more likely to be active in creating friendship contacts than the members of Group 1 are, which we have already concluded, at the aggregate level, by looking at friendship densities. Thus we can regard this as a group characteristic, telling us that people in Group 2 tend to search friendships more actively than do people in Group 1.

3.1.2 The 3-block model

We now fit the SB-model with $c = 3$ groups. The EM-algorithm results in a partition of the 27 actors into 3 groups where the estimated relative sizes of Groups 1, 2 and 3 are, respectively:

$$\hat{\mathbf{p}} = (0.2223, 0.4075, 0.3702),$$

and the recovered group memberships are:

Group 1: 1,3,4,7,8,20,
Group 2: 5,6,14,15,16,19,21,22,23,24,25,
Group 3: 2,9,10,11,12,13,17,18,26,27.

Fig.3 visualizes the Hansell's network data with the 3-block model.

The density table for the 3-block partition is:

Density of directed contacts	Group 1	Group 2	Group 3
Group 1	0.2778	0.2272	0.3667
Group 2	0.0758	0.5124	0.2636
Group 3	0.0500	0.0545	0.0500

We notice that the members of Group 3 rarely seek friendship with other students; however, they receive considerably more contacts from the other two groups then they themselves seek. Thus we can assume that a 3-block partition reveals that Group 3 members are less active in showing a liking for other students.

Comparing this 3-block partition with the previous 2-block structure, we find that all the members of Group 3 belong to Group 1 in the 2-block model. Therefore Group 3 may be regarded as a split from Group 1 in the 2-block model.

To further understand the new group partition, we look at the gender composition:

No. of students	Male	Female
Group 1	5	1
Group 2	2	9
Group 3	6	4

Group 1 consists mainly of male students, and Group 2 of female students, while Group 3 is mixed. This feature was not revealed by the 2-block model, where Groups 1 and 2 had an almost equal number of male and female students. So we conclude that while in the 3-block model there exist relationships between the group partition and the attributes (the gender) of the individual actors, such relationship can not be found in the 2-block model. In this aspect, we think that the 3-block model shows new properties of the network compared to properties obtained when applying the 2-block model.

3.2 Krackhardt's organizational structure

Krackhardt (1987) collected and analyzed cognitive social structure data from a high-tech machine manufacturing company in the USA. Relationships among 21 employees of the company's management team were studied to assess interpersonal dynamics in the organization.

Data were collected by asking each staff member to indicate to whom they typically went for advice and help. Each person described his (or her) advice relationships, as well the relations he (or she) perceived to exist among all other managers. One question was "To whom would you go to seek advice at work?". The data thus consist of twenty-one 21×21-dimensional adjacency matrices of ratings $A1, ..., A21$, one for each of the 21 staff members.

Here we only analyze adjacency matrix $\mathbf{y}$ which we compiled from the 21 original matrices by only including each person's own advice relationship, i.e., $y_{uv} = 1(0)$, if person u will turn (not turn) to person v for advice.

In addition to the adjacency matrix, Krackhardt considered 4 attribute variables, of which we use two, **Age** and **Tenure** (years of employment). In our modeling we make use of these attributes to describe the properties of a recovered group partition.

The data were first modeled using both a 2- and a 3-block model. When fitting the 2-block model the pattern of the advice-searching by members of Group 2 appears to be more heterogeneous than the one exhibited by members of Group 1. In fact, there seemed to be a possible discrepancy between Group 2 members' advice-searching within their own group and their advice searching toward Group 1. In addition, analysis utilizing both AIC and BIC criteria clearly showed that the 2-block did not fit data well compared to the 3-block model. We focus on the results for the 3-block model, and mention briefly the the some advantages of the 3-block model over 2-block model.

The 3-block model gives the following partition of the managers:

Group I: 1,2,6,8,11,12,13,14,16,17,
Group II: 5,10,15,18,19,
Group III: 3,4,7,9,20,21,

with the parameter estimators

$$\hat{\mathbf{p}} = (0.476, 0.238, 0.286).$$

Fig.2 visualizes the Krackhardt network data with the 3-block model.

The 3-block model produces a Group I, which coincides with one of the groups obtained using the 2-block model, while Groups II and III are the result of splitting the other group found in the 2-block model.

We can characterize the properties of these groups in terms of the mean (and the standard deviation) of the age and tenure of the actors, which are given in the table below:

Mean (Standard Deviation)	Age	Tenure
Group I	39.6 (9.8)	13.19 (9.0)
Group II	34.8 (3.56)	6.983 (2.72)
Group III	44 (11.68)	13.306 (8.7)

We note that the two newly formed groups have different characteristics. Group II consists of members who are, on average, younger and have, on average, shorter tenure while Group III consists of older members with the longest tenure. In addition, members of Group II have the smallest standard deviation for age and tenure, thus making this group the most homogeneous group with respect to these two characteristics.

In order to examine the contacts within and between the new groups, we compute the density table below for the 3-block model:

Density of the directed contacts	Group I	Group II	Group III
Group I	0.170	0.160	0.283
Group II	0.740	0.760	0.700
Group III	0.767	0.333	0.417

We conclude that the members of Group I are much less active in seeking advice; however, more of their advice contacts are aimed toward Group III than toward Group II. Moreover, the members of Group II are more inclined to turn to other members for advice, both within their own group and toward the other two groups. One possible explanation could be that this group consists of, on average, the youngest people.

The structure of the advice pattern can now be examined at the individual level by looking at the estimates of the edge parameters given below:

v	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
$\hat{\theta}_{1,v}$	0.4	0.7	0	0.3	0.1	0.2	0.6	0.1	0.1	0.2	0.1	0	0	0.1	0	0.1	0	0.5	0	0	0.7
$\hat{\theta}_{2,v}$	1	1	0.8	0.6	0.8	0.4	0.8	0.8	0.4	0.8	1	0.2	0.8	0.8	0.6	0.8	0.6	0.8	0.8	1	0.6
$\hat{\theta}_{3,v}$	0.67	1	0.17	0.33	0	1	0.5	0.83	0.17	0.5	0.83	1	0	0.83	0.17	0.5	1	1	0	0.5	0.83

We observe that most of the values of $\hat{\theta}_{2,v}$ are higher than those for the remaining two groups ($\hat{\theta}_{1,v}$ and $\hat{\theta}_{3,v}$), which suggests that the contacts from the members of Group II to other members are more likely. This coincides with our previous conclusion that members of Group II are more active than members from the other two groups in searching for advice.

4 The receiver-specific blockmodel

In this section we discuss the receiver-specific blockmodel (RB-model). It is closely related to the SB-model but differs in that it assumes that the probability for a directed edge from, say, vertex u to vertex v depends on the identity of vertex u and the color of vertex v.

The probabilistic model for $(\mathbf{Y}, \mathbf{X})$ assumes that

$$P(\mathbf{Y} = \mathbf{y} \mid \mathbf{X} = \mathbf{x}) = \prod_{1 \le u \ne v \le n} P(Y_{uv} = y_{uv} \mid X_v = x_v)$$

where $(X_u)_{u=1}^n$ are independent, identically distributed rv's, with $P(X_u = i) = p_i, 1 \le i \le c$. Letting, for $1 \le u \ne v \le n$ and $m = 0, 1$,

$$P(Y_{uv} = m \mid x_v) = \psi_{u,x_v}^m (1 - \psi_{u,x_v})^{1-m},$$

we have

$$\begin{aligned} P(\mathbf{Y} = \mathbf{y} \mid \mathbf{X} = \mathbf{x}) &= \prod_{1 \le u \ne v \le n} P(Y_{uv} = y_{uv} \mid X_v = x_v) \\ &= \prod_{1 \le u \ne v \le n} \psi_{u,x_v}^{y_{uv}} (1 - \psi_{u,x_v})^{1-y_{uv}}. \end{aligned} \tag{19}$$

Hence

$$\begin{aligned} P(\mathbf{Y} = \mathbf{y}, \mathbf{X} = \mathbf{x}) &= P(\mathbf{Y} = \mathbf{y} \mid \mathbf{X} = \mathbf{x}) P(\mathbf{X} = \mathbf{x}) \\ &= \prod_{1 \le u \ne v \le n} P(Y_{uv} = y_{uv} \mid X_v = x_v) P(\mathbf{X} = \mathbf{x}) \\ &= \prod_{1 \le u \ne v \le n} \psi_{u,x_v}^{y_{uv}} (1 - \psi_{u,x_v})^{1-y_{uv}} \prod_{j=1}^{c} p_j^{n_j} \\ &= \prod_{u=1}^{n} \prod_{j=1}^{c} p_j^{n_j} \psi_{u,j}^{s_{ju}} (1 - \psi_{u,j})^{n_{uj} - s_{ju}}, \end{aligned} \tag{20}$$

where $n_j = \sum_{v=1}^n I(x_v = j)$, $s_{ju} = \sum_{v=1}^n I(x_v = j) y_{uv}$ and $n_{uj} = n_j - I(x_u = j)$.

Since the analysis of the RB-model parallels that for the SB-model, we provide here only the final formulas to be used later in the data analysis.

To estimate the model parameters, we work with the log likelihood:

$$L(\mathbf{p}, \Psi) = \ln P(\mathbf{y}, \mathbf{x}; \mathbf{p}, \Psi) =$$
$$\sum_{u=1}^{n} \sum_{v: v \ne u} (y_{uv} \ln \psi_{u,x_v} + (1 - y_{uv}) \ln(1 - \psi_{u,x_v})) + \sum_{j=1}^{c} n_j \ln p_j.$$

A straightforward application of the ML-method will produce the following estimators of the parameters in this model:

$$\hat{p}_j = \frac{n_j}{n}, \qquad \hat{\psi}_{u,j} = \frac{s_{ju}}{n_{uj}}. \tag{21}$$

4.1 Parameter estimation in a latent setting

As with the SB-model, the estimation problem becomes much more intricate when it is assumed that group membership $\mathbf{x}$ is unobserved (latent) and our data consist only of the observed network $\mathbf{y}$. An application of the EM-method, which follows the scheme described for the SB-model, provides a two-step estimation algorithm.

Step 1: In Step 1 we calculate (q_{vi}) for $1 \leq v \leq n, 1 \leq i \leq c$ using

$$q_{vi}^{(m)} = \frac{p_i^{(m)} \prod_{u:u\neq v} (\psi_{u,i}^{(m)})^{y_{uv}} (1-\psi_{u,i}^{(m)})^{1-y_{uv}}}{\sum_{j=1}^{c} \left(p_j^{(m)} \prod_{u:u\neq v} (\psi_{u,j}^{(m)})^{y_{uv}} (1-\psi_{u,j}^{(m)})^{1-y_{uv}} \right)}, \tag{22}$$

which satisfies the normalizing condition $\sum_{i=1}^{c} q_{vi}^{(m)} = 1$.

Step 2: In Step 2 the maximum likelihood estimates of the parameters are provided:

$$\hat{p}_i^{(m+1)} = \frac{1}{n} \sum_{v=1}^{n} q_{vi}^{(m)}, \qquad \hat{\psi}_{u,i}^{(m+1)} = \frac{\sum_{v:v\neq u} y_{uv} q_{vi}^{(m)}}{\sum_{v\neq u} q_{vi}^{(m)}}, \tag{23}$$

for $m \geq 0$; $i = 1, ..., c$ and $v = 1, ..., n$. By choosing an arbitrary starting point $(p_1^{(0)}, ..., p_c^{(0)})$ and $(\psi_{u,i}^{(0)})$ for $1 \leq i \leq c; 1 \leq u \leq n$, and iterating equations (22) and (23), we obtain the parameter estimates $(\hat{p}_1, ..., \hat{p}_c)$ and $(\hat{\psi}_{u,i})$ for $1 \leq i \leq c; 1 \leq u \leq n$. The iteration stops when the difference between the estimators obtained in the consecutive rounds is less than a pre-specified threshold value.

4.2 Application of the receiver-specific blockmodel to social network analysis

We return to Hansell's friendship network data. We first explore the group structure of the classmates by fitting the RB-model with two blocks to the data. The EM-algorithm provides the following estimates of the relative sizes of Group 1 resp 2:

$$\hat{\mathbf{p}} = (0.629, 0.371).$$

The color recovery $(\hat{x}_1, ..., \hat{x}_{27})$ of the network vertices is based on the analysis of $q_{vi}, i = 1, 2$, the conditional probabilities of the group membership given the observed network:

v	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
q_{v1}	1	.99	.99	1	1	1	1	1	.99	1	1	.98	1	0	0	0	0	0	1	1	0	0	0	0	0	1	1
q_{v2}	0	.01	.01	0	0	0	0	0	.01	0	0	0.02	0	1	1	1	1	1	0	0	1	1	1	1	1	0	0

Using a threshold value equal to 0.98 to recover the group memberships gives the following partition:

Group 1: 1-13,19,20,26,27,
Group 2: 14-18,21-25.

Fig.4 visualizes the Hansell's network data with the 2-block model.
The students are clearly classified since there are no ambiguous members (that is, members for which $0.02 < q_{v1}, q_{v2} < 0.98$), thus revealing a clear-cut block structure.

Looking at gender composition of the groups, we observe that

No. of students	Male	Female
Group 1	13	4
Group 2	0	10

Now Group 2 consists of only female students, while Group 1 consists of all the male students and four female students. This group composition is very different from that of the 2-block structure found by fitting the SB-model: there the genders are mixed in both groups.

We now describe the structure of the friendship patterns using the group partition above. The friendship patterns classified with respect to the group memberships are given below:

Density of directed contacts	Group 1	Group 2
Group 1	0.1903	0.2176
Group 2	0.0882	0.5000

The above density table can now be compared with the density table below, where the classification is based on gender:

Density of directed contacts	Male	Female
Male	0.2663	0.1374
Female	0.1044	0.3469

We observe that Group 1 members seek friendship contacts both within their own group and toward the students in Group 2, while Group 2 members mainly seek contacts within their own group. In the partition obtained by gender, contacts exist mainly within each group.

Next, the EM-algorithm gives the following estimators of the parameter vector ψ after 100 rounds:

u	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$\hat{\psi}_{u,1}$	0.236	0.059	0.236	0.648	0.293	0.236	0.236	0.412	0.177	0	0	0.177	0	0.059
$\hat{\psi}_{u,2}$	0.199	0.099	0.299	0.099	0.800	0.599	0	0	0	0	0.099	0	0.199	0.599

u	15	16	17	18	19	20	21	22	23	24	25	26	27
$\hat{\psi}_{u,1}$	0	0.589	0	0	0.117	0.411	0.059	0.177	0	0	0	0	0
$\hat{\psi}_{u,2}$	0.498	0.898	0.099	0.199	0.898	0.401	0.399	0.499	0.699	0.299	0.798	0	0

The estimated parameters $(\hat{\psi}_{u,1})$ for $u = 1, ..., 27$ and $(\hat{\psi}_{u,2})$ for $u = 1, ..., 27$ measure the ability of each vertex to seek friendship contacts with students in Group 1 and Group 2, respectively. Student 4 is the most active student when contacts toward members of Group 1 are concerned since

$$\hat{\psi}_{4,1} = \max_{u=1,...,27} \hat{\psi}_{u,1} = 0.648,$$

and students 16 and 19 are the most active students searching for contacts with members of Group 2 since

$$\hat{\psi}_{16,2} = \hat{\psi}_{19,2} = \max_{u=1,...,27} \hat{\psi}_{u,2} = 0.898.$$

These students can thus be considered the most active in seeking contacts with group 1 and 2, respectively.

The estimated parameters $\hat{\psi}$ in the RB-model differ from those in the SB-model in that they indicate the ability of each member to seek contacts with both groups instead of the ability to attract contacts from both groups.

5 Discussion

In this paper we propose novel generalizations of the classic blockmodel. Although blockmodels provide a very useful method for explaining the properties of a network's structure such models have shortcomings because they do not include the actors' unique characteristics. That is blockmodel tells us about the properties of the entire population of a network and subpopulations but they do not tell us very much about the properties of individual members of the population. If we want to understand variations in the behavior of individuals, we need to supplement the global perspective with a closer look at actors' individual characteristics.

We propose generalizations of the blockmodel by introducing probabilities of a directed contact between two vertices that depend on the unique label of the sender vertex and the group menbership of the receiver vertex or, alternatively on the group membership of the sender and the unique label of the receiver. These two models, the sender- and receiver specific blockmodels, allow a more explicit model heterogenity in sending or receiving contacts, respectively. They provide tools for describing and quantifying the variation across individuals in the way they are embedded in "local" social structures as given by the ego networks.

Fitting the sender-specific and receiver-specific blockmodels we are able to model both "out" and "in" ego-nets, thus providing a flexibility which is very advantageous since there is no a single "right" way to define an ego neighborhood for every research question. These two models allow practitioners to model the individual attractiveness and the individual activeness of the actors, respectively.

Appendix

A Proof of equations (12)-(13)

According to (11)

$$\hat{p}_i = \frac{n_i}{n}, \quad \hat{\theta}_{i,v} = \frac{s_{iv}}{n_{iv}},$$

where

$$\begin{aligned}(S_{iv} \mid N_{iv}) &\sim \text{Bin}(n_{iv}, \theta_{i,v}),\\ N_{iv} &\sim \text{Bin}(n-1, p_i),\\ N_i &\sim \text{Bin}(n, p_i).\end{aligned}$$

Hence,

$$E(\hat{\theta}_{i,v}) = EE(\hat{\theta}_{i,v} \mid n_{iv}) = E(\frac{1}{n_{iv}} n_{iv}\theta_{i,v}) = \theta_{i,v}.$$

Furthermore,

$$\begin{aligned}Var(\hat{\theta}_{i,v}) &= EVar(\hat{\theta}_{i,v} \mid n_{iv}) + VarE(\hat{\theta}_{i,v} \mid n_{iv})\\ &= E\frac{N_{iv}\theta_{i,v}(1-\theta_{i,v})}{N_{iv}^2} + Var\frac{n_{iv}\theta_{i,v}}{n_{iv}}\\ &= \theta_{i,v}(1-\theta_{i,v})E(\frac{1}{N_{iv}}).\end{aligned}$$

To get the expected value of the reciprocal of a binomial random variable we use the following Taylor approximation:

$$\begin{aligned}\frac{1}{N_{iv}} = \frac{1}{(n-1)p_i} - \frac{[N_{iv}-(n-1)p_i]}{(n-1)^2 p_i^2} + \frac{2}{2!E^3(N_{iv})}\left(N_{iv}-E(N_{iv})\right)^2\\ - \frac{6}{3!\eta_i^4}\left(N_{iv}-E(N_{iv})\right)^3.\end{aligned} \tag{24}$$

Note that η_i is stochastic and $EN_{iv} < \eta_i < N_{iv}$. Therefore

$$\begin{aligned}E(\frac{1}{N_{iv}}) &= \frac{1}{E(N_{iv})} + \frac{Var(N_{iv})}{E^3(N_{iv})} - E\Big(\frac{\left(N_{iv}-E(N_{iv})\right)^3}{\eta_i^4}\Big)\\ &= \frac{1}{(n-1)p_i} + \frac{(n-1)p_i(1-p_i)}{(n-1)^3 p_i^3} + O(\frac{1}{n^3})\\ &= \frac{(n-2)p_i+1}{(n-1)^2 p_i^2} + O(\frac{1}{n^3}),\end{aligned} \tag{25}$$

which implies

$$Var(\hat{\theta}_{i,v}) = E(\hat{\theta}_{i,v}^2) - \theta_{i,v}^2 = (\theta_{i,v} - \theta_{i,v}^2)\left(\frac{(n-2)p_i + 1}{(n-1)^2 p_i^2} + O(\frac{1}{n^3})\right). \quad (26)$$

B Proof of equation (14)

We want to prove the following equation:

$$\begin{aligned}
&\sum_{x_1=1}^{c} \cdots \sum_{x_n=1}^{c} P(\mathbf{x} \mid \mathbf{y}, \mathbf{p}, \Theta) \\
&\quad \sum_{u=1}^{n} \left[\ln p_{x_u} + \sum_{v:v\neq u} (y_{uv} \ln \theta_{x_u,v} + (1-y_{uv}) \ln(1-\theta_{x_u,v})) \right] \\
&= \sum_{u=1}^{n} \sum_{i=1}^{c} P(X_u = i \mid \mathbf{y}, \mathbf{p}, \Theta) \left[\ln p_i + \sum_{v:v\neq u} (y_{uv} \ln \theta_{i,v} + (1-y_{uv}) \ln(1-\theta_{i,v})) \right].
\end{aligned} \quad (27)$$

By letting

$$h(u,i) = \ln p_i + \sum_{v:v\neq u} [y_{uv} \ln \theta_{i,v} + (1-y_{uv}) \ln(1-\theta_{i,v})]$$

and

$$P(x_1, ..., x_n) = P(\mathbf{x} \mid \mathbf{y}, \mathbf{p}, \Theta), \quad (28)$$

it is enough to prove that

$$\sum_{x_1=1}^{c} \cdots \sum_{x_n=1}^{c} P(x_1, ..., x_n) \sum_{u=1}^{n} h(u, x_u) = \sum_{u=1}^{n} \sum_{i=1}^{c} P(X_u = i \mid \mathbf{y}, \mathbf{p}, \Theta) h(u,i). \quad (29)$$

Noting that (for $u = 1, ..., n$)

$$P(x_u) = \sum_{x_1=1}^{c} \cdots \sum_{x_{u-1}=1}^{c} \sum_{x_{u+1}=1}^{c} \cdots \sum_{x_n=1}^{c} P(x_1, ..., x_n),$$

we have

$$
\begin{aligned}
& \sum_{x_1=1}^{c} \cdots \sum_{x_n=1}^{c} P(x_1,...,x_n) \sum_{u=1}^{n} h(u,x_u) \\
&= \sum_{u=1}^{n} \left[\sum_{x_1=1}^{c} \cdots \sum_{x_n=1}^{c} P(x_1,...,x_n) h(u,x_u) \right] \\
&= \sum_{u=1}^{n} \sum_{x_u=1}^{c} h(u,x_u) \left[\sum_{x_1=1}^{c} \cdots \sum_{x_{i-1}=1}^{c} \sum_{x_{i+1}=1}^{c} \cdots \sum_{x_n=1}^{c} P(x_1,...,x_n) \right] \\
&= \sum_{u=1}^{n} \sum_{x_u=1}^{c} h(u,x_u) P(x_u) \\
&= \sum_{u=1}^{n} \sum_{i=1}^{c} P(X_u = i \mid \mathbf{y}, \mathbf{p}, \Theta) h(u,i),
\end{aligned}
$$

from which (27) follows.

C Proof of equation (15)

To calculate $q_{ui}^{(m)}$ we proceed with the following:

$$
\begin{aligned}
& P(\mathbf{y}, X_u = i \mid \mathbf{p}, \Theta) \\
={}& \sum_{x_1=1}^{c} \cdots \sum_{x_n=1}^{c} I(x_u = i) P(\mathbf{y}, \mathbf{x} \mid \mathbf{p}, \Theta) \\
={}& \sum_{x_1=1}^{c} \cdots \sum_{x_n=1}^{c} I(x_u = i) \prod_{w=1}^{n} \left[p_{x_w} \prod_{v:v\neq w} \theta_{x_w,v}^{y_{wv}} (1-\theta_{x_w,v})^{1-y_{wv}} \right] \\
={}& \sum_{x_1=1}^{c} \cdots \sum_{x_n=1}^{c} I(x_u = i) \left[p_{x_u} \prod_{v:v\neq u} \theta_{x_u,v}^{y_{uv}} (1-\theta_{x_u,v})^{1-y_{uv}} \right] \\
& \prod_{w\neq u} \left[p_{x_w} \prod_{v:v\neq w} \theta_{x_w,v}^{y_{wv}} (1-\theta_{x_w,v})^{1-y_{wv}} \right] \\
={}& \sum_{x_u=1}^{c} \left(I(x_u = i) \left[p_{x_u} \prod_{v:v\neq u} \theta_{x_u,v}^{y_{uv}} (1-\theta_{x_u,v})^{1-y_{uv}} \right] \right) \\
& \sum_{x_u=1}^{c} \cdots \sum_{x_{u-1}=1}^{c} \sum_{x_{u+1}=1}^{c} \cdots \sum_{x_n=1}^{c} \left(\prod_{w\neq u} \left[p_{x_w} \prod_{v:v\neq w} \theta_{x_w,v}^{y_{wv}} (1-\theta_{x_w,v})^{1-y_{wv}} \right] \right) \\
={}& \left(p_i \prod_{v:v\neq u} \theta_{i,v}^{y_{uv}} (1-\theta_{i,v})^{1-y_{uv}} \right) \prod_{w\neq u} \sum_{x_w=1}^{c} \left[p_{x_w} \prod_{v:v\neq w} \theta_{x_w,v}^{y_{wv}} (1-\theta_{x_w,v})^{1-y_{wv}} \right] \\
={}& \left(p_i \prod_{v:v\neq u} \theta_{i,v}^{y_{uv}} (1-\theta_{i,v})^{1-y_{uv}} \right) \prod_{w\neq u} \sum_{j=1}^{c} \left[p_j \prod_{v:v\neq w} \theta_{j,v}^{y_{wv}} (1-\theta_{j,v})^{1-y_{wv}} \right]
\end{aligned}
\tag{30}
$$

and then

$$\begin{aligned} P(\mathbf{y} \mid \mathbf{p}, \Theta) &= \sum_{i=1}^{c} P(\mathbf{y}, X_u = i \mid \mathbf{p}, \Theta) \\ &= \sum_{i=1}^{c} \left[\left(p_i \prod_{v:v\neq u} \theta_{i,v}^{y_{uv}} (1-\theta_{i,v})^{1-y_{uv}} \right) \prod_{w\neq u} \sum_{j=1}^{c} \left(p_j \prod_{v:v\neq w} \theta_{j,v}^{y_{wv}} (1-\theta_{j,v})^{1-y_{wv}} \right) \right] \\ &= \prod_{w=1}^{n} \sum_{j=1}^{c} \left(p_j \prod_{v:v\neq w} \theta_{j,v}^{y_{wv}} (1-\theta_{j,v})^{1-y_{wv}} \right). \end{aligned} \tag{31}$$

Using (30) and (31), the formula for $q_{ui}^{(m)}$ is given by:

$$q_{ui}^{(m)} = \frac{p_i^{(m)} \prod_{v:v\neq u} (\theta_{i,v}^{(m)})^{y_{uv}} (1-\theta_{i,v}^{(m)})^{1-y_{uv}}}{\sum_{j=1}^{c} \left(p_j^{(m)} \prod_{v:v\neq u} (\theta_{j,v}^{(m)})^{y_{uv}} (1-\theta_{j,v}^{(m)})^{1-y_{uv}} \right)}, \tag{32}$$

which satisfies the normalizing condition $\sum_{i=1}^{c} q_{ui}^{(m)} = 1$.

D Profile predictive likelihood

An alternative approach to recovering the latent group membership vector $\mathbf{x}$ is to use the profile predictive likelihood method, where the likelihood function we need to maximize is defined by:

$$L_p(\mathbf{x} \mid \mathbf{y}) = \sup_{\mathbf{p},\Theta} P(\mathbf{y}, \mathbf{x} \mid \mathbf{p}, \Theta). \tag{33}$$

The function can be simplified by first calculating the ML-estimates of the parameters based on the original likelihood function $P(\mathbf{y}, \mathbf{x}; \mathbf{p}, \Theta)$, and then plugging these estimates into $L_p(\mathbf{x} \mid \mathbf{y})$ in order to get an expression for the profile predictive likelihood.

We recall that, in the case that both the edge structure and the vertex colors are known, the ML-estimate for the parameters in the directed SB-model are:

$$\hat{p}_i = \frac{n_i}{n}, \quad \hat{\theta}_{i,v} = \frac{s_{iv}}{n_{iv}}.$$

By plugging these expressions into (33) we get the following formula for the profile likelihood function:

$$\begin{aligned}L_p(\mathbf{x} \mid \mathbf{y}, \hat{\mathbf{p}}, \hat{\Theta}) &= \prod_{u=1}^{n} \left(\hat{p}_{x_u} \prod_{v:v\neq u} \hat{\theta}_{x_u,v}^{y_{uv}} (1-\hat{\theta}_{x_u,v})^{1-y_{uv}} \right) \\ &= \prod_{u=1}^{n} \Big(\frac{1}{n} \sum_{v=1}^{n} I(x_v = x_u) \prod_{v:v\neq u} \Big(\frac{\sum_{w:w\neq v} y_{wv} I(x_w = x_u)}{\sum_{w\neq v} I(x_w = x_u)} \Big)^{y_{uv}} \\ &\quad \Big(1 - \frac{\sum_{w:w\neq v} y_{wv} I(x_w = x_u)}{\sum_{w\neq v} I(x_w = x_u)} \Big)^{1-y_{uv}} \Big).\end{aligned} \tag{34}$$

The partition of vertices into c groups can now be carried out by maximizing $L_p(\mathbf{x} \mid \mathbf{y}, \hat{\mathbf{p}}, \hat{\Theta})$ thereby selecting the vector $\mathbf{x}$, which provides the highest profile likelihood function value. The direct maximization can be carried out by enumerating $L_p(\mathbf{x} \mid \mathbf{y}, \hat{\mathbf{p}}, \hat{\Theta})$ for all possible 2^n values of $\mathbf{x}$. This operation is, however, very time-consuming since, even for moderate values of n, the number of enumerations is very large.

An alternative approach for n values so large that the enumeration approach is impractical could be to first generate a random starting point for the color vector, and then do the following iteration: we maximize (34) as a function of $x_u \in \{1, ..., c\}$ for each $u = 1, ..., n$, when all the coordinates $(x_1, ..., x_n)$ are updated. We repeat this process for a new round, and maximize (34) as a function of x_u for each u. The algorithm stops when the updating of the $\mathbf{x}$ values no longer increases the likelihood function (34).

E The figures and outputs

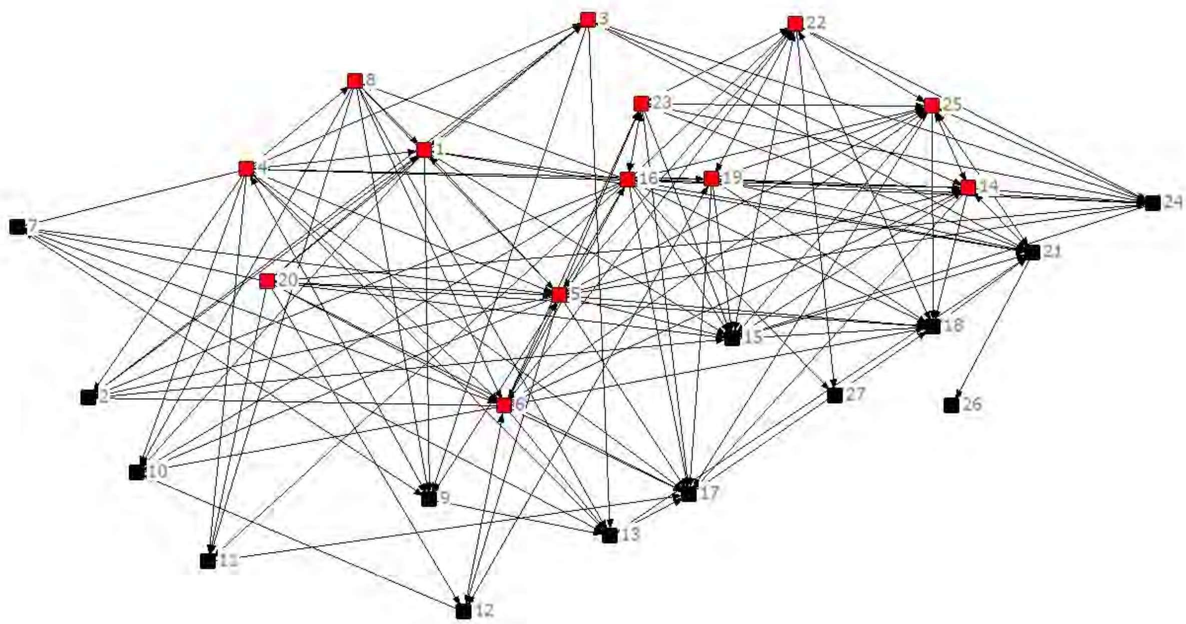

Figure 1: Hansell's friendship data: 2-block model

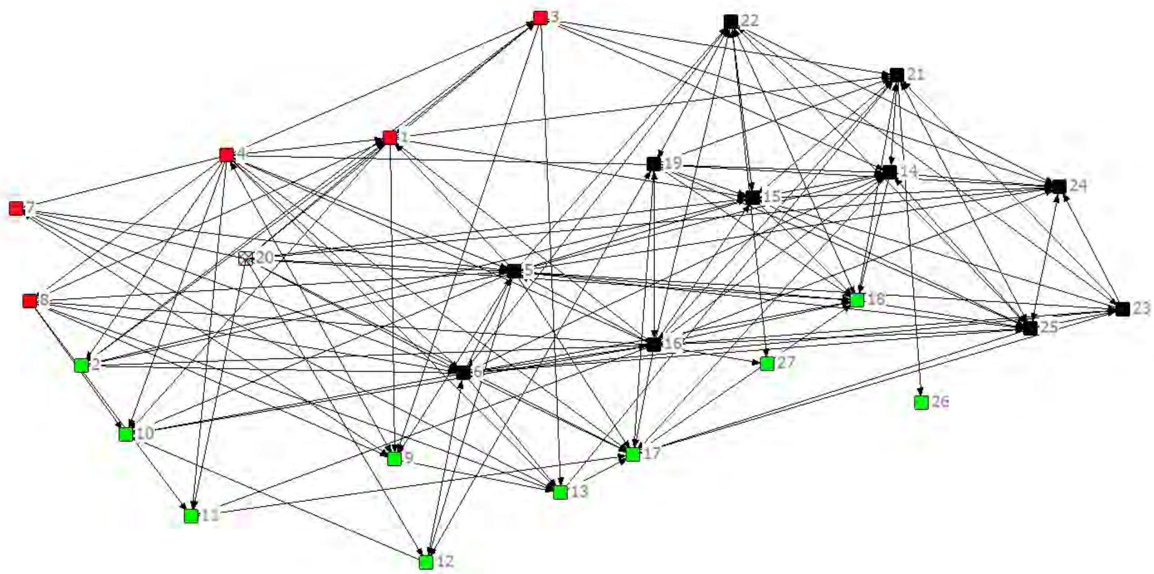

Figure 2: Hansell's friendship data: 3-block model

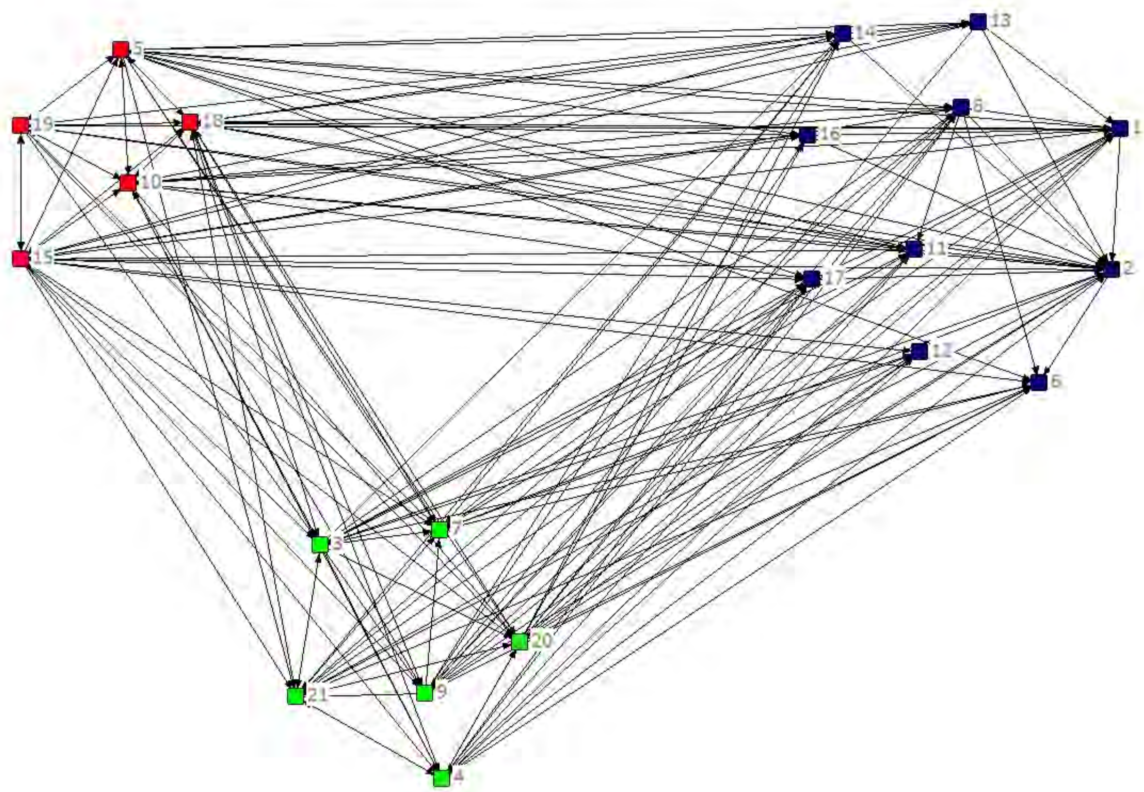

Figure 3: Krackhardt's high-tech managers data: 3-block model

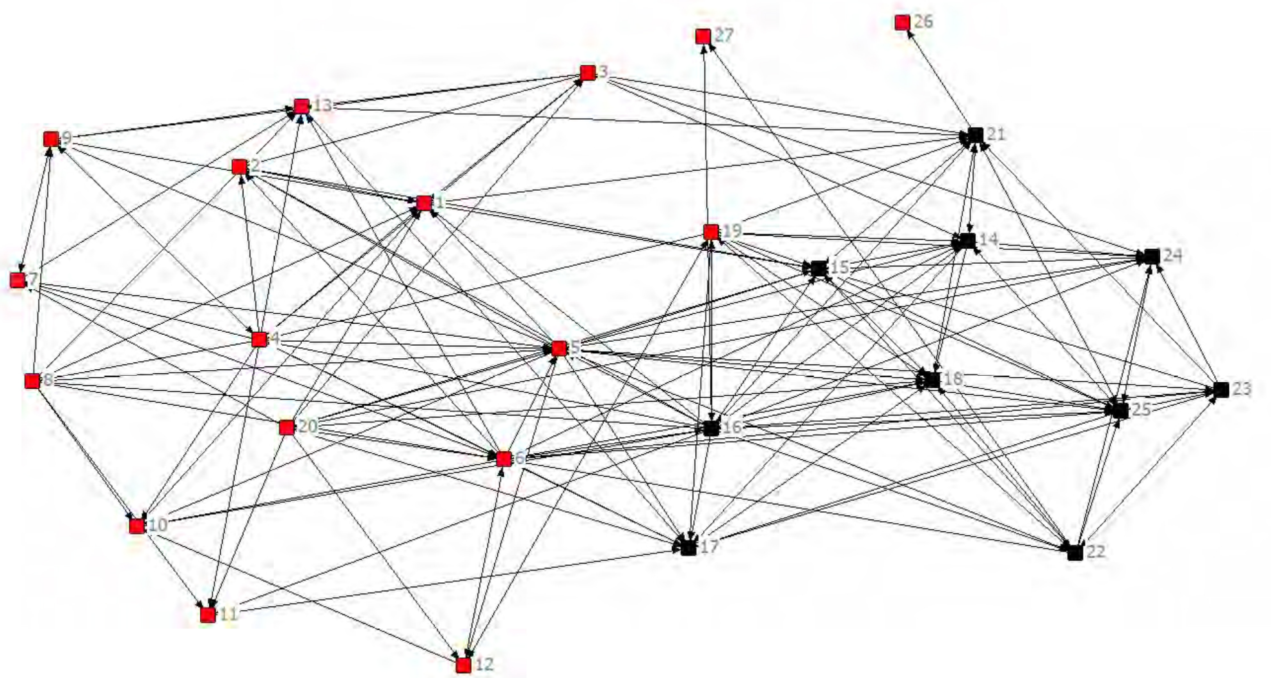

Figure 4: Hansell's friendship data: 2-block receiver model

Acknowledgement

The authors would like to thank Prof Ove Frank for valuable suggestions and comments. We would also like to thank the referee whose insightful comments and suggestions led to considerable improvement in the presentation of this paper.

References

[1] Andersen, E. B. (1980). Discrete Statistical Models with Social Science Applications. NorthHolland, Amsterdam.

[2] Boer, P., Huisman, M., Snijders, T. and Zeggelink, E.P.H. (2003). StOCNET: an open software system for the advanced statistical analysis of social networks. Version 1.4. Groningen: ProGAMMA / ICS.

[3] Burnham, K. P. and Anderson, D. R. (2002). Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach (2nd ed.). Springer-Verlag, ISBN 0-387-95364-7.

[4] Crossley, N., Bellotti, E., Edwards, G., Everett, M. G., Koskinen, J. and Tranmer, M. (2015) Social Network Analysis, An Actor Centred Approach. Sage Publications Ltd, ISBN: 9781446267769.

[5] Dampster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society, B (39):1-38.

[6] Daudin, J. Picard, F. and Robin, S. (2008). A mixture model for random graph. Statistics and computing (18):173-183.

[7] Hansell, S. (1984). Cooperative groups, weak ties, and the integration of peer friendships. Social Psychology Quarterly, 47, 316-328.

[8] Kaiser, M. (2008). Mean clustering coefficients: the role of isolated nodes and leafs on clustering measures for small-world networks. New J. Phys. 10 083042.

[9] Kaiser, M. (2008). Mean clustering coefficients: the role of isolated nodes and leafs on clustering measures for small-world networks. New J. Phys. 10 083042.

[10] Krackhardt, D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

[11] Luce, R.D. and Perry, A.D. (1949). A method of matrix analysis of group structure. Psychometrika 14 (1): 95-116.

[12] Newman, M. (2006). Finding community structure in networks using the eigenvectors of matrices. Phys. Rev. E 74, 036104.

[13] Newman, M. and Leicht, E. (2007). Mixture models and exploratory analysis in networks. PNAS 104:9564-9569.

[14] Nowicki, K. and Snijders, T. (2001). Estimation and prediction for stochastic blockstructures. J. Amer. Statist. Assoc., 96(455):1077-1087.

[15] Picard, F. (2008). An Introduction to mixture models. SSB preprint (7).

[16] Snijders, T. and Nowicki, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. Journal of Classification, 14, 75 -100.

[17] Wasserman, S. and Faust, K. (1994). Social Network Analysis: Methods and Application. Cambridge: Cambridge University Press.

[18] G. B. Folland. Higher-Order Derivatives and Taylor's Formula in Several Variables.

[19] Watts, D.J. and Strogatz, S.H. (1998). Collective dynamics of 'small-world' networks. Nature 393 (6684): 440-442.

[20] Yang, Y. (2005). Can the strengths of AIC and BIC be shared?. Biometrika 92: 937-950.

[21] Zachary, W. (1977). An information flow model for conflict and fission in small groups. Journal of Anthropological Research 33, 452-473.