

Bayesian Community Detection

Morten Mørup

mm@imm.dtu.dk

Mikkel N. Schmidt

mns@imm.dtu.dk

*Section for Cognitive Systems, Technical University of Denmark,
2800 Lyngby, Denmark*

Many networks of scientific interest naturally decompose into clusters or communities with comparatively fewer external than internal links; however, current Bayesian models of network communities do not exert this intuitive notion of communities. We formulate a nonparametric Bayesian model for community detection consistent with an intuitive definition of communities and present a Markov chain Monte Carlo procedure for inferring the community structure. A Matlab toolbox with the proposed inference procedure is available for download. On synthetic and real networks, our model detects communities consistent with ground truth, and on real networks, it outperforms existing approaches in predicting missing links. This suggests that community structure is an important structural property of networks that should be explicitly modeled.

1 Introduction

The analysis of complex networks is an important challenge spurred by many types of networked data arising in practically all fields of science, including biology, social science, and technology (Girvan & Newman, 2002; Sun, Ling, Zhang, Li, & Chen, 2003; Watts & Strogatz, 1998). Many networks naturally decompose into clusters or communities characterized in this way (Fortunato, 2010):

Definition 1. *The organization of vertices in clusters, with many edges joining vertices of the same cluster and comparatively few edges joining vertices of different clusters.*

Such communities have been found to correspond to behavioral or functional units (Karrer, Levina, & Newman, 2008; Gulbahce & Lehmann, 2008; Fortunato, 2010). For example, in networks of protein interaction, communities might comprise proteins with similar functions, and in social networks, communities are groups of closely related people. This suggests that we might gain insight into networks whose function is less well understood

by discovering and examining their communities (Karrer et al., 2008; Fortunato, 2010). A large variety of methods for community detection have been proposed over the years; however, no standards have been agreed on (Gulbahce & Lehmann, 2008). A widely used approach is to optimize some measure of cluster structure such as modularity (Newman & Girvan, 2004) or mutual information (Rosvall & Bergstrom, 2007). An alternative approach is to formulate a Bayesian generative network model, with the advantage of making the definition of the model's notion of a community explicit. The most prominent generative model of network communities is the stochastic block model (Nowicki & Snijders, 2001) and its extensions (Karrer & Newman, 2011; Hofman & Wiggins, 2008; Miller, Griffiths, & Jordan, 2009; Mørup, Schmidt, & Hansen, 2010). Notably, the infinite relational model (IRM) (Kemp, Tenenbaum, Griffiths, Yamada, & Ueda, 2006; Xu, Tresp, Yu, & Krieger, 2006) is a nonparametric extension that allows the number of communities to be inferred from data. In these models, clusters are sets of nodes with homogeneous probabilities of linking within and to other clusters. However, this is subtly different from the intuitive definition of communities above. For example, a group of nodes that link consistently to other groups of nodes but have no internal links between them will be identified as a cluster. This flexible notion of group structure makes it impossible to interpret clusters as communities in the sense conveyed in definition 1.

In this letter, we propose a nonparametric Bayesian generative model of networks with community structure that operationalizes the intuitive definition of communities in definition 1. We derive an efficient inference procedure based on Markov chain Monte Carlo (MCMC). Our model includes a parameter that defines the extent to which a community has comparatively fewer external than internal links. This parameter is learned from data, giving an indication of the strength of the community structure. We compare our model to the IRM and demonstrate how the communities detected by the two models differ on synthetic and real data. Finally, we analyze 17 real networks and show that our model outperforms competing approaches as measured by link prediction. Our results demonstrate that community structure is an important property of networks that should be modeled explicitly. A Matlab toolbox with our proposed method is available for download (Mørup & Schmidt, 2011).

2 Methods

Given a network, we address the problem of partitioning the nodes into clusters, consistent with the notion of communities in definition 1. To simplify the presentation, we consider an undirected network represented by a binary, strictly upper triangular adjacency matrix A , where $A_{ij} = 1$ indicates that a link between node i and j is present; however, the ideas presented

can trivially be extended to directed and weighted networks as well as multigraphs.

To initiate the discussion, we describe the infinite relational model (IRM) (Kemp et al., 2006; Xu et al., 2006) which can be characterized by the following generative process. First, each node is assigned to a cluster according to a Chinese restaurant process (CRP). The CRP is a metaphor for building a partition ground up by assigning the first node (i.e., customer in a restaurant to a table) and subsequent nodes (customers arriving at the restaurant) to an existing cluster with a probability proportional to how many existing customers are placed at the table and at a new table with a probability proportional to α . Customers tend to sit at most popular tables, making the popular tables even more popular, an effect noted as the rich gets richer. The partition of the nodes induced by the CRP is exchangeable in that the order in which the customers arrive does not influence the probability of the partition (Pitman, 2006). Next, link probabilities are generated that specify the probability of observing a link within and between each cluster, and, finally, the links in the network are generated according to these probabilities:

$$\text{Cluster assignment: } \mathbf{z} \sim \text{CRP}(\alpha), \quad (2.1)$$

$$\text{Link probability: } \eta_{\ell m} \sim \text{Beta}(\beta, \beta), \quad (2.2)$$

$$\text{Link: } A_{ij} \sim \text{Bernoulli}(\eta_{z_i z_j}). \quad (2.3)$$

Inference in the IRM model (i.e., determining the posterior distribution of the cluster assignments) entails marginalizing over the link probabilities, which can be done analytically. This is a major advantage of the IRM model, enabling inference by Markov chain Monte Carlo (MCMC) sampling over the cluster assignments alone.

However, the notion of a cluster in the IRM model is not consistent with definition 1. A cluster ℓ is defined by the probability $\eta_{\ell\ell}$ of observing a link between two nodes inside the cluster and the probabilities $\eta_{\ell m}$ for $\ell \neq m$ of observing links between nodes in cluster ℓ and m . This is subtly different from the intuitive notion of communities with more internal than external links.

2.1 Bayesian Community Detection. In this letter, we propose a nonparametric Bayesian model of network communities that strictly follows definition 1. Our model can be described by the following generative process. First, a cluster assignment is generated, partitioning the nodes into K clusters. For each cluster, a within-cluster link probability, $\eta_{\ell\ell}$, is generated specifying the probability of observing a link between two nodes in the cluster. Then for each cluster, a cluster gap, γ_ℓ , between 0 and 1 is generated. The cluster gap multiplied by the within-cluster link probability determines the maximum allowable between-cluster link probability. Next, the probability

of links between each pair of different clusters ℓ and m is considered. For each pair, a between-cluster link probability $\eta_{\ell m}$ is generated, such that it is strictly less than both clusters' within-cluster link probability multiplied by the gap. Finally, links between nodes in the same or in different clusters are generated according to the within- and between-cluster link probabilities:

$$\begin{aligned} \text{Cluster assignment: } z &\sim \text{CRP}(\alpha), \\ \text{Within-cluster link probability: } \eta_{\ell\ell} &\sim \text{Beta}(\beta, \beta), \\ \text{Cluster gap: } \gamma_\ell &\sim \text{Beta}(\vartheta, \vartheta), \\ \text{Between-cluster link probability, } x_{\ell m} &= \min[\gamma_\ell \eta_{\ell\ell}, \gamma_m \eta_{m m}]: \\ \eta_{\ell m} &\sim \text{BetaInc}(\beta, \beta, x_{\ell m}), \\ \text{Link: } A_{ij} &\sim \text{Bernoulli}(\eta_{z_i z_j}), \end{aligned}$$

where $\text{BetaInc}(a, b, x)$ denotes a beta distribution constrained to the interval $[0, x]$, with density,

$$p(\theta) = \frac{1}{B_x(a, b)} \theta^{a-1} (1 - \theta)^{b-1},$$

where $B_x(a, b)$ is the incomplete beta function. The incomplete beta distribution has previously been considered for binomial sampling in Weiler (1965), and aspects regarding numerical evaluation can be found in Dutka (1981). We denote the hyperparameters of the model by $\psi = \{\alpha, \vartheta, \beta\}$.

We name this model *Bayesian community detection* (BCD). According to the BCD model, clusters are groups of nodes where the probability of links between them is at least γ_ℓ times less compared to the probability of links within them. In addition to the model where each cluster has a separate γ_ℓ , we also consider the special case where all clusters share a common $\gamma_1 = \gamma_2 = \dots = \gamma$. Figure 1 shows random networks generated according to the IRM and the proposed BCD model with shared γ . As γ decreases, the relative link densities between clusters drop, and for $\gamma = 0$, no links are generated between clusters. We note that we recover the IRM model for $\gamma \rightarrow \infty$ in the BCD, but in order for the BCD model to adhere to definition 1, we require $\gamma \leq 1$.

2.2 Inference in BCD. Let $\hat{\eta} = \{\eta_{\ell m} | \ell = m\}$ and $\check{\eta} = \{\eta_{\ell m} | \ell \neq m\}$ denote the sets of link probabilities within and between clusters, respectively. Inference in BCD amounts to computing the marginal posterior distribution of the cluster assignments, $p(z|A)$, which entails integrating over the joint

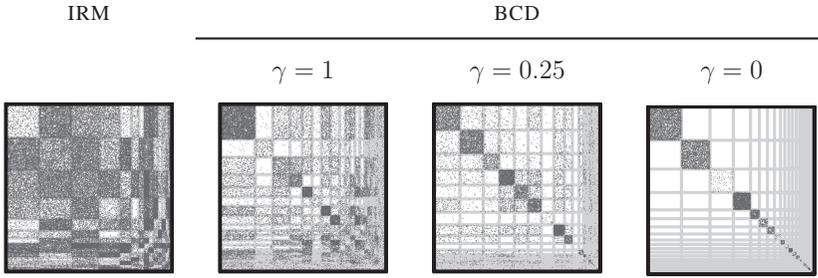


Figure 1: Examples of random graphs according to IRM and BCD for different values of γ .

distribution given by

$$\begin{aligned}
 p(A, z, \eta, \boldsymbol{\gamma} | \boldsymbol{\psi}) &= p(A | z, \eta) p(\tilde{\eta} | \tilde{\boldsymbol{\eta}}, \boldsymbol{\gamma}, \beta) p(\tilde{\boldsymbol{\eta}} | \beta) p(\boldsymbol{\gamma} | \vartheta) p(z | \alpha) \\
 &= \left[\prod_{i=1}^{N-1} \prod_{j=i+1}^N \eta_{z_i z_j}^{A_{ij}} (1 - \eta_{z_i z_j})^{1-A_{ij}} \right] \\
 &\quad \times \left[\prod_{\ell=1}^{L-1} \prod_{m=\ell+1}^L \frac{\eta_{\ell m}^{\beta-1} (1 - \eta_{\ell m})^{\beta-1}}{B_{x_{\ell m}}(\beta, \beta)} \right] \\
 &\quad \times \left[\prod_{\ell=1}^L \frac{\eta_{\ell \ell}^{\beta-1} (1 - \eta_{\ell \ell})^{\beta-1}}{B(\beta, \beta)} \right] \times \left[\prod_{\ell=1}^L \frac{\gamma_{\ell}^{\vartheta-1} (1 - \gamma_{\ell})^{\vartheta-1}}{B(\vartheta, \vartheta)} \right] \\
 &\quad \times \left[\frac{\alpha^L \Gamma(\alpha)}{\Gamma(N + \alpha)} \prod_{\ell=1}^L \Gamma(M_{\ell}) \right],
 \end{aligned}$$

where L is the number of clusters, M_{ℓ} is the number of nodes in the ℓ th cluster, and $N_{\ell m}^+$ and $N_{\ell m}^-$ are the number of links and nonlinks between nodes in cluster ℓ and m such that

$$N_{\ell m}^- = \begin{cases} M_{\ell} M_m - N_{\ell m}^+ & \text{for } l \neq m \\ M_{\ell} (M_{\ell} - 1) / 2 - N_{\ell \ell}^+ & \text{for } l = m \end{cases} .$$

2.2.1 Between-Cluster Link Probabilities. In the IRM model, all link probabilities can be marginalized analytically. This is not the case for BCD since within- and between-cluster link probabilities are dependent; however, the vast majority of these parameters, namely, the between-cluster link probabilities ($\tilde{\boldsymbol{\eta}} = \{\eta_{\ell m} | \ell \neq m\}$), can be marginalized analytically. We therefore

have:

$$\begin{aligned}
 p(A, z, \dot{\eta}, \boldsymbol{\gamma} | \boldsymbol{\psi}) &= \int p(A, z, \boldsymbol{\eta}, \boldsymbol{\gamma} | \boldsymbol{\psi}) d\dot{\boldsymbol{\eta}} \\
 &= \left[\prod_{\ell=1}^L \frac{\eta_{\ell\ell}^{N_{\ell\ell}^+ + \beta - 1} (1 - \eta_{\ell\ell})^{N_{\ell\ell}^- + \beta - 1}}{B(\beta, \beta)} \right] \\
 &\quad \times \left[\prod_{\ell=1}^{L-1} \prod_{m=\ell+1}^L \frac{B_{x_{\ell m}}(N_{\ell m}^+ + \beta, N_{\ell m}^- + \beta)}{B_{x_{\ell m}}(\beta, \beta)} \right] \\
 &\quad \times \left[\prod_{\ell=1}^L \frac{\gamma_{\ell}^{\vartheta - 1} (1 - \gamma_{\ell})^{\vartheta - 1}}{B(\vartheta, \vartheta)} \right] \times \left[\frac{\alpha^L \Gamma(\alpha)}{\Gamma(N + \alpha)} \prod_{\ell=1}^L \Gamma(M_{\ell}) \right].
 \end{aligned}$$

Because we cannot integrate the posterior with respect to the remaining parameters, z , $\dot{\boldsymbol{\eta}}$, and $\boldsymbol{\gamma}$, we sample from their posterior distribution using MCMC.

2.2.2 *Within-Cluster Link Probabilities.* When we eliminate terms that do not depend on $\eta_{\ell\ell}$, the marginal posterior reduces to

$$\begin{aligned}
 p(\eta_{\ell\ell} | A, z, \dot{\boldsymbol{\eta}} \setminus \eta_{\ell\ell}, \boldsymbol{\gamma}, \boldsymbol{\psi}) &\propto \eta_{\ell\ell}^{N_{\ell\ell}^+ + \beta - 1} (1 - \eta_{\ell\ell})^{N_{\ell\ell}^- + \beta - 1} \\
 &\quad \cdot \prod_{m \neq \ell} \frac{B_{x_{\ell m}}(N_{\ell m}^+ + \beta, N_{\ell m}^- + \beta)}{B_{x_{\ell m}}(\beta, \beta)}.
 \end{aligned}$$

We generate samples from this distribution using Metropolis-Hastings with the following proposals:

$$\begin{aligned}
 q_1(\eta_{\ell\ell}^*) &= \text{Beta}(\beta, \beta), \\
 q_2(\eta_{\ell\ell}^*) &= \text{Beta}(N_{\ell\ell}^+ + \beta, N_{\ell\ell}^- + \beta), \\
 q_3(\eta_{\ell\ell}^* | \eta_{\ell\ell}) &= \text{Beta}(\eta_{\ell\ell} C, (1 - \eta_{\ell\ell}) C).
 \end{aligned}$$

The first is the prior, the second is the marginal distribution of $\eta_{\ell\ell}$ disregarding dependencies to other clusters, and the third is a random walk centered on the current value of $\eta_{\ell\ell}$ with concentration parameterized by C . Combining these proposals, we attain acceptance rates around 50%, and to improve mixing, we repeat the sampling 10 times in each iteration.

2.2.3 *Cluster Assignment Variables.* Regarding, z , the BCD model is similar to a Dirichlet process (DP) mixture model, and it is thus possible to use standard MCMC methods for inference in DP mixtures. Because we are

not able to analytically integrate $\dot{\eta}$, we have to resort to nonconjugate sampling approaches. We use two MCMC transition kernels: an auxiliary variable Gibbs sampler (Neal, 2000, algorithm 8) and a split-merge Metropolis-Hastings move (Jain & Neal, 2004), to sample from $p(z|A, \dot{\eta}, \boldsymbol{\gamma}, \boldsymbol{\psi})$. In the auxiliary variable Gibbs sampler T , new within-community densities and gap parameters are drawn from their priors. These auxiliary variables represent possible values for the parameters of the within-community density of nodes that are not associated with any other observations. The node assignment is then updated by Gibbs sampling with respect to the distribution that includes these auxiliary parameters (see also Neal, 2000, algorithm 8, for details). Similarly, the split-merge Metropolis-Hastings algorithm proposes in a split move a new within-community density and gap parameter from the priors as part of defining the launch state for the restricted Gibbs sampler that Jain and Neal (2004) described.

The conditional distribution of the cluster assignment of a single node required for the Gibbs sampler, as well as the restricted Gibbs sweeps in the split-merge sampler, is given by

$$p(z_i = \ell | A, \mathbf{z} \setminus z_i, \dot{\eta}, \boldsymbol{\gamma}, \boldsymbol{\psi}) \propto \eta_{\ell\ell}^{n_{i\ell}^+} (1 - \eta_{\ell\ell})^{n_{i\ell}^-} \alpha^L M_\ell \cdot \prod_{m \neq \ell} \frac{B_{x_{\ell m}}(N_{\ell m}^+ + n_{im}^+ + \beta, N_{\ell m}^- + n_{im}^- + \beta)}{B_{x_{\ell m}}(N_{\ell m}^+ + \beta, N_{\ell m}^- + \beta)}, \quad (2.4)$$

where $n_{i\ell}^+$ and $n_{i\ell}^-$ are the number of links and nonlinks from node i to nodes in cluster ℓ , and (by slight abuse of notation) $N_{\ell m}^+, N_{\ell m}^-$, and M_ℓ are computed excluding links involving node i .

2.2.4 Cluster Gaps. When terms that do not depend on γ_ℓ are eliminated, the posterior reduces to

$$p(\gamma_\ell | A, \mathbf{z}, \dot{\eta}, \boldsymbol{\gamma} \setminus \gamma_\ell, \boldsymbol{\psi}) \propto \gamma_\ell^{\vartheta-1} (1 - \gamma_\ell)^{\vartheta-1} \cdot \prod_{m \neq \ell} \frac{B_{x_{\ell m}}(N_{\ell m}^+ + \beta, N_{\ell m}^- + \beta)}{B_{x_{\ell m}}(\beta, \beta)}.$$

To generate samples from $\boldsymbol{\gamma}$, we use Metropolis-Hastings with the following proposal densities,

$$q_1(\gamma_\ell^*) = \text{Beta}(\vartheta, \vartheta), \quad q_2(\gamma_\ell^* | \gamma_\ell) = \text{Beta}(\gamma_\ell C, (1 - \gamma_\ell)C),$$

where the first is the prior distribution, and the second is a random walk centered at the current value of γ_ℓ with concentration parameterized by C . To improve mixing, we repeat the sampling 10 times in each iteration.

2.3 Predicting Missing Links. Until now it has been assumed that the data are a fully observed network; however, the model can easily be adapted to accommodate a partially observed network. When it is unknown whether there is a link between some sets of pairs of nodes in the network, these are simply not included in the evaluation of the likelihood ($p(A|z, \eta)$), or equivalently excluded in the computation of the link- and nonlink-counts, $N_{\ell m}^+$ and $N_{\ell m}^-$.

When the network is not fully observed, the model can be used to predict if there should be a link between two nodes by computing the posterior predictive distribution, $p(A_{ij}|A)$, where A denotes the observed part of the network. Based on R posterior samples, the predictive distribution is approximated as

$$p(A_{ij}|A) \approx \frac{1}{R} \sum_{r=1}^R p(A_{ij}|A, z^{(r)}, \hat{\eta}^{(r)}, \boldsymbol{\gamma}^{(r)}),$$

where

$$p(A_{ij} = 1|A, z, \hat{\eta}, \boldsymbol{\gamma}) = \begin{cases} \eta_{z_i z_j}, & z_i = z_j, \\ \frac{B_{x_{z_i z_j}}(N_{z_i z_j}^+ + \beta + 1, N_{z_i z_j}^- + \beta)}{B_{x_{z_i z_j}}(N_{z_i z_j}^+ + \beta, N_{z_i z_j}^- + \beta)}, & z_i \neq z_j. \end{cases}$$

2.4 Extensions of the BCD Model. BCD can trivially be extended to handle self-links, in which case the adjacency matrix is nonzero on the diagonal, as well as directed networks, where the adjacency matrix is a full binary matrix. Extension to integer weighted networks is also trivial, replacing the Bernoulli likelihood and beta priors by Poisson and gamma distributions. Furthermore, BCD can trivially be extended to handle multiple types of links by sharing the cluster assignments across link types and introducing different link probability and gap parameters for each link type. The details of all these possible extensions are described below.

2.4.1 Extension to Weighted Graphs. For the modeling of weighted graphs based on integer count co-occurrences, we have the following generative process for the IRM model:

$$\text{Cluster assignment: } z \sim \text{CRP}(\alpha), \tag{2.5}$$

$$\text{Link probability: } \eta_{\ell m} \sim \text{Gamma}(\beta, \beta), \tag{2.6}$$

$$\text{Link: } A_{ij} \sim \text{Poisson}(\eta_{z_i z_j}), \tag{2.7}$$

and for the BCD,

Cluster assignment: $z \sim \text{CRP}(\alpha)$,

Within-cluster link probability: $\eta_{\ell\ell} \sim \text{Gamma}(\beta, \beta)$,

Cluster gap, $\gamma_\ell \sim \text{Beta}(\vartheta, \vartheta)$,

Between-cluster link probability, $x_{lm} = \min(\gamma_\ell \eta_{\ell\ell}, \gamma_m \eta_{mm})$,

$\eta_{\ell m} \sim \text{GammaInc}(\beta, \beta, x_{lm})$,

Link: $A_{ij} \sim \text{Poisson}(\eta_{z_i z_j})$.

Let $G_x(a, b)$ denote the normalization constant of the incomplete gamma distribution ($\text{GammaInc}(a, b, x)$) (i.e., constrained to the interval $[0, x]$). Marginalizing the between-cluster rate η_{lm} , we obtain

$$\begin{aligned}
 p(A, z, \dot{\eta}, \boldsymbol{\gamma} | \boldsymbol{\psi}) &= \int p(A, z, \boldsymbol{\eta}, \boldsymbol{\gamma} | \boldsymbol{\psi}) d\dot{\eta} \\
 &= \left[\prod_{\ell=1}^L \frac{\eta_{\ell\ell}^{N_{\ell\ell}^+ + \beta - 1} \exp[-\eta_{\ell\ell} (N_{\ell\ell}^{\text{tot}} + \beta)]}{G(\beta, \beta)} \right] \\
 &\quad \times \left[\prod_{\ell=1}^{L-1} \prod_{m=\ell+1}^L \frac{G_{x_{\ell m}}(N_{\ell m}^+ + \beta, N_{\ell m}^{\text{tot}} + \beta)}{G_{x_{\ell m}}(\beta, \beta)} \right] \\
 &\quad \times \left[\prod_{\ell=1}^L \frac{\gamma_\ell^{\vartheta-1} (1 - \gamma_\ell)^{\vartheta-1}}{B(\vartheta, \vartheta)} \right] \times \left[\frac{\alpha^L \Gamma(\alpha)}{\Gamma(N + \alpha)} \prod_{\ell=1}^L \Gamma(M_\ell) \right].
 \end{aligned}$$

2.4.2 *Extension to Directed Graphs.* For directed graphs, we have for the marginalized likelihood given above for integer-weighted graphs,

$$\begin{aligned}
 p(A, z, \dot{\eta}, \boldsymbol{\gamma} | \boldsymbol{\psi}) &= \int p(A, z, \boldsymbol{\eta}, \boldsymbol{\gamma} | \boldsymbol{\psi}) d\dot{\eta} \\
 &= \left[\prod_{\ell=1}^L \frac{\eta_{\ell\ell}^{N_{\ell\ell}^+ + \beta - 1} \exp[-\eta_{\ell\ell} (N_{\ell\ell}^{\text{tot}} + \beta)]}{G(\beta, \beta)} \right] \\
 &\quad \times \left[\prod_{\ell=1}^L \prod_{m \neq \ell}^L \frac{G_{x_{\ell m}}(N_{\ell m}^+ + \beta, N_{\ell m}^{\text{tot}} + \beta)}{G_{x_{\ell m}}(\beta, \beta)} \right] \\
 &\quad \times \left[\prod_{\ell=1}^L \frac{\gamma_\ell^{\vartheta-1} (1 - \gamma_\ell)^{\vartheta-1}}{B(\vartheta, \vartheta)} \right] \times \left[\frac{\alpha^L \Gamma(\alpha)}{\Gamma(N + \alpha)} \prod_{\ell=1}^L \Gamma(M_\ell) \right].
 \end{aligned}$$

2.4.3 *Extension to Multigraphs.* We model multigraphs by sharing the cluster assignments z across link types but introducing link-type specific probabilities or rates η and gap parameters γ :

$$p(\{A^{(1)}, A^{(2)}, \dots, A^{(K)}\}, z, \{\eta^{(1)}, \eta^{(2)}, \dots, \eta^{(K)}\}, \{\gamma^{(1)}, \gamma^{(2)}, \dots, \gamma^{(K)}\} | \psi) \\ = \left[\prod_k p(A^{(k)} | z, \eta^{(k)}) p(\ddot{\eta}^{(k)} | \dot{\eta}^{(k)}, \gamma^{(k)}, \beta) p(\dot{\eta}^{(k)} | \beta) p(\gamma^{(k)} | \vartheta) \right] p(z | \alpha).$$

3 Results and Discussion

We analyzed the proposed BCD model on a variety of synthetic and real networks. Unless otherwise stated, we set $\beta = \vartheta = 1$, used $T = 3$ auxiliary components in the Gibbs sampler (Neal, 2000, algorithm 8), set $\alpha = \log N$, and used three rounds of restricted Gibbs sampling in the split-merge sampler. We ran the MCMC sampler for 500 iterations and discarded the first 400 samples for burn-in. The displayed solutions are the highest likelihood sample obtained by the sampler across the 100 samples. By definition, MAP estimation involves taking the modal, or most commonly occurring, sample. However, it is practically certain that every sample will be unique (i.e., every sample has frequency 1), and therefore the highest likelihood sample is used as a proxy for the MAP solution. The displayed graphs are sorted according to the sizes of the extracted clusters.

3.1 Synthetic Data. To examine the properties of BCD, we studied the performance of the model on two synthetic data sets, chosen to highlight key properties of the model also found on real networks.

The first data set was formed by two groups, each containing 20 nodes. Within each of the two groups, half of the nodes were fully connected to each other, while the remaining nodes were completely disconnected from each other but linked to the fully connected core of nodes with a link density of 0.5. These nodes thereby formed so-called satellite communities (Newman & Girvan, 2004) to the main community. Two links were further placed between the core nodes of the two groups bridging the two main groups. Figure 2 shows the result of the analysis for fixed $\gamma = 0.01$ and $\gamma = 1$ in the BCD model, as well as the results of analysis by IRM. IRM extracts four communities: the two strongly connected core groups and the two satellite communities. The clusters found by IRM are not communities in the sense described in definition 1 because the clusters of satellite do not constitute a community with more internal than external links. In fact, they have no internal links whatsoever. In BCD, the satellites are grouped together with the core nodes of each community ($\gamma = 0.01$) or when the prior is chosen to strongly favor many communities with a small gap ($\gamma = 1, \alpha = 250$), each satellite node is assigned to its own community.

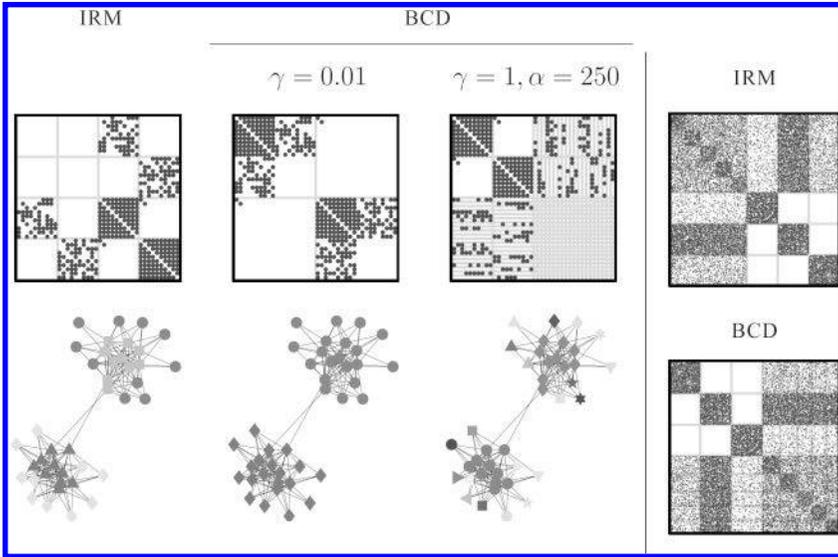


Figure 2: (Left) IRM and BCD analysis of a synthetically generated network of 40 nodes with two groups of 20 nodes each subdivided into 10 nodes forming a core and 10 nodes defining the periphery (i.e., satellite communities). (Right) IRM and BCD analysis of synthetically generated network with five smaller and three larger communities.

Subsequently, we analyzed the performance of IRM and BCD on a network consisting of eight communities (five smaller and three larger), all with an internal link density of 0.6. The five smaller communities have a link probability between them of 0.3 and an identical link probability to the remaining three communities (0.1, 0.55, and 0.2). Figure 2 shows the results for IRM and BCD. Although there are eight communities in the data, the IRM finds only four communities, grouping the five smaller communities together. The reason is that the five smaller communities have identical link probabilities toward the rest of the network. The BCD model correctly identifies the eight communities, since this solution leads to larger within-cluster than between-cluster link probabilities.

3.2 Zachary's Karate Club and the Bottleneck Dolphin Networks. Zachary's Karate Club and the bottleneck dolphins of Doubtful Sound (Lusseau et al., 2003) are often-used benchmark networks for community detection (Newman & Girvan, 2004; Fortunato, 2010). The karate network describes social interactions between members of a karate club that eventually split in two due to a dispute, and the dolphin network consists of relations among 62 dolphins that were observed to split into two large

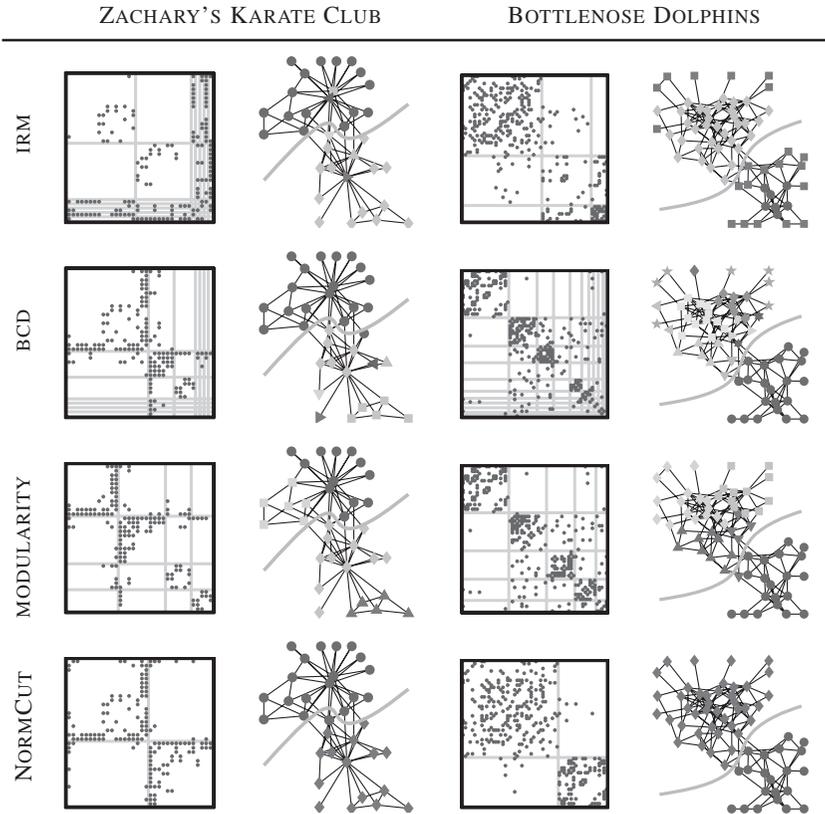


Figure 3: Analysis of the karate and dolphin networks by IRM, BCD, MODULARITY, and NormCut.

groups. We compare the community structure identified by the BCD, IRM, MODULARITY (Newman & Girvan, 2004), and normalized graph cut of Shi and Malik (2000). The normalized graph cut (NormCut) was based on clustering the values of the second smallest eigenvalue of the normalized graph Laplacian into two clusters using k -means. The results of analyzing the networks using IRM, BCD, MODULARITY, and NormCut are shown in Figure 3.

In the karate network, IRM assigns central persons to individual clusters and finds two large clusters of satellites, BCD detects one large community (with one misclassified person) and subdivides the other community into two groups and four satellites, MODULARITY splits each of the two communities into two subcommunities, and the NormCut of Shi and Malik (2000) incorrectly assigns one node to a wrong community.

In the dolphins network, IRM finds two densely connected core clusters but assigns all other nodes to one big satellite cluster, and BCD detects one of the communities (except for one node, which is given a cluster of its own) and splits the other into four subcommunities plus some satellite communities. MODULARITY gives a similar result, correctly identifying one community (except for one node assigned a different community) and splitting the other into four subcommunities, whereas the NormCut identifies a correct split of the network.

For both considered networks, the NormCut works well. We note, however, that it is in general an open problem for spectral approaches such as the considered NormCut to define how many eigenvectors to include in the analysis, as well as the number of clusters to be extracted by k -means (Von Luxburg, 2007). A benefit of the IRM, BCD, and MODULARITY is that the models automatically determine the model orders.

Comparing the discovered clusters with the true communities, it appears that IRM has a tendency to assign a densely connected core and its satellite nodes into separate clusters, whereas the BCD model has a tendency to assign some satellite nodes to separate clusters. Both of these issues can be explained as an unwarranted subdivision of communities containing nodes with largely varying degrees, suggesting that the models could be improved by taking node degree into account. For example, in the BCD analysis of the karate network, one node is incorrectly assigned to the top cluster (dark gray circles) although it has more links to the bottom cluster. This happens because the node has a substantially lower degree than the other nodes in the bottom cluster.

3.3 Link Prediction in Real Networks. Link prediction as measured by the area under curve AUC of the receiver operator characteristic has become a standard for the evaluation of network models (Clauset, Moore, & Newman, 2008; Miller et al., 2009). We compared the performance of the proposed BCD model with prominent variants of the stochastic block model (Nowicki & Snijders, 2001; Kemp et al., 2006; Xu et al., 2006; Hofman & Wiggins, 2008; Mørup et al., 2010). IHW denotes a nonparametric extension of the model in Hofman and Wiggins (2008) where the relational matrix η is defined by two parameters—a within-community parameter ρ_1 and a between-community parameter ρ_0 ; IDW is a model with separate within-community parameters ρ_k and a shared between-community density ρ_0 , as discussed in Mørup et al. (2010); and IRMCB corresponds to the IRM model where the prior is modified to favor communities,

$$\eta_{\ell m} \sim \begin{cases} \text{Beta}(10, 1) & \text{if } l = m \\ \text{Beta}(1, 10) & \text{otherwise} \end{cases}.$$

We analyzed 14 networks, summarized in Table 1, which also shows the number of parameters found by IRM and BCD, the inferred cluster

Table 1: Network Properties and Results of Analysis.

	Network Properties			Clusters			AUC (%)		
	N	N ⁺	Weighted and Directed	IRM	BCD	γ	IRM	BCD	
USAir97	332	2126		15.4(5)	18.6(5)	1.00(0)	95.7(10)	95.9(5)	
USPowerGrid	4941	6594		7.2(3)	34.2(17)	0.04(0)	52.0(16)	77.0(36)	
Football	115	613		10.8(2)	13.0(0)	0.09(0)	88.2(22)	89.3(24)	
Celegans	306	2345	✓	23.2(6)	30.0(4)	1.00(0)	76.2(25)	80.4(82)	
yeast	2361	6646		21.8(7)	32.4(6)	1.00(0)	88.9(4)	88.7(7)	
lesmis	77	254	✓	13.0(3)	15.0(8)	0.97(1)	95.5(8)	94.7(14)	
Geom	7343	11,898	✓	59.4(8)	76.2(17)	0.99(0)	86.3(3)	89.4(8)	
netscience	1589	2742	✓	6.4(4)	10.8(13)	0.91(4)	55.5(24)	67.1(20)	
cond-mat	16,726	47,594	✓	34.0(4)	44.4(6)	1.00(0)	69.6(9)	72.8(6)	
SciMet	3084	10,413		12.2(2)	24.0(6)	1.00(0)	55.3(15)	89.9(6)	
smaGri	1059	4919	✓	7.4(2)	20.0(3)	1.00(0)	54.0(3)	88.7(5)	
smallW	396	994	✓	8.0(0)	12.0(4)	1.00(0)	91.7(24)	97.1(11)	
NIPS	234	598		7.8(3)	30.6(2)	0.02(0)	85.6(45)	94.1(33)	
NIPSCW	2865	4733	✓	19.8(4)	111.1(38)	0.15(13)	87.1(11)	90.0(6)	

Notes: Network properties: number of nodes and links and whether network is weighted and directed; standard deviation on last digit given in parentheses. Results of analysis: number of clusters, value of the shared gap parameter γ , and area under curve (AUC) link prediction scores (%). All networks except the NIPS and NIPSCW networks were obtained from <http://www.cise.ufl.edu/research/sparse/mat/>.

Table 2: Area Under Curve (AUC) [%] Link Prediction Score.

	IHW	IDM	IRM	IRMCB	BCD Shared γ	BCD Separate γ
USAir97	75.0(23)	82.4(16)	95.7(10)	95.1(9)	95.9(5)	95.4(4)
USPowerGrid	76.9(17)	61.9(13)	52.0(16)	51.5(15)	77.0(36)	80.9(18)
Football	83.7(31)	83.3(31)	88.2(22)	88.7(21)	89.3(24)	89.4(25)
Celegans	57.7(17)	54.9(9)	76.2(25)	74.8(15)	80.4(82)	73.6(94)
yeast	68.0(3)	82.5(7)	88.9(4)	88.7(4)	88.7(7)	87.3(4)
lesmis	70.1(55)	82.4(27)	95.5(8)	92.2(23)	94.7(14)	93.8(16)
Geom	66.8(16)	71.1(7)	86.3(3)	86.6(3)	89.4(8)	89.3(4)
netscience	58.7(34)	57.3(14)	55.5(24)	49.4(24)	67.1(20)	65.3(27)
cond-mat	58.9(9)	61.5(8)	69.6(9)	66.5(5)	72.8(6)	73.4(8)
SciMet	75.1(12)	82.8(9)	55.3(15)	63.4(16)	89.9(6)	89.2(4)
smaGri	75.2(16)	83.4(8)	54.0(3)	65.4(17)	88.7(5)	88.2(7)
smallW	84.7(9)	92.4(18)	91.7(24)	90.9(25)	97.1(11)	97.5(8)
NIPS	83.5(39)	87.4(46)	85.6(45)	88.0(34)	94.1(33)	94.0(19)
NIPSCW	81.7(25)	85.2(13)	87.1(11)	86.0(10)	90.0(6)	91.3(2)

Note: Standard deviation on last digit given in parentheses.

gap, as well as AUC scores of link prediction. AUC scores were computed by treating 2.5% of all the links and an equivalent number of nonlinks as missing for prediction. The values shown are for five random initializations, each with different random sets of link and nonlinks left out for prediction. This framework is in line with sampling a fixed fraction of dyads in the graph for prediction (Kok & Domingos, 2007; Miller et al., 2009; Menon & Elkan, 2011) as the AUC is invariant to the relative sizes of the link and nonlink classes. However, an issue with these frameworks for sampling links is that they tend to deweight errors on the periphery as this region of the graph does not contain many edges relative to its size. This potential bias is the same for the models considered, and we therefore use the applied link prediction approach primarily to evaluate the relative performance of the considered models.

In none of the experiments IHW, IDM, or IRMCB perform best (quantitative results of these analyses are included in Table 2). The proposed BCD model outperformed all other models for the majority of the considered networks. We also estimated models with community-specific parameters (separate values of γ); however, this gave similar results to using a shared γ parameter (see also Table 2).

In Table 1, the performance of the IRM and BCD model with shared γ parameter is given. Of the 14 considered networks the USPowerGrid, Football, NIPS and NIPSCW had a large comparative difference in within-community and between-community link density (i.e., $\gamma < 0.2$). The NIPS network is a binary graph of the top 234 collaborating NIPS authors in NIPS volumes 1 to 17 (also considered in Miller et al., 2009). The NIPSCW network is the complete integer weighted network of all NIPS authors in

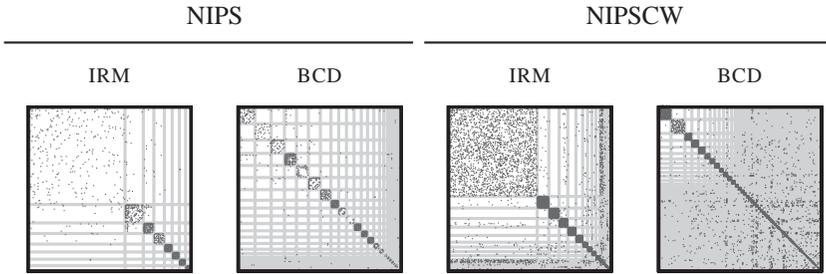


Figure 4: Result of an IRM and BCD analysis of the small binary NIPS network of the 234 most collaborating authors and the complete weighted NIPS network including all authors as well as the number of articles that authors have collaborated on.

volumes 1 to 17 compiled by Globerson, Chechik, Pereira, and Tishby (2005), where weights indicate the number of papers authors have collaborated on. Interestingly, the analysis indicates that of all the networks considered, the NIPS author collaboration exhibits the strongest comparative gap between internal and external link densities— $\gamma = 0.02$ for NIPS—while a substantial gap of $\gamma = 0.15$ is found for NIPSCW. Figure 4 gives an example of the results obtained when analyzing the small-sized and full NIPS author collaboration networks. From the figure, it can be seen that the IRM model extracts a large low-density community, whereas the BCD model subdivides this large group into smaller community-structured groups similar to the effect demonstrated on the synthetic network example in Figure 2. An inspection of the communities shows that the extracted communities of BCD correspond well with known machine learning research communities (see the appendix). While the IRM model defines a large-periphery class, the BCD accounts for the network structure in terms of communities. Both are valid accounts and explain different types of network structure. If the aim is to account for structure, the model of choice should be the one that performs the best in terms of link prediction; if the aim is to separate the nodes into strict communities, the BCD model is more appropriate than the IRM; and if the aim is a flexible model that uses few parameters, the IRM model is in general more favorable than the proposed BCD.

For all 14 networks in Table 1, it can be seen that the BCD model extracts a larger number of communities than the IRM. To investigate if this is a general property of the BCD, we generated three data sets: the first according to the IRM model without community structure, the second according to the BCD model, and the third according to an Erdős-Rényi random graph (corresponding to the IRM and BCD model with one giant community). In the analysis of all three data sets, the IRM model resulted in substantially fewer clusters than the proposed BCD model. For the data set generated according to the IRM model without community structure, the BCD

generated a large number of small clusters that combined nodes from IRM clusters with strong between-cluster link density or split existing clusters into smaller components. From these results, it is evident when inspecting the permuted graphs that the clusters extracted by BCD do not form valid communities according to definition 1. For the data generated according to the BCD model, the IRM merged many of the relatively small or low-density communities together. In the analysis of the Erdős-Rényi random graph, the BCD model was able to identify community structure emerging at random. When evaluating the models' performance in terms of their ability to predict links, it was observed that the model that generated the data outperformed the alternative model, while both models performed no better than random guessing on the Erdős-Rényi random graph. This indicates that link prediction can be used to quantify the better model (the three experiments are given in *demo3.m*, *demo4.m* and *demo5* in the accompanying Matlab toolbox; Mørup & Schmidt, 2011).

Table 3 shows the average per iteration CPU time for each of the methods. A benefit of the IHW, IDM, IRM, and BCD is that they scale in the number of edges rather than the size of the network. All the methods rely on the sufficient statistics N_{cm}^+ and N_{cm}^- that can be efficiently computed considering only the edge list of the graph rather than the full adjacency matrix of size N^2 . From the CPU times, it can be seen that the BCD in general is about 10 times slower than the other approaches. Part of this additional computational cost can be attributed to the evaluation of the log of the normalized incomplete beta and gamma functions based on Matlab's *betainc.m* and *gammainc.m* functions that invokes about twice the cost of the regular *betaln.m* and *gammaln.m* functions. The remaining additional cost we attribute to the computational cost invoked by nonconjugate split merge sampling of \mathbf{z} , as well as a sampling of $\boldsymbol{\eta}$ and $\boldsymbol{\gamma}$.

We finally compared the proposed BCD model to the mixed membership stochastic block model (MMSB) (Airoldi, Blei, Fienberg, & Xing, 2008) and latent feature relational model (LFRM) (Miller et al., 2009) on the same data sets and link-prediction setup considered in Miller et al. (2009) (see Table 4). The number of components in the MMSB model can be estimated based on maximizing the likelihood on holdout data (Airoldi et al., 2008). BCD is the best-performing model on the Alyawarra data set while being on par with the best-performing LFRM model for the NIPS data. It is also better performing than the Bayesian clustered tensor factorization model of Sutskever, Salakhutdinov, and Tenenbaum (2009) having an AUC of 0.90 and the multiple relational clusterings of Kok and Domingos (2007) resulting in an AUC score of approximately 0.85. However, the model does not perform well on the countries data. We attribute this to the data not being well described by community structure, since some relations are positive (e.g., *military alliance with* and *exports to*) while other are negative (e.g., *protests* and *negative communications*) (Kemp et al., 2006) which does not comply with the notion of community structure assumed in BCD.

Table 3: Average per Iteration CPU-Time.

	IHW	IDM	IRM	IRMCB	BCD Shared γ	BCD Separate γ
USAir97	0.33(5)	0.43(2)	0.35(3)	0.39(3)	2.69(18)	3.44(24)
USPowerGrid	2.18(14)	6.85(56)	6.32(8)	7.43(55)	69.38(526)	70.73(288)
Football	0.05(0)	0.06(0)	0.07(0)	0.08(0)	0.78(5)	1.22(9)
Celegans	0.14(0)	0.18(0)	0.27(0)	0.4(3)	4.89(33)	5.82(38)
yeast	1.45(15)	3.16(27)	3.95(25)	6.72(62)	35.13(116)	41.03(150)
Iesmis	0.04(0)	0.05(0)	0.06(0)	0.06(0)	0.58(4)	0.82(6)
Geom	14.23(263)	11.61(152)	22.75(322)	27.85(318)	304.42(832)	407.61(3582)
netscience	1.07(30)	2.28(15)	2.09(12)	2.41(7)	16.62(391)	24.23(219)
cond-mat	48.94(1360)	47.2(1413)	34.27(46)	57.68(1079)	422.49(5737)	901.53(7856)
SciMet	3.89(84)	3.35(5)	4.34(38)	6.04(46)	67.91(363)	79.97(362)
smaGri	1.26(22)	1.46(6)	1.89(14)	2.07(16)	15.71(108)	17.14(132)
smallW	0.41(7)	0.57(5)	0.58(4)	0.64(3)	4.52(28)	4.86(41)
NIPS	0.09(1)	0.14(1)	0.21(0)	0.25(1)	1.96(10)	2.93(17)
NIPSCW	1.98(29)	2.13(20)	3.05(14)	3.6(18)	72.89(1046)	45.23(261)

Note: The standard deviation is on the last digit given in parentheses.

Table 4: Area Under Curve (AUC) (%) Link Prediction Results.

	LFRM			BCD		
	IRM Initialization	Random Initialization	IRM	MMSB	Shared Gap	Separate Gap
Countries (global)	87.1(10)	70.7(53)	85.0(3)	86.4(8)	83.7(7)	78.2(7)
Alyawarra (global)	91.8(11)	71.3(30)	89.4(30)	91.4(10)	93.7(2)	93.5(1)
NIPS	95.1(13)	94.7(13)	89.1(13)	87.1(13)	91.2(6)	93.8(5)

Notes: Nonparametric latent feature relational model (LFRM) (Miller et al., 2008), infinite relational model (IRM) (Kemp et al., 2006), mixed membership stochastic block model (MMSB) (Airoldi et al., 2008), and the proposed infinite community model (BCD). Standard deviation on last digit given in parentheses.

4 Conclusion

Many networks of scientific interest naturally decompose into clusters or communities. In this letter, we proposed the BCD model to explicitly model community structure in networks and demonstrated that modeling community structure improves on link prediction compared to the IRM and related nonparametric models of networks. Our analysis supports the observation that community structure is an important structural property of networks (Fortunato, 2010) that, as currently demonstrated, can be incorporated into the nonparametric Bayesian modeling of networks. The proposed BCD model trivially extends to other Bayesian generative models such as the multiple-membership modeling of Miller et al. (2009) and Mørup et al. (2010) as well as the degree-corrected stochastic block model of Karrer and Newman (2011). A Matlab toolbox with the proposed inference procedure is available for download (Mørup & Schmidt, 2011).

Appendix A: NIPS Author Communities

Table 5 lists the extracted groups by the IRM and BCD for the small binary NIPS network of the 234 top collaborating NIPS authors. From the extracted groups, it can be seen that the large first cluster of the IRM model has been split into several clusters by the BCD (clusters 2, 3, 9, 10, 15–28) representing dense subcommunities of authors who collaborate closely. The remaining smaller clusters extracted by IRM are very similar to the clusters defined by BCD: IRM cluster 2 \sim BCD cluster 1, IRM cluster 3 \sim BCD cluster 4, IRM cluster 4 \sim BCD cluster 7, IRM cluster 5 \sim BCD cluster 8, IRM cluster 6 \sim BCD cluster 11, IRM cluster 7 \sim BCD cluster 13, and IRM cluster 8 \sim BCD cluster 12. When generating synthetic data according to the BCD model, a similar effect is observed: the IRM merges some of the small or low-density communities into large low-density clusters, whereas the BCD

Table 5: NIPS Small Network Communities.

IRM	BCD
<p>Cluster 1: Allen R, Alspecter J, Amari S, Atlas L, Baldi P, Ballard D, Bartlett M, Barto A, Bengio S, Bengio Y, Bialek W, Bienenstock E, Bishop C, Black M J, Bourlard H, Bower J, Brunak S, Buhmann J, Cauwenberghs G, Chauvin Y, Chechik G, Chen H, Cohen M, Cole R, Coolen A, Cottrell G, Cowan J, Crammer K, Dayan P, De Ruyter Van Steveninck R, Donoghue J, Doucet A, Doya K, Dreyfus G, Dugas C, Edelman S, Etienne-Cummings R, Franco H, Freeman W, Frey B J, Friedmann N, Ghahramani Z, Giles C, Goodman R, Gordon G, Hansen L, Hanson S, Hastie T, Herbrich R, Hertz J, Hinton G, Horn D, Horne B, Intrator N, Jabri M, Johnson D, Jordan M, Jung T, Jaakkola T, Kawato M, Kearns M, Keeler J, Koller D, Krogh A, Lee D, Lee T, Lee Y, Leen T, Littlewort G, MacKay D, Makeig S, Marks R, Martin G, Mayhew J, Meila M, Meir R, Mjolsness E, Moody J, Morgan N, Movellan J R, Movellan J, Mozer M C, Mozer M, Murray A, Murray-Smith R, Maass W, Nadal J, Ng A Y, Nowlan S, Obermayer K, Ohmi T, Oppen M, Parr R, Pawelzik K, Pearlmutter B, Perona P, Personnaz L, Pouget A, Precup D, Ritter H, Roth D, Roweis S, Ruderman D, Rumelhart D, Ruppert E, Russell S, Schapire R, Sejnowski T, Seung H, Shibata T, Siegelmann H, Singer Y, Singh S, Smyth P, Spence C, Sutton R, Saad D, Tarassenko L, Tenenbaum J, Tishby N, Touretzky D, Tresp V, VandenBout D, Vincent P, Viola P, Vaadia E, Wang X, Weinschall D, Weiss Y, Wiles J, Williams C, Winther O, Wolniewicz R, Wolpert D, Yamashita T, Yang H, Yuille A, Zemel R, Ziehe A.</p>	<p>Cluster 1: Bartlett P, Bousquet O, Chapello O, Cristianini N, Elisseeff A, Herbrich R, Kowalczyk A, Logothetis N, Mika S, Mukherjee S, Muller K, Platt J, Poggio T, Pontil M, Rasmussen C, Ratsch G, Scholkopf B, Scholz M, Shawe-Taylor J, Smola A J, Smola A, Warmuth M K, Weston J, Williamson R, Zhou D.</p> <p>Cluster 2: Amari S, Chechik G, Coolen A, Doucet A, Edelman S, Hansen L, Horn D, Intrator N, Leen T, Moody J, Oppen M, Pawelzik K, Roth D, Ruppert E, Siegelmann H, Spence C, Saad D, Tresp V, Wang X, Weinschall D, Winther O, Yang H, Ziehe A.</p> <p>Cluster 3: Barto A, Bishop C, Freeman W, Frey B J, Ghahramani Z, Jordan M, Jaakkola T, Kawato M, Kearns M, Meila M, Ng A Y, Precup D, Roweis S, Russell S, Singh S, Sutton R, Tenenbaum J, Weiss Y, Wolpert D.</p> <p>Cluster 4: Baird H, Boser B, Bottou L, Burges C, Denker J, Gardner W, Graf H, Guyon I, Henderson D, Howard R, Hubbard W, Jackel L, LeCun Y, Sackinger E, Simard P, Solla S, Vapnik V.</p> <p>Cluster 5: Bialek W, Chen H, Cottrell G, Crammer K, De Ruyter Van Steveninck R, Friedmann N, Giles C, Horne B, Jabri M, Koller D, Lee Y, Parr R, Ruderman D, Schapire R, Singer Y, Tishby N, Vaadia E.</p> <p>Cluster 6: Bartlett M, Dayan P, Doya K, Hinton G, Jung T, Lee T, Littlewort G, Makeig S, Movellan J R, Movellan J, Mozer M, Nowlan S, Pouget A, Sejnowski T, Viola P, Williams C, Zemel R.</p> <p>Cluster 7: Bair W, DeWeerth S, Douglas R, Harris J, Horiuchi T, Koch C, Lazzaro J, Lippmann R, Liu S, Luo J, Mahowald M, Mead C, Moore A, Sivilotti M, Wawrzynek J.</p>

Table 5: (Continued.)

IRM	BCD
<p>Cluster 2: Bartlett P, Bousquet O, Burges C, Chapelle O, Cristianini N, Elisseeff A, Kowalczyk A, Logothetis N, Mika S, Mukherjee S, Muller K, Platt J, Poggio T, Pontil M, Rasmussen C, Ratsch G, Scholkopf B, Scholz M, Shawe-Taylor J, Smola A J, Smola A, Vapnik V, Warmuth M K, Weston J, Williamson R, Zhou D.</p>	<p>Cluster 8: Baker C, Burgard W, Ferguson D, Hahnel D, Morris A, Omohundro Z, Reverte C, Thayer S, Thrun S, Whittaker C, Whittaker W.</p> <p>Cluster 9: Baldi P, Ballard D, Brunak S, Chauvin Y, Goodman R, Hertz J, Krogh A, Mjolsness E, Perona P, Smyth P, Yuille A.</p>
<p>Cluster 3: Baird H, Boser B, Bottou L, Denker J, Gardner W, Graf H, Guyon I, Henderson D, Howard R, Hubbard W, Jackel L, LeCun Y, Sackinger E, Simard P, Solla S.</p>	<p>Cluster 10: Bower J, Buhmann J, Hanson S, Murray A, Murray-Smith R, Maass W, Ohmi T, Ritter H, Shibata T, Tarassenko L, Yamashita T.</p>
<p>Cluster 4: Bair W, DeWeerth S, Douglas R, Harris J, Horiuchi T, Koch C, Lazzaro J, Lippmann R, Liu S, Luo J, Mahowald M, Mead C, Moore A, Sivilotti M, Wawrzynek J.</p>	<p>Cluster 11: Boonyanit K, Burr J, Kritayakirana K, Leung M, Murray M, Peterson A, Schwartz E, Stork D, Watanabe T, Wolff G.</p>
<p>Cluster 5: Baker C, Burgard W, Ferguson D, Hahnel D, Morris A, Omohundro Z, Reverte C, Thayer S, Thrun S, Whittaker C, Whittaker W.</p>	<p>Cluster 12: Jain A, McNair A, Osterholtz L, Saito H, Schmidbauer O, Sloboda T, Tebelskis J, Waibel A, Woszczyna M.</p>
<p>Cluster 6: Boonyanit K, Burr J, Kritayakirana K, Leung M, Murray M, Peterson A, Schwartz E, Stork D, Watanabe T, Wolff G.</p>	<p>Cluster 13: Blackman D, Chiu T, Clare T, Dao J, Donham C, Hsieh T, Loiaz M, Mueller P, Vander Spiegel J.</p>
<p>Cluster 7: Blackman D, Chiu T, Clare T, Dao J, Donham C, Hsieh T, Loiaz M, Mueller P, Vander Spiegel J.</p>	<p>Cluster 14: Bourlard H, Cohen M, Franco H, Johnson D, Keeler J, Martin G, Morgan N, Rumelhart D.</p>
<p>Cluster 8: Jain A, McNair A, Osterholtz L, Saito H, Schmidbauer O, Sloboda T, Tebelskis J, Waibel A, Woszczyna M.</p>	<p>Cluster 15: Allen R, Alspector J, Bienenstock E, Black M J, Donoghue J, Meir R.</p>
	<p>Cluster 16: Cauwenberghs G, Cowan J, Lee D, MacKay D, Seung H.</p>
	<p>Cluster 17: Dreyfus G, Nadal J, Personnaz L.</p>
	<p>Cluster 18: Atlas L, Cole R, Marks R.</p>
	<p>Cluster 19: Bengio Y, Dugas C, Vincent P.</p>
	<p>Cluster 20: Gordon G, Touretzky D.</p>
	<p>Cluster 21: Obermayer K, Mayhew J.</p>
	<p>Cluster 22: Mozer M C, Wolniewicz R.</p>
	<p>Cluster 23: Etienne-Cummings R.</p>
	<p>Cluster 24: VandenBout D.</p>
	<p>Cluster 25: Bengio S.</p>
	<p>Cluster 26: PearlMutter B.</p>
	<p>Cluster 27: Wiles J.</p>
	<p>Cluster 28: Hastie T.</p>

model correctly identifies the generated communities (see also *demo.4* in the accompanying Matlab toolbox: Mørup & Schmidt, 2011).

Acknowledgments

We thank the two anonymous reviewers for their very useful comments and suggestions to improve this letter.

References

- Airoldi, E. M., Blei, D. M., Fienberg, S. E., & Xing, E. P. (2008). Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.*, 9, 1981–2014.
- Clauset, A., Moore, C., & Newman, M. E. J. (2008). Hierarchical structure and the prediction of missing links in networks. *Nature*, 453(7191), 98–101.
- Dutka, J. (1981). The incomplete beta function: A historical profile. *Archive for History of Exact Sciences*, 24, 11–29.
- Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3–5), 75–174.
- Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12), 7821–7826.
- Globerson, A., Chechik, G., Pereira, F., & Tishby, N. (2005). Euclidean embedding of co-occurrence data. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in neural information processing systems*, 17 (pp. 497–504). Cambridge, MA: MIT Press.
- Gulbahce, N., & Lehmann, S. (2008). The art of community detection. *BioEssays*, 30(10), 934–938.
- Hofman, J. M., & Wiggins, C. H. (2008). Bayesian approach to network modularity. *Physical Review Letters*, 100(25), 258701.
- Jain, S., & Neal, R. M. (2004). A split-merge Markov chain Monte Carlo procedure for the Dirichlet process mixture model. *Journal of Computational and Graphical Statistics*, 13(1), 158–182.
- Karrer, B., Levina, E., & Newman, M. E. J. (2008). Robustness of community structure in networks. *Physical Review E*, 77(4), 1–10.
- Karrer, B., & Newman, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1), 016107.
- Kemp, C., Tenenbaum, J. B., Griffiths, T. L., Yamada, T., & Ueda, N. (2006). Learning systems of concepts with an infinite relational model. In *Proceedings of the National AAAI Conference on Artificial Intelligence*. Palo Alto, CA: AAAI Press.
- Kok, S., & Domingos, P. (2007). Statistical predicate invention. In *Proceedings of the Twenty-Fourth International Conference on Machine Learning*. Palo Alto, CA: AAAI Press.
- Lusseau, D., Schneider, K., Boisseau, O. J., Haase, P., Slooten, E., & Dawson, S. M. (2003). The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. *Behavioral Ecology and Sociobiology*, 54(4), 396–405.

- Menon, A., & Elkan, C. (2011). Link prediction via matrix factorization. In *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 437–452). Berlin: Springer-Verlag.
- Miller, K. T., Griffiths, T. L., & Jordan, M. I. (2009). Nonparametric latent feature models for link prediction. In Y. Bengio, D. Schuurmans, J. Lafferty, C.K.I. Williams, & A. Culotta (Eds.), *Advances in neural information processing systems*, 22 (pp. 1276–1284). Red Hook, NY: Curran Associates.
- Mørup, M., & Schmidt, M. N. (2011). *Matlab code for Bayesian community detection*. Available online at: http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=6147.
- Mørup, M., Schmidt, M. N., & Hansen, L. K. (2010). Infinite multiple membership relational modeling for complex networks. In *IEEE International Workshop on Machine Learning for Signal Processing*. Piscataway, NJ: IEEE.
- Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9, 249–265.
- Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E*, 69(2), 026113-1-15.
- Nowicki, K., & Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455), 1077–1087.
- Pitman, J. (2006). *Combinatorial stochastic processes*. Berlin: Springer-Verlag.
- Rosvall, M., & Bergstrom, C. T. (2007). An information-theoretic framework for resolving community structure in complex networks. *Proceedings of the National Academy of Sciences*, 104(18), 7327–7331.
- Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), 888–905.
- Sun, S., Ling, L., Zhang, N., Li, G., & Chen, R. (2003). Topological structure analysis of the protein-protein interaction network in budding yeast. *Nucleic Acids Research*, 31(9), 2443–2450.
- Sutskever, I., Salakhutdinov, R., & Tenenbaum, J. (2009). Modelling relational data using Bayesian clustered tensor factorization. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), *Advances in neural information processing systems*, 22 (pp. 1821–1828). Red Hook, NY: Curran Associates.
- Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, 17(4), 395–416.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of “small-world” networks. *Nature*, 393(6684), 440–442.
- Weiler, H. (1965). The use of incomplete beta functions for prior distributions in binomial sampling. *Technometrics*, 7, 335–347.
- Xu, Z., Tresp, V., Yu, K., & Kriegel, H. P. (2006). Learning infinite hidden relational models. In *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence*. AUAI Press.

This article has been cited by:

1. Kasper Winther Andersen, Kristoffer H. Madsen, Hartwig Roman Siebner, Mikkel N. Schmidt, Morten Mørup, Lars Kai Hansen. 2014. Non-parametric Bayesian graph models reveal community structure in resting state fMRI. *NeuroImage* **100**, 301-315. [[CrossRef](#)]
2. Farshad Nourbakhsh, Samuel Rota Bulò, Marcello Pelillo. 2014. A matrix factorization approach to graph compression with partial information. *International Journal of Machine Learning and Cybernetics* . [[CrossRef](#)]