

# Bayesian Community Detection

**Morten Mørup and Mikkel N. Schmidt**<sup>1</sup>

<sup>1</sup>Section for Cognitive Systems, Technical University of Denmark

Asmussens Alle, Bygning 321, 2800 Kgs. Lyngby

{mm,mns}@imm.dtu.dk.

**Keywords:** Community detection, Complex networks, Infinite relational Model, Stochastic Block-model, Modularity, Normalized Cut.

## Abstract

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 non-parametric 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 modelled.

# 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 [5, 25, 28]. Many networks naturally decompose into clusters or communities characterized by [4]:

**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 [10, 7, 4]. 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 [10, 4]. A large variety of methods for community detection have been proposed over the years; however, no standard have been agreed upon [7]. A widely used approach is to optimize some measure of cluster structure such as modularity [20] or mutual information [23]. 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 blockmodel [21] and its extensions [11, 8, 16, 17]. Notably, the infinite relational model (IRM) [12, 30] is a non-parametric extension that allows the number of communities to be inferred from data. In these models, clusters are sets of nodes with homogenous 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 which 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 def. 1.

In this paper we propose a non-parametric Bayesian generative model of networks with community structure, that operationalizes the intuitive definition of communities in def. 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 less 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 [18].

## 2 Methods

Given a network, we address the problem of partitioning the nodes into clusters, consistent with the notion of communities in def. 1. To simplify the presentation, we consider an undirected network represented by a binary, strictly upper triangular adjacency matrix  $\mathbf{A}$ , 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 multi-graphs.

To initiate the discussion, we describe the *infinite relational model* (IRM) [12, 30] 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 Chinese restaurant process 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 probability proportional to how many existing customers are placed at the table and at a new table with a probability proportional to  $\alpha$ . Customers thereby 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 the customers arrive does not influence the probability of the partition[22]. Next, link probabilities are generated which 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.

$$z \sim \text{CRP}(\alpha), \quad \text{Cluster assignment}, \quad (1)$$

$$\eta_{\ell m} \sim \text{Beta}(\beta, \beta), \quad \text{Link probability}, \quad (2)$$

$$A_{ij} \sim \text{Bernoulli}(\eta_{z_i z_j}), \quad \text{Link}. \quad (3)$$

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 def. 1. A cluster  $\ell$  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  $\ell$  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 paper, we propose a non-parametric 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*,  $\gamma_\ell$ , 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  $\ell$  and  $m$  are 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.

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

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 [29] and aspects regarding numerical evaluation can be found in [3]. 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 are at least  $\gamma_\ell$  times less compared to the probability of links within them. In addition to the model where each cluster has a separate  $\gamma_\ell$ , we also consider the special case where all clusters share a common  $\gamma_1 = \gamma_2 = \dots = \gamma$ . Fig. 1 shows random networks generated according to the IRM and the proposed BCD model with shared  $\gamma$ . As  $\gamma$  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 def. 1 we require  $\gamma \leq 1$ .

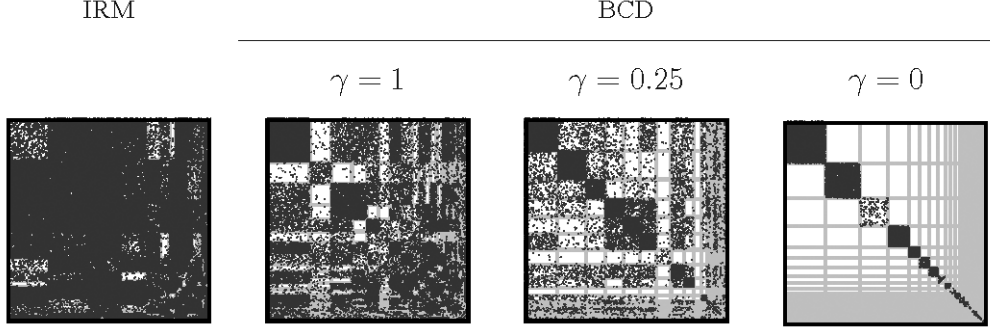


Figure 1: Examples of random graphs according to IRM and BCD for different values of  $\gamma$ .

## 2.2 Inference in BCD

Let  $\dot{\eta} = \{\eta_{\ell m} | \ell = m\}$  and  $\ddot{\eta} = \{\eta_{\ell m} | \ell \neq m\}$  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,  $p(\mathbf{z} | \mathbf{A})$ , which entails integrating over the joint distribution given by

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

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

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

## 2.1 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 ( $\dot{\eta} = \{\eta_{\ell m} | \ell \neq m\}$ ), can be marginalized analytically. I.e. we have

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

As we can not integrate the posterior with respect to the remaining parameters,  $\mathbf{z}$ ,  $\dot{\eta}$ , and  $\gamma$ , we sample from their posterior distribution using MCMC.

## 2.2 Within-cluster link probabilities

Eliminating terms that do not depend on  $\eta_{\ell\ell}$ , the marginal posterior reduces to

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

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

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

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.

### 2.3 Cluster assignment variables

Regarding,  $\mathbf{z}$ , 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. As we are not able to analytically integrate  $\dot{\boldsymbol{\eta}}$  we have to resort to non-conjugate sampling approaches. We use two MCMC transition kernels: an auxiliary variable Gibbs sampler [19, Algorithm 8], and a split-merge Metropolis-Hastings move [9], to sample from  $p(\mathbf{z}|\mathbf{A}, \dot{\boldsymbol{\eta}}, \boldsymbol{\gamma}, \boldsymbol{\psi})$ . In the auxiliary variable Gibbs sampler  $T$  new within community densities and gap parameters are drawn from their priors. These auxiliary variables represents 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 [19, Algorithm 8] for details. Similarly, the split-merge Metropolis-Hastings algorithm propose 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 described in [9].

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

$$p(z_i = \ell | \mathbf{A}, \mathbf{z} \setminus z_i, \dot{\boldsymbol{\eta}}, \boldsymbol{\gamma}, \boldsymbol{\psi}) \propto \eta_{\ell\ell}^{n_{i\ell}^+} (1 - \eta_{\ell\ell})^{n_{i\ell}^-} \alpha^L M_\ell \cdot \prod_{m \neq \ell} \frac{B_{x_{\ell m}}(N_{\ell m}^+ + n_{im}^+ + \beta, N_{\ell m}^- + n_{im}^- + \beta)}{B_{x_{\ell m}}(N_{\ell m}^+ + \beta, N_{\ell m}^- + \beta)}, \quad (4)$$

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

### 2.4 Cluster gaps

Eliminating terms that do not depend on  $\gamma_\ell$ , the posterior reduces to

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



To generate samples from  $\gamma$ , 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  $\gamma_\ell$  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 is a fully observed network; however, the model can easily be adapted to accommodate partially observed network. When it is unknown whether there is a link or not between some set of pairs of nodes in the network, these are simply not included in the evaluation of the likelihood ( $p(\mathbf{A} | \mathbf{z}, \boldsymbol{\eta})$ ), 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,  $p(A_{ij} | \mathbf{A})$ , where  $\mathbf{A}$  denotes the observed part of the network. Based on  $R$  posterior samples, the predictive distribution is approximated as

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

where

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

### 2.4 Extensions of the BCD model

BCD can trivially be extended to handle *self-links*, in which case the adjacency matrix is non-zero 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 links types and introducing different link probability and gap parameters for each link type. The details of all these possible extensions are described below.

## 2.1 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

$$z \sim \text{CRP}(\alpha), \quad \text{Cluster assignment}, \quad (5)$$

$$\eta_{\ell m} \sim \text{Gamma}(\beta, \beta), \quad \text{Link probability}, \quad (6)$$

$$A_{ij} \sim \text{Poisson}(\eta_{z_i z_j}), \quad \text{Link}. \quad (7)$$

and for the BCD

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

$$\eta_{\ell\ell} \sim \text{Gamma}(\beta, \beta), \quad \text{With.-clust. link prob.},$$

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

$$\eta_{\ell m} \sim \text{GammaInc}(\beta, \beta, x_{lm}), \quad \text{Betw.-clust. link prob.},$$

$$x_{lm} = \min(\gamma_\ell \eta_{\ell\ell}, \gamma_m \eta_{mm})$$

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

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  $\eta_{lm}$  we obtain

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

## 2.2 Extension to directed graphs

For directed graphs we have for the marginalized likelihood given above for integer weighted graphs

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

## 2.3 Extension to multi-graphs

We model multi-graphs by sharing the cluster assignments  $\mathbf{z}$  across links types but introducing link type specific probabilities/rates  $\boldsymbol{\eta}$  and gap parameters  $\boldsymbol{\gamma}$ , i.e.

$$\begin{aligned}
p(\{\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \dots, \mathbf{A}^{(K)}\}, \mathbf{z}, \{\dot{\boldsymbol{\eta}}^{(1)}, \dot{\boldsymbol{\eta}}^{(2)}, \dots, \dot{\boldsymbol{\eta}}^{(K)}\}, \{\boldsymbol{\gamma}^{(1)}, \boldsymbol{\gamma}^{(2)}, \dots, \boldsymbol{\gamma}^{(K)}\} | \boldsymbol{\psi}) &= \\
&= \left[ \prod_k p(\mathbf{A}^{(k)} | \mathbf{z}, \boldsymbol{\eta}^{(k)}) p(\dot{\boldsymbol{\eta}}^{(k)} | \dot{\boldsymbol{\eta}}^{(k)}, \boldsymbol{\gamma}^{(k)}, \beta) p(\boldsymbol{\eta}^{(k)} | \beta) p(\boldsymbol{\gamma}^{(k)} | \vartheta) \right] p(\mathbf{z} | \alpha)
\end{aligned}$$

### 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 [19, Algorithm 8], set  $\alpha = \log N$ , and used 3 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 datasets, chosen to highlight key properties of the model also found on real networks.

The first dataset 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 [20] to the main community. Two links were further placed between the core nodes of the two groups bridging the two main groups. Fig. 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 as well the two satellite communities. The clusters found by IRM are not communities in the sense described in def. 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

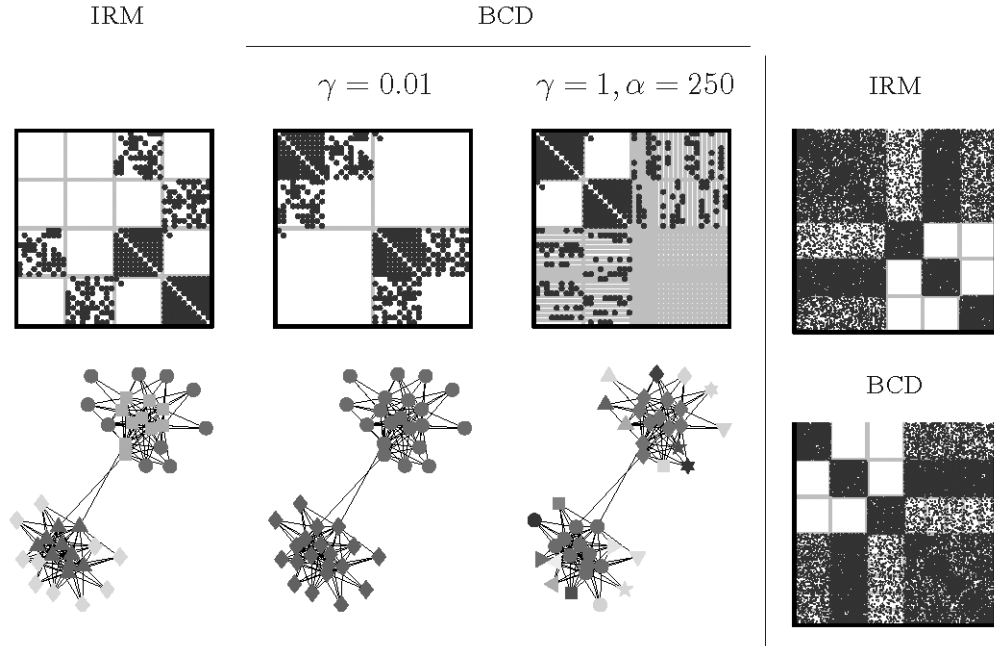


Figure 2: Left: IRM and BCD analysis of a synthetically generated network of 40 nodes with two groups of 20 nodes each subdivided into ten nodes forming a core and ten nodes defining the periphery (i.e. satellite communities). Right: IRM and BCD analysis of synthetically generated network with five smaller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

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.

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 identical link probability to the remaining three communities (0.1, 0.55 and 0.2). Fig. 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 towards 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 Bottlenose Dolphin networks

Zachary’s Karate Club and the bottlenose dolphins of Doubtful Sound [14] are often-used benchmark networks for community detection [20, 4]. The karate network describes social interactions between members of a karate club which eventually split in two due to a dispute, and the dolphin network consist of relations between 62 dolphins that were observed to split into two large groups. We compare the community structure identified by the BCD, IRM, MODULARITY [20] and normalized graph cut of [24]. 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 fig. 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 sub-communities, and the NORMCUT of [24] 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, BCD detects one of the communities (except for one node that is given a cluster of its own) and splits the other into four sub-communities 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 sub-communities whereas the NORMCUT correctly splits the Bottlenose dolphin network while the methods

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 [27]. A benefit of the IRM, BCD and MODULARITY is that the models automatically determine the model orders.

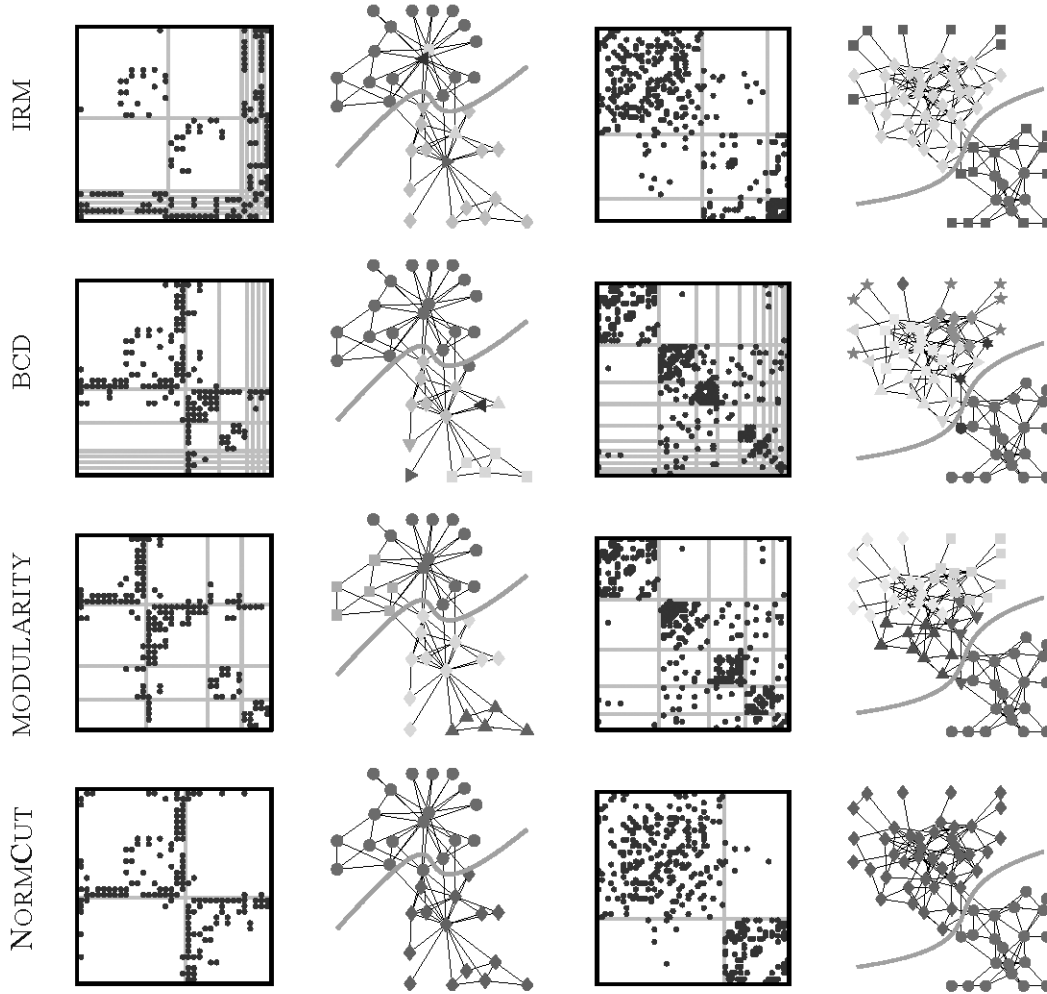


Figure 3: Analysis of the karate and dolphin networks by IRM, BCD, MODULARITY and NORMCUT.

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 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 (red 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 [2, 16]. We compared the performance of the proposed BCD model with prominent variants of the stochastic block model [21, 12, 30, 8, 17]. IHW denotes a non-parametric extension of the model in [8] where the relational matrix  $\eta$  is defined by two parameters, a within community parameter  $\rho_1$  and a between community parameter  $\rho_0$ ; IDW is a model with separate within community parameters  $\rho_k$  and a shared between community density  $\rho_0$  as discussed in [17]; 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 tab. 1 which also shows the number of parameters found by IRM and BCD, the inferred cluster 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 non-links as missing for prediction. The shown values are for 5 random initialization each with different random sets of link and non-links left out for prediction. This framework is in line with sampling a fixed fraction of dyads in the graph for prediction [13, 16, 15] as the AUC is invariant to the relative sizes of the link and non-link classes. However, An issue with these frameworks for sampling links is that they tends to de-weight errors on the periphery as this regions of the graph does not contain many edges relative to its size. This potential bias is the same for the considered models 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 performed best (quantitative results of these analyses are included in tab. 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  $\gamma$ ), however this gave similar



Table 1: Network properties (number of nodes and links and whether network is weighted and directed, standard deviation on last digit given in parenthesis.) and results of analysis (number of clusters, value of the shared gap parameter  $\gamma$ , 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/>.

	NETWORK PROPERTIES				CLUSTERS		$\gamma$	AUC [%]	
	$N$	$N^+$	WGH.	DIR.	IRM	BCD		IRM	BCD
USAir97	332	2126			15.4(5)	18.6(5)	1.00(0)	95.7(10)	<b>95.9(5)</b>
USPowerGrid	4941	6594			7.2(3)	34.2(17)	0.04(0)	52.0(16)	<b>77.0(36)</b>
Football	115	613			10.8(2)	13.0(0)	0.09(0)	88.2(22)	<b>89.3(24)</b>
Celegans	306	2345	✓	✓	23.2(6)	30.0(4)	1.00(0)	76.2(25)	<b>80.4(82)</b>
yeast	2361	6646			21.8(7)	32.4(6)	1.00(0)	<b>88.9(4)</b>	88.7(7)
lesmis	77	254	✓		13.0(3)	15.0(8)	0.97(1)	<b>95.5(8)</b>	94.7(14)
Geom	7343	11898	✓		59.4(8)	76.2(17)	0.99(0)	86.3(3)	<b>89.4(8)</b>
netscience	1589	2742	✓		6.4(4)	10.8(13)	0.91(4)	55.5(24)	<b>67.1(20)</b>
cond-mat	16726	47594	✓		34.0(4)	44.4(6)	1.00(0)	69.6(9)	<b>72.8(6)</b>
SciMet	3084	10413		✓	12.2(2)	24.0(6)	1.00(0)	55.3(15)	<b>89.9(6)</b>
smaGri	1059	4919		✓	7.4(2)	20.0(3)	1.00(0)	54.0(3)	<b>88.7(5)</b>
smallW	396	994		✓	8.0(0)	12.0(4)	1.00(0)	91.7(24)	<b>97.1(11)</b>
NIPS	234	598			7.8(3)	30.6(2)	0.02(0)	85.6(45)	<b>94.1(33)</b>
NIPSCW	2865	4733	✓		19.8(4)	111.1(38)	0.15(13)	87.1(11)	<b>90.0(6)</b>

Table 2: Area under curve (AUC) [%] link prediction score (standard deviation on last digit given in parenthesis).

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

results to using a shared  $\gamma$  parameter (see also table 2).

In tab. 1 the performance of the IRM and BCD model with shared  $\gamma$  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 1-17 also considered in [16]. The NIPSCW network is the complete integer weighted network of all NIPS 1-17 authors compiled by [6] 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, i.e.  $\gamma = 0.02$  for NIPS while a substantial gap of  $\gamma = 0.15$  is found for NIPSCW. Fig. 4 gives an example

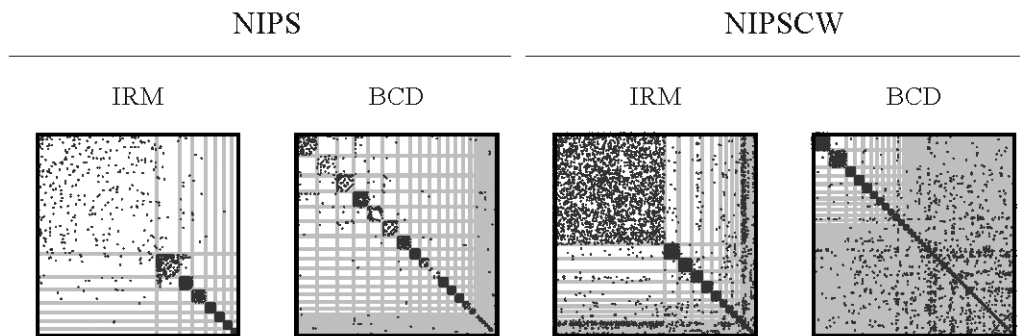


Figure 4: Result of an IRM and BCD analysis of the small binary NIPS network of the 234 most collaborating authors and the complete weighted NIPS network including all authors as well as the number of articles authors have collaborated on.

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 fig. 2. Inspecting the communities the extracted communities of BCD correspond well with known machine learning research communities (see Appendix A). While the IRM model defines a large periphery class the BCD account 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 model that perform 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 whereas if the aim is a flexible model that use few parameters the IRM model is in general more favorable than the proposed BCD.

For all the 14 considered networks in tab. 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 datasets; 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

Table 3: Average pr. iteration cpu-time (standard deviation on last digit given in parenthesis).

	IHW	IDM	IRM	IRMCB	BCD shared $\gamma$	BCD separate $\gamma$
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)
lesmis	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)

with one giant community). In the analysis of all three datasets the IRM model resulted in substantial fewer clusters than the proposed BCD model. For the dataset 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 def. 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 [18]).

Table 4: Area under curve (AUC) [%] link prediction results from [16] for the non-parametric latent feature relational model (LFRM) [16], infinite relational model (IRM) [12], mixed membership stochastic blockmodel (MMSB) [1], and the proposed infinite community model (BCD). (Standard deviation on last digit given in parenthesis).

	LFRM		IRM	MMSB	BCD	
	IRM init.	Random init.			Shared gap	Separate gap
Countries (global)	<b>87.1(10)</b>	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)	<b>93.7(2)</b>	93.5(1)
NIPS	<b>95.1(13)</b>	94.7(13)	89.1(13)	87.1(13)	91.2(6)	93.8(5)

In tab. 3 is given the average pr. iteration cpu-time for each of the methods. A benefit of the IHW, IDM, IRM and BCD are that they scale in the number of edges rather than the size of the network. I.e. all the methods rely on the sufficient statistics  $N_{\ell m}^+$  and  $N_{\ell m}^-$  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 non-conjugate split merge sampling of  $\mathbf{Z}$ , as well as sampling of  $\hat{\eta}$  and  $\gamma$ .

We finally compared the proposed BCD model to the mixed membership stochastic block-model (MMSB) [1] and latent feature relational model (LFRM) [16] on the same datasets and link-prediction setup considered in [16] (see tab. 4.) The number of components in the MMSB model can be estimated based on maximizing the likelihood on hold out data [1]. BCD is the best performing model on the Alyawarra dataset 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 [26] having an AUC of 0.90 and the Multiple Relational Clusterings MRC of [13] 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*) [12] which does not comply with the notion of community structure assumed in BCD.

## 4 Conclusions

Many networks of scientific interest naturally decompose into clusters or communities. In this paper 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 non-parametric models of networks. Our analysis support the observation that community structure is an important structural property of networks [4] that as currently demonstrated can be incorporated into the non-parametric Bayesian modeling of networks. The proposed BCD model trivially extends to other Bayesian generative models such as the multiple-membership modeling of [16, 17] as well as the degree corrected stochastic block model of [11]. A Matlab toolbox with the proposed inference procedure is available for download [18].

## Appendix A: NIPS author communities

In tab. 5 is given 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 (i.e. cluster 2, 3, 9, 10, 15-28) representing dense sub communities of authors that collaborate closely. The remaining smaller clusters extracted by IRM are very similar to the clusters defined by BCD, i.e. 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 model correctly identifies the generated communities (see also *demo.4* in the accompanying Matlab toolbox [18]).

## Acknowledgments

The authors would like to thank two anonymous reviewers for their very useful comments and suggestions to improve the paper.

## References

- [1] Edoardo M. Airoldi, David M. Blei, Stephen E. Fienberg, and Eric P. Xing. Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.*, 9:1981–2014, 2008.
- [2] Aaron Clauset, Cristopher Moore, and M. E. J. Newman. Hierarchical structure and the prediction of missing links in networks. *Nature*, 453(7191):98–101, 2008.
- [3] Jacques Dutka. The incomplete beta function a historical profile. *Archive for History of Exact Sciences*, 24:11–29, 1981. 10.1007/BF00327713.
- [4] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75 – 174, 2010.
- [5] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, June 2002.
- [6] Amir Globerson, Gal Chechik, Fernando Pereira, and Naftali Tishby. Euclidean embedding of co-occurrence data. In *Advances in Neural Information Processing Systems 17*, pages 497–504. MIT Press, 2005.
- [7] Natali Gulbahce and Sune Lehmann. The art of community detection. *BioEssays*, 30(10):934–938, October 2008.

Table 5: NIPS small network communities.

IRM	BCD
<p><b>Cluster 1:</b> 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, DeRuyter-VanSteveninck R, Donoghue J, Doucet A, Doya K, Dreyfus G, Dugas C, Edelman S, Etienne-Cummings R, Franco H, Freeman W, Frey B J, Friedman 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, Opper M, Parr R, Pawelzik K, Pearlmuter B, Perona P, Personnaz L, Pouget A, Precup D, Ritter H, Roth D, Roweis S, Ruderman D, Rumelhart D, Ruppin 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, Weinsall 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><b>Cluster 2:</b> 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><b>Cluster 3:</b> 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><b>Cluster 4:</b> 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><b>Cluster 5:</b> Baker C, Burgard W, Ferguson D, Hahnel D, Morris A, Omohundro Z, Reverte C, Thayer S, Thrun S, Whittaker C, Whittaker W.</p> <p><b>Cluster 6:</b> Boonyanit K, Burr J, Kritayakirana K, Leung M, Murray M, Peterson A, Schwartz E, Stork D, Watanabe T, Wolff G.</p> <p><b>Cluster 7:</b> Blackman D, Chiu T, Clare T, Dao J, Donham C, Hsieh T, Loinaz M, Mueller P, VanderSpiegel J.</p> <p><b>Cluster 8:</b> Jain A, McNair A, Osterholtz L, Saito H, Schmidbauer O, Sloboda T, Tebel-skis J, Waibel A, Woszczyna M.</p>	<p><b>Cluster 1:</b> Bartlett P, Bousquet O, Chapelle 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><b>Cluster 2:</b> Amari S, Chechik G, Coolen A, Doucet A, Edelman S, Hansen L, Horn D, Intrator N, Leen T, Moody J, Opper M, Pawelzik K, Roth D, Ruppin E, Siegelmann H, Spence C, Saad D, Tresp V, Wang X, Weinsall D, Winther O, Yang H, Ziehe A.</p> <p><b>Cluster 3:</b> 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><b>Cluster 4:</b> 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><b>Cluster 5:</b> Bialek W, Chen H, Cottrell G, Crammer K, DeRuyterVanSteveninck R, Friedman 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><b>Cluster 6:</b> 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><b>Cluster 7:</b> 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><b>Cluster 8:</b> Baker C, Burgard W, Ferguson D, Hahnel D, Morris A, Omohundro Z, Reverte C, Thayer S, Thrun S, Whittaker C, Whittaker W.</p> <p><b>Cluster 9:</b> Baldi P, Ballard D, Brunak S, Chauvin Y, Goodman R, Hertz J, Krogh A, Mjolsness E, Perona P, Smyth P, Yuille A.</p> <p><b>Cluster 10:</b> 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><b>Cluster 11:</b> Boonyanit K, Burr J, Kritayakirana K, Leung M, Murray M, Peterson A, Schwartz E, Stork D, Watanabe T, Wolff G.</p> <p><b>Cluster 12:</b> Jain A, McNair A, Osterholtz L, Saito H, Schmidbauer O, Sloboda T, Tebel-skis J, Waibel A, Woszczyna M.</p> <p><b>Cluster 13:</b> Blackman D, Chiu T, Clare T, Dao J, Donham C, Hsieh T, Loinaz M, Mueller P, VanderSpiegel J.</p> <p><b>Cluster 14:</b> Bourlard H, Cohen M, Franco H, Johnson D, Keeler J, Martin G, Morgan N, Rumelhart D.</p> <p><b>Cluster 15:</b> Allen R, Alspecter J, Bienenstock E, Black M J, Donoghue J, Meir R.</p> <p><b>Cluster 16:</b> Cauwenberghs G, Cowan J, Lee D, MacKay D, Seung H.</p> <p><b>Cluster 17:</b> Dreyfus G, Nadal J, Personnaz L.</p> <p><b>Cluster 18:</b> Atlas L, Cole R, Marks R.</p> <p><b>Cluster 19:</b> Bengio Y, Dugas C, Vincent P.</p> <p><b>Cluster 20:</b> Gordon G, Touretzky D.</p> <p><b>Cluster 21:</b> Obermayer K, Mayhew J.</p> <p><b>Cluster 22:</b> Mozer M C, Wolniewicz R.</p> <p><b>Cluster 23:</b> Etienne-Cummings R.</p> <p><b>Cluster 24:</b> VandenBout D.</p> <p><b>Cluster 25:</b> Bengio S.</p> <p><b>Cluster 26:</b> Pearlmuter B.</p> <p><b>Cluster 27:</b> Wiles J.</p> <p><b>Cluster 28:</b> Hastie T.</p>



- [8] Jake M. Hofman and Chris H. Wiggins. Bayesian approach to network modularity. *Physical Review Letters*, 100(25), Jun 2008.
- [9] Sonia Jain and Radford M. Neal. A split-merge markov chain monte carlo procedure for the dirichlet process mixture model. *Journal of Computational and Graphical Statistics*, 13(1):158–182, 2004.
- [10] Brian Karrer, Elizaveta Levina, and M E J Newman. Robustness of community structure in networks. *Physical Review E*, 77(4):1–10, 2008.
- [11] Brian Karrer and M. E. J. Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1):016107+, January 2011.
- [12] Charles Kemp, Joshua B. Tenenbaum, Thomas L. Griffiths, Takeshi Yamada, and Naonori Ueda. Learning systems of concepts with an infinite relational model. In *Artificial Intelligence, Proceedings of the National AAI Conference on*, 2006.
- [13] Stanley Kok and Pedro Domingos. Statistical Predicate Invention. In *Proceedings of the Twenty-Fourth International Conference on Machine Learning*, Corvallis, Oregon, 2007. AAAI Press.
- [14] David Lusseau, Karsten Schneider, Oliver J. Boisseau, Patti Haase, Elisabeth Slooten, and Steve M. Dawson. The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. *Behavioral Ecology and Sociobiology*, 54(4):396–405, 2003.
- [15] A. Menon and C. Elkan. Link prediction via matrix factorization. *Machine Learning and Knowledge Discovery in Databases*, pages 437–452, 2011.
- [16] Kurt T. Miller, Thomas L. Griffiths, and Michael I. Jordan. Nonparametric latent feature models for link prediction. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1276–1284, 2009.
- [17] M. Mørup, M. N. Schmidt, and L. K. Hansen. Infinite multiple membership relational modeling for complex networks, 2010.

- [18] Morten Mørup and Mikkel N. Schmidt. Matlab code for bayesian community detection. [http://www2.imm.dtu.dk/pubdb/views/publication\\_details.php?id=6147](http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=6147), 2011.
- [19] Radford M. Neal. Markov chain sampling methods for dirichlet process mixture models. *Computational and Graphical Statistics, Journal of*, 9:249–265, 2000.
- [20] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69(2):026113–1–15., 2004.
- [21] Krzysztof Nowicki and Tom A. B. Snijders. Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455):1077–1087, 2001.
- [22] J. Pitman. *Combinatorial stochastic processes*, volume 1875. Springer-Verlag, 2006.
- [23] Martin Rosvall and Carl T. Bergstrom. An information-theoretic framework for resolving community structure in complex networks. *Proceedings of the National Academy of Sciences*, 104(18):7327–7331, May 2007.
- [24] J. Shi and J. Malik. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):888–905, 2000.
- [25] Shiwei Sun, Lunjiang Ling, Nan Zhang, Guojie Li, and Runsheng Chen. Topological structure analysis of the protein-protein interaction network in budding yeast. *Nucleic Acids Research*, 31(9):2443–2450, 2003.
- [26] Ilya Sutskever, Ruslan Salakhutdinov, and Joshua Tenenbaum. Modelling relational data using bayesian clustered tensor factorization. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems 22*, pages 1821–1828. 2009.
- [27] U. Von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

- [28] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, June 1998.
- [29] H. Weiler. The use of incomplete beta functions for prior distributions in binomial sampling. *Technometrics*, pages 335–347, 1965.
- [30] Zhao Xu, Volker Tresp, Kai Yu, and Hans-Peter Kriegel. Learning infinite hidden relational models. *Uncertainty in Artificial Intelligence (UAI2006)*, 2006.