

Structural regularity exploration in multidimensional networks via Bayesian inference

Yi Chen¹ · Xiaolong Wang^{1,2} · Buzhou Tang¹

Received: 4 July 2016 / Accepted: 28 April 2017 / Published online: 8 June 2017
© The Natural Computing Applications Forum 2017

Abstract Multidimensional networks, networks with multiple kinds of relations, widely exist in various fields in the real world, such as sociology, chemistry, biology and economics. One fundamental task of network analysis is to explore network structure, including assortative structure (i.e., community structure), disassortative structure (e.g., bipartite structure) and mixture structure, that is, to find structural regularities in networks. There are two aspects of structural regularity exploration: (1) group partition—how to partition nodes of networks into different groups, and (2) group number—how many groups in networks. Most existing structural regularity exploration methods for multidimensional networks need to pre-assume the structure type (e.g., the community structure) and to give the group number of networks, among which the structure type is a guide to group partition. However, the structure type and group number are usually unavailable in advance. To explore structural regularities in multidimensional

networks well without pre-assuming which type of structure they have, we propose a novel feature aggregation method based on a mixture model and Bayesian theory, called the multidimensional Bayesian mixture (MBM) model. To automatically determine the group number of multidimensional networks, we further extend the MBM model using Bayesian nonparametric theory to a new model, called the multidimensional Bayesian nonparametric mixture (MBNPM) model. Experiments conducted on a number of synthetic and real multidimensional networks show that the MBM model outperforms other related models on most networks and the MBNPM model is comparable to the MBM model.

Keywords Multidimensional networks · Network structure · Structural regularity exploration · Mixture model · Bayesian nonparametric theory

1 Introduction

In recent years, with the rapid development of social networking sites (e.g., Twitter, Facebook, Flickr and YouTube), people are available to a large number of real-world networks. Network analysis has attracted considerable attention from more and more researchers in various fields, including sociology [1], biology [2], ecology [3] and engineering [4]. A network can be denoted as a set of nodes with connections between them. Most existing research focuses on a monodimensional network, that is, a network is composed of a single type of relationship (i.e., dimension). However, in many real-world networks, there are also many multidimensional networks where more than one type of relationship (i.e., dimension) exist between nodes. For example, in a social network, various types of

Electronic supplementary material The online version of this article (doi:10.1007/s00521-017-3041-5) contains supplementary material, which is available to authorized users.

✉ Buzhou Tang
tangbuzhou@hitsz.edu.cn

Yi Chen
chenyi@hitsz.edu.cn

Xiaolong Wang
wangxl@insun.hit.edu.cn

¹ Department of Computer Science and Technology, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518055, China

² School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China

relationships such as friend, colleague, schoolmate, romance, neighbor and family may coexist between people.

One fundamental task of network analysis is to explore network structure, that is, to find structural regularities in networks, including assortative structure, disassortative structure [5] and mixture structure [6]. The task of structural regularity exploration is to determine how to partition the nodes of networks into different groups (group partition) and how many groups in networks (group number). In the last decades, a large number of methods have been proposed for group partition or group number. Most of them focus on monodimensional networks, but only a small number of them focus on multidimensional networks. The main reason may lie in that methods for monodimensional networks are not applied to multidimensional networks directly.

A great challenge for structural regularity exploration in a multidimensional network is how to aggregate information of each dimension of the network. Recently, several methods have been proposed for community detection in multidimensional networks. For example, Zhu and Li proposed a unified network transferred from multiplex networks for community detection [7]. Tang and Liu proposed a two-phase strategy for community detection in multidimensional networks, called the principal modularity maximization (PMM) [8]. Boden et al. introduced a multilayer coherent subgraph (MLCS) model and a best-first search algorithm MiMAG to mine coherent subgraphs in multilayer graphs [9]. Mucha et al. presented an extended modularity framework to detect the community structure of arbitrary multiplex networks [10]. All these methods encounter a dilemma that they have to pre-assume the structure type (e.g., the community structure), but this information is, however, usually unavailable before the structures have been found successfully. Recently, several statistical inference-based models concentrate on network analysis and the parameters of models can be used for structure exploration. DuBois presented a generative method for modeling multidimensional networks [11]. Sinkkonen presented an infinite mixture model for multi-relational networks [12]. Xu et al. proposed a multi-relational Gaussian process model to learn multi-relational networks [13]. Although these models have the ability of handling multidimensional networks, they usually perform poorly on many real-world networks for structural regularity exploration.

In order to explore structural regularities in multidimensional networks well without pre-assuming which type of structure they have, in this paper, we propose a novel feature aggregation method based on a mixture model and Bayesian theory, called the multidimensional Bayesian

mixture (MBM) model. To get the MBM model, we first extend the Newman's mixture model (NMM) [5], a model for structural regularity exploration in monodimensional networks, for multidimensional networks, called the multidimensional Newman's mixture model (MNM), and then extend the MNM model using Bayesian theory. Furthermore, to automatically determine the group number of multidimensional networks, we further extend the MBM model using Bayesian nonparametric theory [14] to a new model, called the multidimensional Bayesian nonparametric mixture (MBNPM) model. Experiments conducted on a number of synthetic and real multidimensional networks show that the MBM model outperforms other related models on most networks and the MBNPM model is comparable to the MBM model. As far as we known, the MBNPM model is one of the earliest models that can determine not only the group partition and but also group number of multidimensional networks.

The remainder of this paper is organized as follows. Section 2 briefly presents related work on the subject. Section 3 presents the MBM model and a Markov chain Monte Carlo (MCMC) method for approximate inference. Section 4 presents the MBNPM model and a MCMC method for approximate inference. Experiments are presented in Sect. 5. Section 6 draws conclusions.

2 Related work

2.1 Community detection in multidimensional networks

Community detection in multidimensional networks needs to aggregate the information of all dimensions. According to the information aggregation way, existing methods for community detection in multidimensional networks may fall into two categories: network aggregation and feature aggregation. The network aggregation methods first aggregate all dimensions into a unified network and then employ existing community detection methods on the unified network. Battiston et al. [15] presented a simple unified network in which two nodes are connected if there exists at least one connection between the nodes in all simplex networks. The method is easy to understand, but it clearly loses most of the multiplex information. Tang et al. [16] described another simple united network, called the network integration, via calculating the average interactions among nodes in multidimensional networks. The limitation of this method is that it treats each dimension as being equivalent, which may lose the difference of dimensions. Aiming at solving this limitation, Zhu and Li

[7] proposed a new unified network which takes both the importance of simplex networks and node similarity into consideration. Berlingerio et al. [17] presented a mapping function considering the node neighbors to transform a multidimensional network into a monodimensional one. Wu et al. [18] proposed a co-ranking framework that makes full use of the mutual influence between relations and nodes to transform multi-relational networks to single-relational networks. The feature aggregation methods first extract features or clusters from each dimension and then aggregate them for structure detection. Tang and Liu [8] proposed a principal modularity maximization method to handle community detection in multidimensional networks via a two-phase strategy, where the first phase extracts the structural features from each dimension and the second phase concatenates these features to detect the representative communities. Boden et al. [9] first introduced a multilayer coherent subgraph model to cluster nodes of densely connected in a subset of the graph layers and then used a best-first search algorithm to select the most interesting, non-redundant final clusters. The focus of the method is to identify meaningful groups rather than to provide a group partition for all nodes. Mucha et al. [10] defined interslice connections as connections across different dimensions and extended the modularity to simultaneously model interslice connections and intraslice connections for community detection in a multiscale or multiplex network. This method aims at providing a partition for each dimension of the network rather than for the multidimensional network. Carchiolo et al. [19] further presented a method to improve efficiency of the extended modularity. However, both the network aggregation and feature aggregation methods focus on handling a network with community structure.

2.2 Statistical inference-based models for structural regularity exploration

Several statistical inference-based models have been proposed for handling networks, which may fall into two categories: (1) models special for structural regularity exploration and (2) models designed for network analysis. The models in the first category concentrate on exploring structural regularities in monodimensional networks. Newman and Leicht [5] first introduced a general definition of network structure—a set of nodes that have similar patterns of connection to others, and presented a probabilistic mixture model to explore structural regularities in complex networks. Shen et al. [20] proposed a stochastic block-based model to show clear information on structural regularities in complex networks. Based on the Shen's model, Chai et al. [6] first proposed a popularity–

productivity stochastic block model for general structure detection via introducing popularity and productivity two random variables and then extended it to identify structural regularities in text-associated networks. Chen et al. [21] presented a signed probabilistic mixture model to identify network structure in signed networks, where the model can identify assortative structure in a signed network with only positive links and identify disassortative structure in a signed network with only negative links. The main limitations of these models lie in that they cannot handle multidimensional networks in a direct way and need to pre-specify the group number. The models in the second category concentrate on network analysis, and the parameters of models can be used to explore structure and link prediction. DuBois and Smyth [11] presented a generative method for modeling dyadic events which can be treated as multidimensional networks as each event is denoted by a triplet (sender, recipient, type). Khoshneshin and Street [22] proposed a graphical model which use latent feature networks (LFN) as a framework for multi-relational network analysis. The two network analysis models also suffer from a limitation that specifying the group number is necessary. To allow a model to automatically determine the group number, researchers recently tried Bayesian non-parametric models. Kemp et al. [23] proposed an infinite relational model to learn systems of concepts in relational datasets, where object-feature data can be profitably viewed as a relation between two sets of entities. Sinkkonen et al. [12] presented a simple infinite model for general multi-relational data, where co-occurrences within a single categorical variable can be viewed as multiple relations between two sets of entities. Xu et al. [13] proposed a multi-relational Gaussian process model to learn multi-relational data. Although these models have the ability of handling multidimensional networks, they usually perform poorly on many real-world networks for structural regularity exploration.

3 Multidimensional Bayesian mixture (MBM) model

Generally, a multidimensional network with D dimensions and N nodes can be represented as $A = \{A^{(1)}, \dots, A^{(d)}, \dots, A^{(D)}\}$. For the d th dimension, $A_{ij}^{(d)} = 1$, ($1 \leq i, j \leq N$) if there is a link from node i to node j and 0 otherwise. And we use $L_i^{(d)}$ to denote the out links of node i in the d th dimension (i.e., a set of neighbor edges of node i in an undirected network). In this section, we first extend the NMM model to the MNM model for handling multidimensional networks and then derive the MBM model from

the MNM model using Bayesian theory. For convenience, we only introduce them on directed networks in detail.

3.1 Multidimensional Newman’s mixture (MNM) model

A directed monodimensional network with K groups (K is a predefined number) can be generated by the NMM model with two parameters $u_{ij} \in E_U$ and θ_{kj} , where π_k denotes the probability of a node in group k ($k \in \{1, 2, \dots, K\}$) and θ_{kj} denotes the probability of a link from a node in group k connecting to node j , subjected to the normalization constraint $\sum_{j=1}^N \theta_{kj} = 1$. The vector θ_k represents the characteristic of nodes in group k linking to other nodes. According to θ_k , nodes connecting to other nodes in similar patterns are grouped together. A network is generated in the following way: For each node i and its links L_i , (1) node i falls into a group z_i with probability π_{z_i} and (2) each link $A_{ij} \in L_i$ selects node j with probability $\theta_{z_i j}$. The probability of a network A with N nodes can be written as

$$p(A, z|\pi, \theta) = p(A, z|\theta) \cdot p(z|\pi) = \prod_{i=1}^N \left(\pi_{z_i} \cdot \prod_{j=1}^{L_i} \theta_{z_i j}^{A_{ij}} \right) \quad (1)$$

where $z_i \in \{1, \dots, K\}$ is a hidden variable that needs to be inferred. The logarithm of Eq. (1) is as follows:

$$\ln p(A, z|\pi, \theta) = \sum_{i=1}^N \left(\ln \pi_{z_i} + \sum_{j=1}^{L_i} A_{ij} \ln \theta_{z_i j} \right) \quad (2)$$

Because of hidden variables $z_i, i = 1, \dots, N$, parameters π_k and $1 \leq i, j \leq n$ cannot be estimated using likelihood maximization estimation. An expectation–maximization (EM) [24] algorithm is proposed to estimate them. In the E step, the expectation of Eq. (2) is:

$$\bar{L} = \sum_{i,k} q_{ik} \left[\ln \pi_k + \ln \sum_j A_{ij} \ln \theta_{kj} \right] \quad (3)$$

with

$$q_{ik} = \frac{\pi_k \prod_j \theta_{kj}^{A_{ij}}}{\sum_k \pi_k \prod_j \theta_{kj}^{A_{ij}}}, \quad (4)$$

where q_{ik} denotes the probability of node i belonging to group k . In the M step, the parameters π and θ can be re-estimated by optimizing Eq. (3) as

$$\pi_k = \frac{\sum_i q_{ik}}{N}, \quad (5)$$

$$\theta_{kj} = \frac{\sum_i A_{ij} q_{ik}}{\sum_j \sum_i A_{ij} q_{ik}}, \quad (6)$$

The θ_{kj} represents the link preference or feature of a group connecting nodes. In a community structure, nodes with more links connect the nodes of the same group, leading to a group with larger θ_{kj} to choose nodes within the group than other nodes; in a bipartite structure, nodes with more links connect across groups, leading to a group with larger θ_{kj} to choose nodes in the contrary group; in a mixture structure of community structure and bipartite structure, some groups with larger θ_{kj} to choose nodes within the group and other groups with larger θ_{kj} to choose nodes in the contrary group. So we use

$$M = \begin{bmatrix} \theta_{11} & \dots & \theta_{k1} & \dots & \theta_{K1} \\ \dots & \dots & \dots & \dots & \dots \\ \theta_{1j} & \dots & \theta_{kj} & \dots & \theta_{Kj} \\ \dots & \dots & \dots & \dots & \dots \\ \theta_{1N} & \dots & \theta_{kN} & \dots & \theta_{KN} \end{bmatrix}$$

as the network structural features.

The above model has extracted structural features from a monodimensional network. For multidimensional networks, we follow the PMM model [8] to aggregate the structural features of all dimensions as follows

$$M = [M^{(1)}, \dots, M^{(d)}, \dots, M^{(D)}], \quad (7)$$

and perform k -means on the M to get the final node partition.

In the NMM model, node i may not fall into group k because of u_{p_i} in Eq. (4) and will be wrongly assigned to another group. Specially, when $q_{i1} = \dots = q_{iK} = 0$, node i may not fall into any group, which results in that the NMM model fails to handle this network. To avoid this problem, we extend the MNM model using Bayesian theory.

3.2 Multidimensional Bayesian mixture (MBM) model

In a Bayesian setting, the model parameters will themselves be random variables, and prior distributions will be placed over the variables. Currently, the predominant choice is the use of conjugate priors. The conjugate prior for parameters π and θ_k is the Dirichlet distribution

$$\pi \sim \text{Dirichlet}(\alpha), \quad (8)$$

$$\theta_k \sim \text{Dirichlet}(\beta), \quad (9)$$

where α and β are the Dirichlet hyperparameters.

The generative process for a network is summarized as follows:

1. Draw π from Dirichlet(α);
2. For each group k in K groups; Draw θ_k from Dirichlet(β);
3. For each new node i :
 - (a) Draw a latent group z_i from π ;
 - (b) For each link in L_i :

Draw an end node j from θ_{z_i} .

Then, the probability of a network A with N nodes is (refer to Eq. 1):

$$p(A, z|\alpha, \beta) = p(A, z|\pi, \theta)p(\pi|\alpha)P(\theta|\beta) \tag{10}$$

Due to the conjugacy between the Dirichlet and Multinomial distributions, Eq. (10) can be simplified as:

$$p(A, z|\alpha, \beta) = p(A|z, \beta)p(z|\alpha) \tag{11}$$

with

$$p(z|\alpha) = \int p(z|\pi)p(\pi|\alpha)d\pi \tag{12}$$

$$p(A|z, \beta) = \int p(A|z, \theta)p(\theta|\beta)d\theta. \tag{13}$$

3.3 Inference

The latent variable z of the Bayesian model (Eq. 11) cannot be exactly inferred, but can be approximately inferred by a wide variety of approximate inference algorithms such as Laplace approximation, variational approximation and Markov chain Monte Carlo (MCMC). In this study, we follow Palla et al.’s work [25] to use MCMC for parameter estimation, which follows an iterative procedure to achieve posterior inference over the latent variable z by the Gibbs and slice sampling. The sampler iterates as follows:

3.3.1 Sampling z

Gibbs sampling [26] is used to obtain samples of z . For each node i , given the group assignment for all other nodes, the group probability of the omitted node i choosing group k is as follows:

$$p(z_i = k|z_i, A) \propto \prod_{j=1}^{L_i} \frac{m_{k,i}^j + \beta}{m_{k,i} + N\beta + j - 1} \cdot \frac{n_k + \alpha}{N + K\alpha} \tag{14}$$

where n_k denotes the number of nodes belongs to group k , $m_{k,i}$ denotes the number of out links from nodes that belong to k except node i and $m_{k,i}^j$ denotes the number of

out links from nodes that belong to k except node i to node j . The probability $p_{ij} \in E_P$ that has the same meaning of q_{ik} (see Eq. 4) avoids the zero problem as $\frac{m_{k,i}^j + \beta}{m_{k,i} + N\beta + j - 1} > \frac{m_{k,i}^j}{m_{k,i} + (N-1)\beta + j - 1} \geq 0$.

3.3.2 Hyperparameters

We use slice sampling [27] for hyperparameters α and β and limit them in (0, 1).

The parameter $m_{k,i}^j$ denotes the number of out links from nodes falling into group k to node j . It plays a similar role as θ_{kj} in NMM. So we use

$$M = \begin{bmatrix} m_1^1 & \dots & m_k^1 & \dots & m_K^1 \\ \vdots & & & & \vdots \\ m_1^j & \dots & m_k^j & \dots & m_K^j \\ \vdots & & \vdots & & \vdots \\ m_1^N & \dots & m_k^N & \dots & m_K^N \end{bmatrix}$$

as network structural features.

Algorithm 1 shows the procedure in full. For each dimension, the initial group of each node is randomly assigned, and the data are sampled after the log-likelihood reaches a stationary state. While 20–30 iterations often appear sufficient for convergence, we run 2000 iterations for all networks in the following experiments. For each dimension, the sample with highest log-likelihood corresponds exactly to the expected “true” structural feature.

Algorithm 1: MCMC for MBM model

```

Input: a multidimensional network with  $N$  nodes and  $D$  dimensions, pre-
specified group number  $K$ .
Output: group partition  $P = \{g_1, \dots, g_n, \dots, g_N\}$ .
1 Initialize feature matrix  $M=[]$ ,  $\alpha = 0.5$ ,  $\beta = 0.5$ 
2 for  $d=1$  to  $D$  do
3   for  $i=1$  to  $N$  do
4     Initialize  $g_i$  with random integer between 1 and  $K$ 
5   end for
6   Compute count matrices  $m;n$ 
7   while not converged do
8     for  $i=1$  to  $N$  do
9       Decrement  $m[g_i, L_i]$ ,  $n[g_i]$ 
10      for  $k=1$  to  $K$  do
11         $\eta[k] \leftarrow \prod_{j=1}^{L_i} \frac{m_{k,i}^j + \beta}{m_{k,i} + N\beta + j - 1} \cdot \frac{n_k + \alpha}{N + K\alpha}$ 
12      end for
13       $g_i \leftarrow \text{IndexOf}(\text{RandomMultinomial}(1; \eta))$  // Gibbs sampling
14      Increment  $m[g_i, L_i]$ ,  $n[g_i]$ 
15    end while
16     $\alpha \leftarrow \text{slicesample}(\alpha)$  // Slice sampling
17     $\beta \leftarrow \text{slicesample}(\beta)$  // Slice sampling
18  end while
19   $M = [M \ m]$ 
20 end for
21  $\{g_1, \dots, g_i, \dots, g_N\} = \text{kmeans}(M, K)$ 

```

The MBM model can also be extended to handle undirected networks like the NMM model. The difference between directed networks and undirected networks just lies in that any link between two nodes is symmetrical in undirected networks while not symmetrical in directed networks. Therefore, in an undirected network, when node i in z_i connects to another node j , node j in z_j also connects to node i . Again, we use $\varphi_{z_{ij}}$ to denote the probability that a node in group z_i choosing to connect to node j . $A_{ij} \in L_i$ is generated according to the multinomial distribution $\text{Multinomial}(\theta_{z_{ij}}\theta_{z_{ji}})$, where $\theta_{z_{ij}}\theta_{z_{ji}}$ denotes the probability of an edge between node i and j , given nodes i and j in groups $T_{ij}^{(t)}$ and z_j , respectively. This probability is normalized by the constraint $\sum_{i=1}^N \sum_{j=1}^N \varphi_{z_{ij}}\varphi_{z_{ji}} = \left[\sum_{i=1}^N \varphi_{z_{ji}} \right] \cdot \left[\sum_{j=1}^N \varphi_{z_{ij}} \right] = 1$. Since both z_i and z_j vary from 1 to K , $\sum_{i=1}^N \varphi_{z_{ji}} = \sum_{j=1}^N \varphi_{z_{ij}}$. Thus, $\sum_{j=1}^N \varphi_{z_{ij}} = 1$ exactly the same as in the directed case. The remainder of the generative process follows as before.

4 Multidimensional Bayesian nonparametric mixture (MBNPM) model

To allow the MBM model to determine the group number, we use a common Bayesian nonparametric method, called the Chinese restaurant process (CRP), for nonparametric distribution of latent groups. The CRP, a vivid metaphor for building a partition of nodes, assigns a new node (i.e., a new customer entering the restaurant) to a new group (i.e., table) or the existing groups (tables) with a probability below

$$p(z_i = k | z_1, \dots, z_{i-1}) = \begin{cases} \frac{\alpha}{i-1+\alpha} & k \text{ is a new group} \\ \frac{n_k}{i-1+\alpha} & n_k > 0 \end{cases}, \tag{15}$$

where n_k denotes the number of customers already assigned to table k and α is a hyperparameter. The characteristics of the CRP are as follows: (1) nodes tend to fall into popular

groups and make the popular groups more popular, and (2) new nodes always have a chance to fall into new groups. These characteristics allow the group number K to approach infinity, but only a finite number of groups are used to generate the observed network.

In MBNPM, the generative process for a network is summarized as follows:

1. Draw π from $\text{CRP}(\alpha)$;
Draw a latent group z_i from $\text{CRP}(\alpha)$;
2. For each group k in ∞ groups;
Draw θ_k from $\text{Dirichlet}(\beta)$;
3. For each new node i :
 - (a) Draw a latent group z_i from π ;
 - (b) For each link in L_i :

Draw an end node j from θ_{z_i} .

Then, the probability of a network A (refer to Eq. 11) is:

$$p(A, z, K | \alpha, \beta) = p(A | z, \beta) p_{\text{CRP}}(z, K | \alpha). \tag{16}$$

The hyperparameter α , the priori probability of π , impacts on the number of groups. Although the group number is potentially infinite, the CRP gives an extremely uneven distribution over groups and ensures that the number of groups K is much smaller than the number of nodes N with an appropriate small value α . The hyperparameter β , the prior probabilities of θ , describes the degree distribution of a node within groups.

In the process of inference, the group probability of the omitted node i choosing group k is changed as (refer to Eq. 14):

$$p(z_i = k | z_i, A) \propto \prod_{j=1}^{L_i} \frac{m_{k,i}^j + \beta}{m_{k,i} + N\beta + j - 1} \cdot \frac{F(n_k, \alpha)}{N + K\alpha} \tag{17}$$

where $F(n_k, \alpha) = n_k$, if $n_k > 0$; otherwise $F(0, \alpha) = \alpha$, meaning that a new group is generated.

Algorithm 2 shows the procedure in full. The initialization of K can be any positive integer. For the aggregated structural features M , we adopt the density peaks (DP)-based method [28] instead of K -means to automatically determine the group number and group partition.

Algorithm 2: MCMC for MBNPM model

Input: a multidimensional network with D dimensions and N nodes.
Output: group number K , group partition $P = \{g_1, \dots, g_i, \dots, g_N\}$.

```

1 Initialize feature matrix  $M=[\ ]$ ,  $\alpha = 0.5$ ,  $\beta = 0.5$ ,  $K=1$ 
2 for  $d=1$  to  $D$  do
3   for  $i=1$  to  $N$  do
4     Initialize  $g_i$  with random integer between 1 and  $K$ 
5   end for
6   Compute count matrices  $m;n$ 
7   while not converged do
8     for  $i=1$  to  $N$  do
9       Decrement  $m[g_i, L_i]$ ,  $n[g_i]$ 
10      Remove empty group when  $n[g_i]=0$ 
11       $KK \leftarrow K+1$ 
12      for  $k=1$  to  $KK$  do
13         $\eta[k] \leftarrow \prod_{j=1}^{L_i} \frac{m_{k,-i}^j + \beta}{m_{k,-i} + N\beta + j - 1} \cdot \frac{n_k + \alpha}{N + K\alpha}$ 
14      end for
15       $g_i \leftarrow \text{IndexOf}(\text{RandomMultinomial}(1; \eta))$  // Gibbs sampling
16      if  $g_i > K$  do
17        Add new group  $m[g_i, :] \leftarrow 0$ ,  $n[g_i] \leftarrow 0$ 
18      end if
19      Increment  $m[g_i, L_i]$ ,  $n[g_i]$ 
20    end for
21     $\alpha \leftarrow \text{slicesample}(\alpha)$  // Slice sampling
22     $\beta \leftarrow \text{slicesample}(\beta)$  // Slice sampling
23  end while
24   $M=[M \ m]$ 
25 end for
26  $K, \{g_1, \dots, g_i, \dots, g_N\} = \text{DP}(M)$ 

```

5 Experiments

To investigate the effectiveness of the MBM and MBNPM models for structural regularity exploration in multidimensional networks, we compare them with other state-of-the-art methods on nine synthetic and real multidimensional networks, including a synthetic network and a real network with community structure, a synthetic network with tripartite structure, a real network with bipartite structure, and a synthetic network and four real networks with mixture structure. The nine networks are all multidimensional networks with gold standard structures we can collect from other studies. The detailed information of these networks is shown in Table 1.

To test the performance of the MBM and MBNPM models, we compare them with seven models for multidimensional networks, including three network aggregation methods (i.e., the Unify-Infomap (UI), Unify-OSLOM (UO) and Unify-Louvain (UL) proposed by [7]), a feature aggregation method (i.e., the PMM proposed by [8]), two

statistical inference-based models (i.e., the Marginal Product Mixture Model (MPMM) proposed by [11] and the simple relational component model (SRCM) proposed by [12]) and MNM. The source code of the PMM model has been released by the authors. We use the released source codes in this study and implement UI, UO, UL, MPMM, SRCM and MNM models by ourselves. For the PMM, MNM, MPMM and MBM models that require a pre-specified group number, we adopt the gold standard. As the PMM, MNM and MBM may converge to local optima, we run each model ten times and take the solution giving the highest evaluation.

The performance of all models is measured by the normalized mutual information (NMI) [29], which is widely used for evaluating the structure detection:

$$P_{nmi}(G, G') = \frac{2MI(G, G')}{H(G) + H(G')}, \quad (18)$$

where $G = (G_1, G_2, \dots, G_K)$ are defined groups in a network, $G' = (G'_1, G'_2, \dots, G'_K)$ are groups detected by an

Table 1 Detailed information of the nine networks used in our study

Name	N	D	E	K	Directed	Structure types	Source
Syn-com	350	4	(17,858; 15,180; 10,728; 2716)	3	No	Community	Synthetic
Ckm ^a	246	3	(898; 996; 846)	4	No	Community	Real
Syn-tri	150	4	(1274; 1258; 2056; 1288)	3	No	Tripartite	Synthetic
DutchElite ^b	4747	4	(1090; 304; 104; 8260)	2	No	Bipartite	Real
Syn-mix	150	4	(1730; 1340; 1312; 1562)	5	No	Mixture	Synthetic
Lazega ^a	71	3	(612; 855; 756)	2	Yes	Mixture	Real
Gd99 ^b	234	4	(124; 102; 72; 308)	8	No	Mixture	Real
Bay ^b	125	2	(1969; 1938)	7	Yes	Mixture	Real
Email ^c	4408	4	(154,1826; 137,104; 12,040; 12,504)	13	No	Mixture	Real

^a <http://moreno.ss.uci.edu/data.html>

^b <http://vlado.fmf.uni-lj.si/pub/networks/data/>

^c http://www.poleia.lip6.fr/~jacoby/Research/research_en.html

algorithm, $H(G)$ and $H(G')$ are the entropies of G and G' and $MI(G, G')$ is the mutual information between them. A high P_{nmi} means a good detection. Specially, $P_{nmi} = 1$ means a perfect detection.

Before the comparison, we analyze the convergence of the Bayesian mixture model and nonparametric Bayesian mixture model on a monodimensional network. Figure 1 illustrates examples of the two models on the Syn-com four dimensions shown in Fig. 2a–d (described in the following section). As can be seen, the two models converge to a stationary state within a few iterations on all networks, which means they can quickly explore the structural features. Compared with the nonparametric Bayesian mixture model, the Bayesian mixture model shows more stable log-likelihoods on the corresponding four dimensions, which means it explores clearer structural features with pre-specified group numbers. In addition, both the two models show more stable log-likelihoods in the former one dimension than the latter three dimensions as it is composed of clearer structure.

5.1 Multidimensional networks with assortative structure

Syn-com and Ckm, two multidimensional networks with community structure, are used to test the capability of our models on assortative structure detection. The Syn-com is a common undirected synthetic dataset for community detection in multidimensional networks [7, 8]. The network has three groups, with each having 50, 100 and 200 nodes, respectively. It is composed of four dimensions. For each dimension, edges within groups are randomly generated with an interaction probability and noise edges randomly connecting two nodes are generated with low probability. As shown in Fig. 2a–d, each dimension demonstrates partial communities and four dimensions demonstrate the whole communities. Ckm [30] is a real undirected

multidimensional network concerning the impact of 246 physicians adoption of a new drug. The network contains three dimensions which are generated according to three sociometric questions for each doctor: To whom did he most often turn for advice and information? With whom did he most often discuss his cases in the course of an ordinary week? Who were the friends, among his colleagues, whom he saw most often socially? All physicians come from four towns in Illinois: Peoria, Bloomington, Quincy and Galesburg, which form a community structure. It is a sparse multidimensional network without any edge noise. Table 2 shows the P_{nmi} results when applying nine models on the nine networks, where the numbers in bold are the best P_{nmi} s on the networks. As can be seen, on the Sym-com network, both MBM and MBNPM models achieve a perfect P_{nmi} of 1, that is all nodes are partitioned into the correct groups, which are the same as the MNM model and outperform other models; on the Ckm network, the MBM achieves a P_{nmi} of 0.9385 and the MBNPM achieves a P_{nmi} of 0.7568, which are inferior to the UI model but outperform other models. We further analyze the UI model which achieves a P_{nmi} of 0 and 0.9628 on the Syn-com and Ckm networks, respectively. The UI is a model based on multilevel compression of random walks, which is sensitive to network noise. The Syn-com is accompanied with edge noise in each dimension and the Ckm is a network without any edge noise. So, it is reasonable to us that the UI model makes a tremendous difference on the two networks. Overall, the MBM and MBNPM models show comparable performance with other four models on the multidimensional networks with assortative structure.

5.2 Multidimensional networks with disassortative structure

Syn-tri and Elite, two multidimensional networks with multipartite structure, are used to test the capability of our models on disassortative structure detection. The Syn-tri is

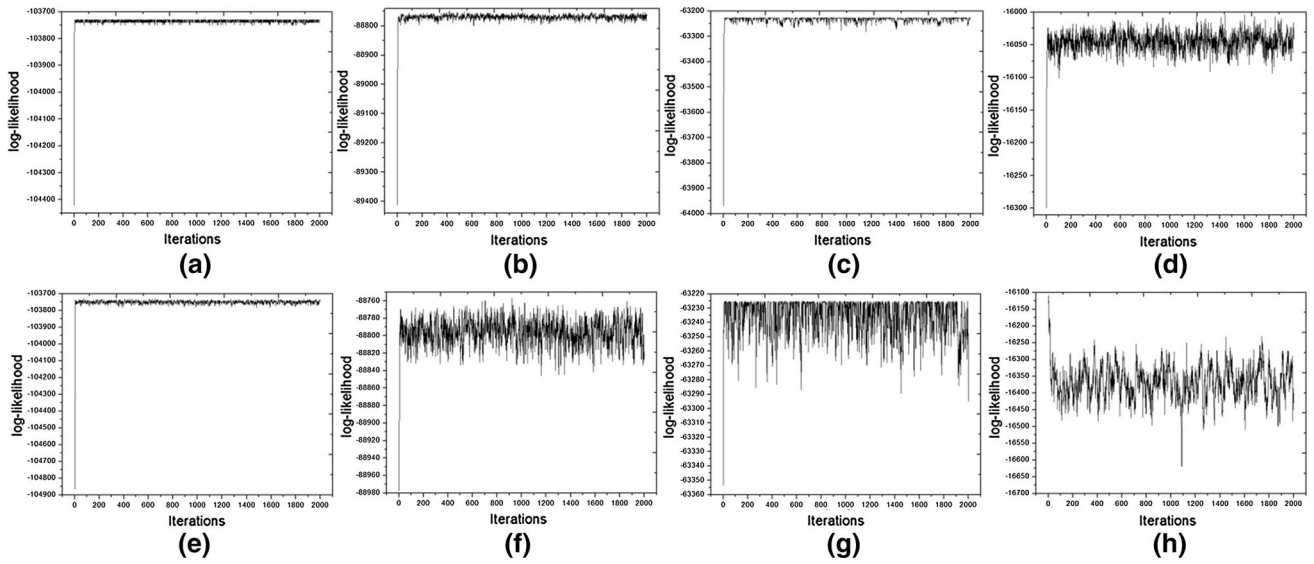


Fig. 1 The log-likelihood of the Bayesian mixture model and nonparametric Bayesian mixture model on the Syn-com four dimensions. **a–d** Bayesian mixture model, **e–h** nonparametric Bayesian mixture model

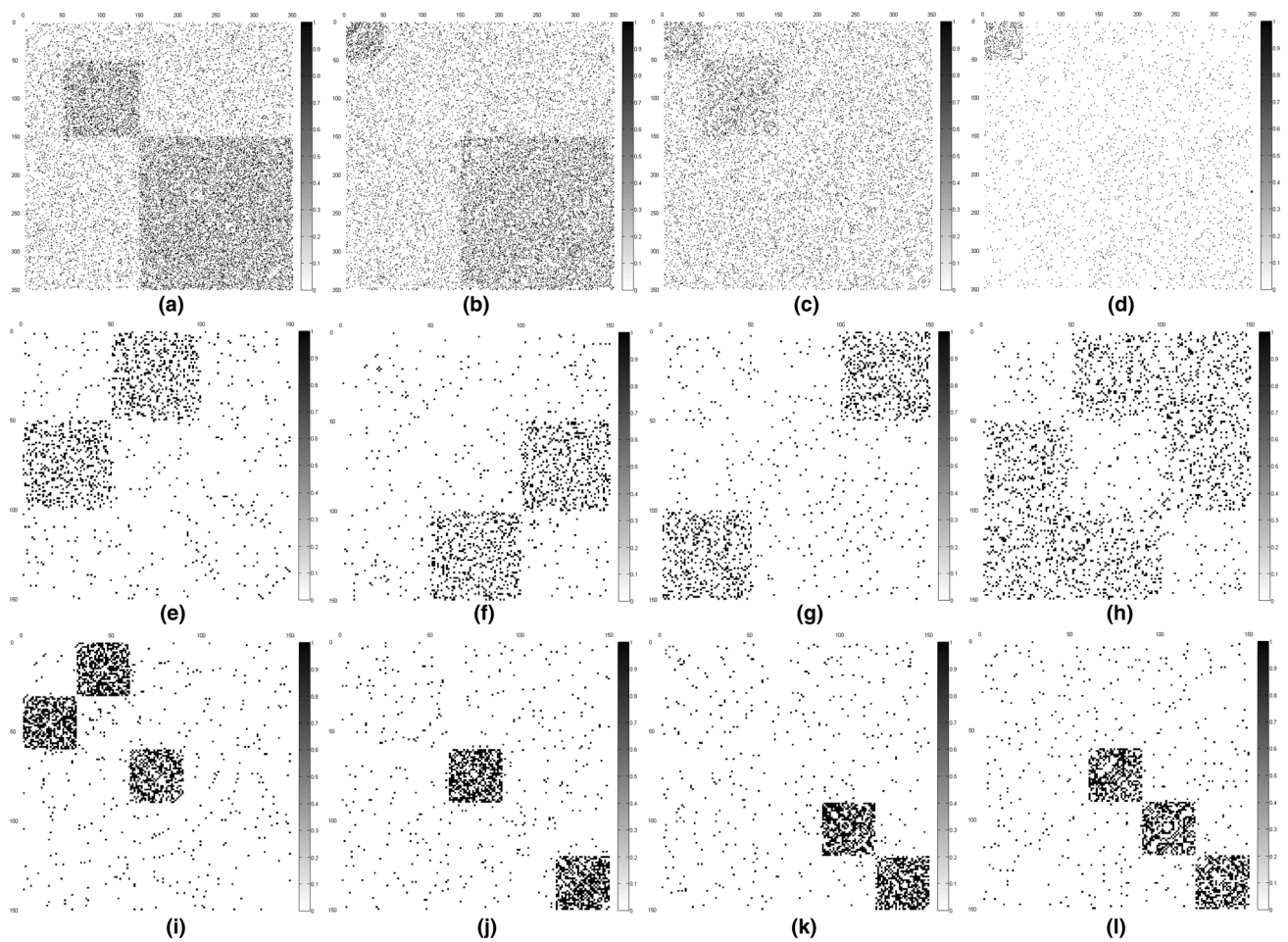


Fig. 2 The adjacency matrix of multidimensional networks. **a–d** Syn-com network, **e–h** Syn-tri network, **i–l** Syn-mix network

Table 2 P_{nmi} s identified by nine models on the nine multidimensional networks

ID	UI	UO	UL	PMM	MPMM	SCRM	MNM	MBM	MBNPM
1	0	0.9822	0.1946	0.9665	0.0074	0.5576	0.9727	1	1
2	0.9628	0.6485	0.5668	0.648	0.0944	0.6892	0.7164	0.9385	0.7568
3	0	0.0451	0.2745	0.0085	0.0027	0.0224	1	1	1
4	0.0233	0.0323	0.0228	0.0081	0.0016	0.0019	0.0182	0.3467	0.1248
5	0.8538	0.7252	0.3977	0.8013	0.1104	0.9057	1	1	1
6	0	0.5294	0.2171	0.0133	0.0002	0.0312	0.0162	0.5699	0.2521
7	0.2012	0.1098	0.2063	0.1087	0.0541	0.0469	0.1964	0.1942	0.1728
8	0	0.0419	0.3773	0.4443	0.1146	0.3158	0.3832	0.4726	0.3419
9	0.2746	0.2964	0.2257	0.3479	0.0016	0.2595	0.151	0.4078	0.3905

Bold values are the best P_{nmi} s on the networks

an undirected synthetic multidimensional network with tripartite structure to test the MBM model. The network has 150 nodes, equally dividing to three groups. It is composed of four dimensions, three of which demonstrates a bipartite structure and one of which demonstrates a tripartite structure shown in Fig. 2e–h. Elite [31] is a real undirected multidimensional network describing the relation between 3810 persons and 937 administrative bodies in the Dutch government. The network contains four dimensions characterizing four major relations: chairman, vice-chairman, treasurer and lid/member. It is a bipartite network with two groups: person and administrative body. As shown in Table 2, on the Syn-tri network, both MBM and MBNPM models achieve a perfect P_{nmi} of 1, which are the same as the MNM model and outperform other models; on the Elite network, the MBM achieves a P_{nmi} of 0.3467 and the MBNPM achieves a P_{nmi} of 0.1248, which are higher than other models by at least 0.3144 and 0.0925, respectively. The reason why the UI, UO, UL, PMM and MPMM models are inferior to the MBM and MBNPM models is that they have pre-assumed the community structure on the two networks, whereas the MBM and MBNPM models have not.

5.3 Multidimensional networks with mixture structure

In addition to assortative structure and disassortative structure, other types of structures also exist in multidimensional networks. Syn-mix, Lazega [32], Gd99 [33], Bay [34] and Email [35], five multidimensional networks with mixture structure, are used to test the capability of our models on mixture structure detection. The Syn-mix is an undirected synthetic multidimensional network with a mixture of community structure and bipartite structure. It contains 150 nodes, equally dividing to five groups, two of which form a bipartite structure and other three of which form a community structure. The network is composed of four dimensions as shown in Fig. 2i–

l. The Lazega is a real directed multidimensional network collected from a study of corporate law partnership between 71 attorneys which was carried out in a Northeastern US corporate law firm. The network is composed of three dimensions, including strong-coworker network, advice network and friendship network. They are, respectively, generated from the following three questionnaires: Would you go through this list and check the names of those with whom you have worked with? Think back over the past year, to whom did you go for basic professional advice? Would you go through this list, and check the names of those you socialize with outside work? All attorneys are divided into two statuses: partner and associate, forming a mixture structure. The Gd99 is a real undirected network characterizing 234 characters and their relations in the long-running German soap opera called “Lindenstrasse.” The network contains four dimensions describing four relations: family, business, friendship and partner. All characters are labeled by eight colors: yellow, green, blue, pink, white, orange, gray and black. The Bay is real directed foodweb networks describing a 125-component budget of the carbon exchanges occurring during the wet and dry seasons in Florida Bay. Nodes in these networks represent major components of the ecosystem, and edges represent the transfers of material or energy between the components. All the 125 nodes are categorized into seven species: primary producers, invertebrates, fishes, birds, reptiles, mammals and detrital compartments. The Email is a real undirected networks extracted from the emails of the authors of this article. It is described by four different types of relations: authorship, recipient, copy and sameday. All emails are categorized into 13 groups: administration, archives, articles, corbeillesauvegarde, enseignement, inbox, mailinglists, personnel, projets, recherche, sent, spam and trash. Table 2 illustrates the results identified by the nine models. As can be seen, the MBM model achieves the best P_{nmi} s on the Syn-mix, Lazega, Bay and Email networks and the MBNPM model achieves the best P_{nmi} s on the Syn-mix and Email networks. Specifically, On the Syn-mix network, both MBM and MBNPM models

Table 3 P_{nmi} s identified by nine models on the three synthetic multidimensional networks except a certain dimension

ID	UI	UO	UL	PMM	MPMM	SCRM	MNM	MBM	MBNPM
C1	0	0.2335	0.1745	0.3991	0.0125	0.3842	0.2758	0.8832	0.9188
C2	0	0.1891	0.1769	0.7751	0.0016	0.504	0.4063	1	1
C3	0	0.227	0.1947	0.5499	0.0044	0.6593	0.6136	1	1
C4	0	0.681	0.2504	0.7388	0.005	0.553	0.4786	0.9843	0.9727
T1	0	0.0393	0.134	0.0128	0.0124	0.0211	0.716	1	0.9473
T2	0	0.042	0.1507	0.0113	0.0083	0.0125	0.6855	1	1
T3	0	0.0302	0.1278	0.0032	0.0095	0.0243	0.6563	1	1
T4	0	0.0197	0.1875	0.0085	0.0276	0.0264	0.6802	1	1
M1	0.6462	0.5524	0.4861	0.8059	0.0519	0.6113	0.7271	0.8714	0.9057
M2	0.8876	0.8651	0.5485	0.7931	0.0815	0.9057	0.8982	1	1
M3	0.8479	0.6163	0.5422	0.6743	0.0865	0.9057	0.8476	1	1
M4	0.7863	0.5026	0.5313	0.6541	0.0392	0.9057	0.7138	1	1

Bold values are the best P_{nmi} s in the networks

achieve a perfect P_{nmi} of 1, which are the same as the MNM model and outperform other models; on the Lazega network, the P_{nmi} of the MBM model is 0.5699, which is higher than other models by at least 0.0405, and the P_{nmi} of the MBNPM model is 0.2521, which is inferior to the UO and MBM models but is superior to UI, UL, PMM, MNM, MPMM and SCRM models; on the Gd99 network, the P_{nmi} s of the MBM and MBNPM models are 0.1942 and 0.1728, respectively, which are inferior to the UI, UL and MNM models but are superior to other models; on the Bay network, the P_{nmi} of the MBM model is 0.4726, which is higher than other models by at least 0.0283, and the P_{nmi} of the MBNPM model is 0.3419, which is inferior to the UL, PMM, MNM and MBM models; on the Email network, the P_{nmi} of the MBM is 0.4078, which is higher than other models by at least 0.0599, and the P_{nmi} of the MBM is 0.3905, which is higher than other models by at least 0.0426. Overall, the MBM and MBNPM models show comparable performance with other four models on the multidimensional networks with mixture structure.

5.4 Robust analysis

We also compare the nine models on synthetic multidimensional networks except a certain dimension for robust analysis. (C1, C2, C3, C4), (T1, T2, T3, T4) and (M1, M2, M3, M4) are used to denote the Syn-com, Syn-tri and Syn-mix networks except the first, second, third and fourth dimension, respectively. Table 3 illustrates the results identified by the nine models on the 12 multidimensional networks, where numbers in bold are the best P_{nmi} s in the networks. It can be seen that the MBM model achieves the best P_{nmi} s on all networks except the C1 and M1, and the MBNPM model achieves the best P_{nmi} s on all networks except the C4 and T1. On the C1 and M1, the MBM model is inferior to the MBNPM but superior to other models; on the C4 and T1, the MBNPM is inferior to the

MBM but superior to other models. The reason why the MBM fails to achieve a P_{nmi} of 1 on the C1 and M1 is that the structure of these networks is not clear; for example, the bipartite structure is nonexistent in M1. Overall, the MBM and MBNPM models outperform other models on all networks.

6 Conclusion

In this paper, we propose a novel feature aggregation method based on a mixture model and Bayesian theory for multidimensional networks without pre-assuming the structure type, called the multidimensional Bayesian mixture (MBM) model. We also further extend the MBM model using Bayesian nonparametric theory to automatically determine the group number, called the multidimensional Bayesian nonparametric mixture (MBNPM) model. Experiments conducted on a number of synthetic and real multidimensional networks show that the MBM and MBNPM models (1) show comparable performance with other state-of-the-art methods in networks with community structure or mixture structure and (2) outperform other methods in networks with disassortative structure. As future work, we will apply our models to structural regularity exploration on real networks and seek possible applications.

Acknowledgements This paper is supported in part by Grants: National 863 Program of China (2015AA015405), NSFCs (National Natural Science Foundations of China) (61573118, 61402128, 61473101 and 61472428), Special Foundation for Technology Research Program of Guangdong Province (2015B010131010), Strategic Emerging Industry Development Special Funds of Shenzhen (20151013161937, JSGG20151015161015297 and JCYJ20160531192358466), Innovation Fund of Harbin Institute of Technology (HIT.NSRIF.2017052), Program from the Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education (93K172016K12) and CCF-Tencent Open Research Fund (RAGR20160102).

Compliance with ethical standards

Conflict of interest The article is an extension of the paper entitled “Structural regularity exploration in multidimensional networks” published in the 22nd International Conference on Neural Information Processing (ICONIP 2015).

References

- de Franciscis S, Caravagna G, d’Onofrio A (2016) Gene switching rate determines response to extrinsic perturbations in a transcriptional network motif. *Sci Rep* 6:26980. doi:10.1038/srep26980
- Newman M, Clauset A (2015) Structure and inference in annotated networks. *Nat Commun* 7:11863. doi:10.1038/ncomms11863
- Coyte KZ, Schluter J, Foster KR (2015) The ecology of the microbiome: networks, competition, and stability. *Science* 350(6261):663–666. doi:10.1126/science.aad2602
- Neftci E et al (2013) Synthesizing cognition in neuromorphic electronic systems. *Proc Natl Acad Sci* 110(37):E3468–E3476. doi:10.1073/pnas.1212083110
- Newman ME, Leicht EA (2007) Mixture models and exploratory analysis in networks. *Proc Natl Acad Sci* 104(23):9564–9569. doi:10.1073/pnas.0610537104
- Chai B-F et al (2013) Combining a popularity-productivity stochastic block model with a discriminative-content model for general structure detection. *Phys Rev E* 88(1):012807. doi:10.1103/PhysRevE.88.012807
- Zhu G, Li K (2014) A unified model for community detection of multiplex networks. In: International conference on web information systems engineering. Springer. doi:10.1007/978-3-319-11749-2_3
- Tang L, Liu H (2009) Uncovering cross-dimension group structures in multi-dimensional networks. In: SDM workshop on analysis of dynamic networks. Sparks, NV
- Boden B et al (2013) RMiCS: a robust approach for mining coherent subgraphs in edge-labeled multi-layer graphs. In: Proceedings of the 25th international conference on scientific and statistical database management. ACM. doi:10.1007/s10618-012-0272-z
- Mucha PJ et al (2010) Community structure in time-dependent, multiscale, and multiplex networks. *Science* 328(5980):876–878. doi:10.1126/science.1184819
- DuBois C, Smyth P (2010) Modeling relational events via latent classes. In: Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining. ACM. doi:10.1145/1835804.1835906
- Sinkkonen J et al (2008) A simple infinite topic mixture for rich graphs and relational data. In: Proceedings of the NIPS workshop on analyzing graphs: theory and applications. Citeseer
- Xu Z, Kersting K, Tresp V (2009) Multi-relational learning with Gaussian processes. In: IJCAI. doi:10.1145/1390334.1390380
- Gershman SJ, Blei DM (2012) A tutorial on Bayesian nonparametric models. *J Math Psychol* 56(1):1–12. doi:10.1016/j.jmp.2011.08.004
- Battiston F, Nicosia V, Latora V (2014) Structural measures for multiplex networks. *Phys Rev E* 89(3):032804. doi:10.1103/PhysRevE.89.032804
- Tang L, Wang X, Liu H (2010) Community detection in multi-dimensional networks. *Computer Science and Engineering*, Arizona State University, Tempe
- Berlingerio M, Pinelli F, Calabrese F (2013) Abacus: frequent pattern mining-based community discovery in multidimensional networks. *Data Min Knowl Disc* 27(3):294–320. doi:10.1007/s10618-013-0331-0
- Wu Z et al (2013) Community detection in multi-relational social networks. In: International conference on web information systems engineering. Springer. doi:10.1007/978-3-642-41154-0_4
- Carchiolo V et al (2011) Communities unfolding in multislice networks. In: Complex networks. Springer, pp 187–195. doi:10.1007/978-3-642-25501-4_19
- Shen H-W, Cheng X-Q, Guo J-F (2011) Exploring the structural regularities in networks. *Phys Rev E* 84(5):056111. doi:10.1103/PhysRevE.84.056111
- Chen Y et al (2014) Overlapping community detection in networks with positive and negative links. *J Stat Mech Theory Exp* 2014(3):P03021. doi:10.1088/1742-5468/2014/03/P03021
- Khoshneshin M, Street N (2013) A graphical model for multi-relational social network analysis. *Relation* 4(K3):K4
- Kemp C et al (2006) Learning systems of concepts with an infinite relational model. In: AAAI. doi:10.1145/1102351.1102407
- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Ser B (Methodological)* 1–38
- Palla K, Knowles DA, Ghahramani Z (2012) An infinite latent attribute model for network data. In: International conference on machine learning
- Andrieu C et al (2003) An introduction to MCMC for machine learning. *Mach Learn* 50(1–2):5–43. doi:10.1023/A:1020281327116
- Neal RM (2003) Slice sampling. *Ann Stat*. doi:10.1214/aos/1056562461
- Rodriguez A, Laio A (2014) Clustering by fast search and find of density peaks. *Science* 344(6191):1492–1496. doi:10.1126/science.1242072
- Ana L, Jain AK (2003) Robust data clustering. In: Proceedings of 2003 IEEE computer society conference on computer vision and pattern recognition, 2003. IEEE. doi:10.1109/CVPR.2003.1211462
- Coleman J, Katz E, Menzel H (1957) The diffusion of an innovation among physicians. *Sociometry* 20(4):253–270. doi:10.2307/2785979
- Nooy W (2006) Interlock formation an actor-oriented approach. In: Politics and interlocking directorates conference. MIT Press
- Snijders TA et al (2006) New specifications for exponential random graph models. *Sociol Methodol* 36(1):99–153. doi:10.1111/j.1467-9531.2006.00176.x
- Handcock MS et al (2003) statnet: an R package for the statistical modeling of social networks. <http://www.csde.washington.edu/statnet>
- Ulanowicz RE, DeAngelis DL (2005) Network analysis of trophic dynamics in South Florida ecosystems. US Geological Survey Program on the South Florida Ecosystem, vol 114
- Jacob Y, Denoyer L, Gallinari P (2011) Classification and annotation in social corpora using multiple relations. In: Proceedings of the 20th ACM international conference on information and knowledge management. ACM. doi:10.1016/j.ins.2009.01.007