

# Copula Mixed-Membership Stochastic Blockmodel

Xuhui Fan, Richard Yi Da Xu, Longbing Cao

FEIT, University of Technology Sydney, Australia

xhfan.ml@gmail.com; {Yida.Xu;Longbing.Cao}@uts.edu.au

## Abstract

The Mixed-Membership Stochastic Blockmodels (MMSB) is a popular framework for modelling social relationships by fully exploiting each individual node's participation (or membership) in a social network. Despite its powerful representations, MMSB assumes that the membership indicators of each pair of nodes (i.e., people) are distributed independently. However, such an assumption often does not hold in real-life social networks, in which certain known groups of people may correlate with each other in terms of factors such as their membership categories. To expand MMSB's ability to model such dependent relationships, a new framework - a Copula Mixed-Membership Stochastic Blockmodel - is introduced in this paper for modeling intra-group correlations, namely an individual Copula function jointly models the membership pairs of those nodes within the group of interest. This framework enables various Copula functions to be used on demand, while maintaining the membership indicator's marginal distribution needed for modelling membership indicators with other nodes outside of the group of interest. Sampling algorithms for both the finite and infinite number of groups are also detailed. Our experimental results show its superior performance in capturing group interactions when compared with the baseline models on both synthetic and real world datasets.

## 1 Introduction

Community modeling is an important but challenging topic which has seen applications in various settings including social-media recommendation [Tang and Liu, 2010][Li et al., 2009], customer partitioning [Wang et al., 2015], discovering social networks [Fan et al., 2015][Fan et al., 2016b], and partitioning protein-protein interaction networks [Girvan and Newman, 2002][Fortunato, 2010]. Quite a few models have been proposed in the last few years to address these problems; some earlier examples include *stochastic blockmodel* [Nowicki and Snijders, 2001], and its infinite community case - *infinite relational model* (IRM) [Kemp et al., 2006], both assume that each node has one latent variable to directly indi-

cate its community membership, dictated by a single distribution of communities. Their aim is to partition a network of nodes into different communities based on the pair-wise, directional binary observations.

A typical need and challenge in community modeling is to capture the complex interactions amongst the nodes in different applications. Accordingly, several variants of IRM were proposed, including the *mixed membership stochastic blockmodel* (MMSB) [Airoldi et al., 2008], in which multiple roles (membership indicators) can possibly be played by one node. Each node has its own "membership distribution", and its relation with all other nodes is generated from it. For any two nodes, having determined their corresponding membership indicator pair, their (directional) interactions are generated from a so-called, "role-compatibility matrix" with its row and column indexed by this pair. One mentionable development of MMSB is the *nonparametric metadata dependent relational model* (NMDR) [Kim et al., 2012], which modifies MMSB by incorporating each node's metadata information into the membership distribution.

However, all of the MMSB-typed models make the assumption that, for each relation between two nodes, their corresponding membership indicator pairs were determined independently. This may limit the way membership indicators can be distributed. In fact, under many social network settings, certain known group members may have higher correlated interactions towards the ones within the same group. For instance, in a company, IT support team members tend to co-interact with each other more than with employees of other departments. Another example is that teenagers may have similar "likes" or "dislikes" on certain topics, compared with the views they may hold towards people of other age groups. MMSB-typed models overlook such interactions within a group and thus cannot fully capture the intrinsic interactions within a network.

In reality, within a social networking context, it is important to incorporate group member interactions (here called intra-group correlations) into the modeling of membership indicators. After introducing these intra-group correlations, it is important that at the same time, we do not alter membership indicators' distributions themselves, so that their interactions to people outside of the known subgroups are unaffected.

Accordingly, in this paper, a Copula function [Nelsen, 2006][McNeil and Nešlehová, 2009] is introduced to MMSB,

forming a *copula Mixed-Membership Stochastic Blockmodels* (cMMSB), for modeling the intra-group correlations. With cMMSB, we can flexibly apply various Copula functions towards different subsets of pairs of nodes while maintaining the original marginal distribution of each of the membership indicators. We develop ways in which a bivariate Copula can be used for two distributions of indicators, enjoying infinitely possible values. Under the framework, we can incorporate different choices of Copula functions to suit the need of the applications. With different Copula functions imposed on the different groups of nodes, each of the Copula function's parameters will be updated in accordance with the data. What is more, we also give two analytical solutions to calculate the conditional marginal density to the two indicator variables, which plays a crucial role in our likelihood calculation and also creates a new way of calculating a deterministic relationship between multiple variables in a graphical model.

## 2 Preliminary knowledge on Copula Model

Here we describe very briefly a bivariate copula function  $C(u, v)$ , which is a Cumulative Distribution Function over the interval  $[0, 1] \times [0, 1]$  with the uniform marginal distribution [Nelsen, 2006]. This correlation representation is extremely useful since we have the following theorem:

**THEOREM 1.** Sklar's Theorem: *Let  $X$  and  $Y$  be random variables with distribution functions  $F$  and  $G$  respectively and joint distribution function  $H$ . Then there exists a Copula  $C$  such that for all  $(x, y) \in R \times R$ :*

$$H(x, y) = C(F(x), G(y)) \quad (1)$$

$C$  is unique if  $F$  and  $G$  are continuous, then the joint probability density function is:

$$h(x, y) = c(F(x), G(y)) \cdot f(x)g(y) \quad (2)$$

Here  $c(u, v) = \partial^2 C(u, v) / \partial u \partial v$  is noted for the copula density function.

Sklar's theorem ensures the uniqueness of copula function  $C(F(x), G(y))$  once the joint distribution  $h(x, y)$  and its two marginal distributions  $f(x)$  and  $g(y)$  are known. The modification of a Copula function does not change the marginal distributions, which serves the purpose of this paper.

The popularity of copula models from various applications also meant the availability of different choices of copula functions to suit various applications. The commonly used copula function includes Gaussian Copula, Archimedean Copula (Clayton, Gumbel, Frank, etc.). We have visualized the

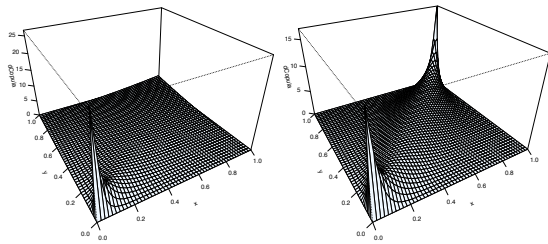


Figure 1: Clayton Copula (2) and Gaussian Copula (0.9) visualization.

probability density function for Clayton Copula and Gaussian Copula in Figure 1. For a comprehensive survey of copula functions, please refer to [Nelsen, 2006].

## 3 Copula Mixed Membership Stochastic Blockmodel (cMMSB)

### 3.1 Notations

The notations and their meanings to be used in this paper are presented in Table 1.

Table 1: Notations for cMMSB	
$n$	number of nodes
$K$	number of discovered communities
$e_{ij}$	directional, binary interactions
$\gamma, \alpha$	concentration parameters for HDP
$s_{ij}$	sender's (from $i$ to $j$ ) membership indicator
$r_{ij}$	receiver's (from $j$ to $i$ ) membership indicator
$\pi_i$	mixed-membership distribution for node $i$ , it generates $s_{i1}, \dots, s_{in}, r_{1i}, \dots, r_{ni}$
$\pi_{ik}$	the "significance" of community $k$ for node $i$
$B$	role-compatibility matrix
$B_{k,l}$	compatibilities between communities $k$ and $l$
$m_{k,l}$	number of links from community $k$ to $l$ i.e. $m_{ik} = \#\{ij : s_{ij} = k, r_{ij} = l\}$
$m_{k,l}^1$	part of $m_{k,l}$ where the corresponding $e_{ij} = 1$ i.e. $m_{k,l}^1 = \sum_{s_{ij}=k, r_{ij}=l} e_{ij}$
$m_{k,l}^0$	part of $m_{k,l}$ where the corresponding $e_{ij} = 0$ $m_{k,l}^0 = m_{k,l} - m_{k,l}^1$
$N_{ik}$	number of times that a node $i$ has participated in community $k$ (either sending or receiving) i.e. $N_{ik} = \#\{j : s_{ij} = k\} + \#\{j : r_{ji} = k\}$
$\theta$	parameter associated with any Copula function

### 3.2 Graphical Model Description

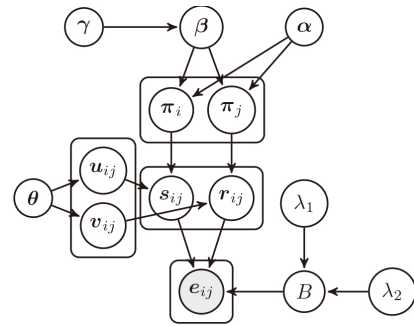


Figure 2: Graphical model of Copula MMSB

The generative process of graphical modeling is illustrated below:

$$C1: \beta \sim GEM(\gamma)$$

$$C2: \{\pi_i\}_{i=1}^n \sim DP(\alpha \cdot \beta)$$

$$\text{C3: } \begin{cases} (u_{ij}, v_{ij}) \sim \text{Copula}(\theta), & g_{ij} = 1; \\ u_{ij}, v_{ij} \sim U(0, 1), & g_{ij} = 0. \end{cases}$$

$$\text{C4: } s_{ij} = \Pi_i^{-1}(u_{ij}), r_{ij} = \Pi_j^{-1}(v_{ij})$$

$$\text{C5: } B_{k,l} \sim \text{Beta}(\lambda_1, \lambda_2), \forall k, l;$$

$$\text{C6: } e_{ij} \sim \text{Bernoulli}(B_{s_{ij}, r_{ij}}).$$

Here  $g_{ij} = 1$  in C3 denotes that the node pair  $(i, j)$  belongs to the sub-group of interest, i.e.,  $s_{ij}, r_{ij}$  are implicitly correlated, while  $g_{ij} = 0$  means  $(s_{ij}, r_{ij})$  are modelled using traditional MMSB. In C4,  $\Pi_i^{-1}(u_{ij}) = \{\min k : \sum_{q=1}^k \pi_{iq} \geq u_{ij}\}$  denotes the interval of  $\pi_i$  that  $u_{ij}$  belongs into, and similar notation is applied to  $\Pi_j^{-1}(v_{ij}) = \{\min k : \sum_{q=1}^k \pi_{jq} \geq v_{ij}\}$ .

For a simplified illustration, we divide the generative model into three sub-models: (1) “mixed membership distribution modelling”, (2) “copula incorporated membership indicator pair” and (3) “binary observation modelling”, with their details elaborated in the following sections.

### Mixed Membership Distribution Modeling

C1-C2 are for the generation of each node’s mixed membership distribution. The number of communities, i.e.,  $k$  is an important factor in mixed membership distribution models. Therefore, we consider two possibilities here. The first is to use a fixed  $k$ . As the graphical model in Fig. 2 shows, for all the mixed-membership distributions  $\{\pi_i\}_{i=1}^n$ , there is a common parent node  $\beta$ , where  $\beta$  typically has a “non-informative” symmetric Dirichlet prior, i.e.,  $(\beta_1, \dots, \beta_k) \sim \text{Dir}(\gamma, \dots, \gamma)$  [Airoldi et al., 2008]. The appropriate choice of  $k$  is determined by the model selection method, such as the BIC criterion [Schwarz, 1978], which is commonly used in [Airoldi et al., 2008][Xing et al., 2010][Fan et al., 2016a].

The second solution is applicable for the uncertain number of communities, which is often the case under many social network settings. The usual approach is to use the Hierarchical Dirichlet Process (HDP) [Teh et al., 2006] prior with  $\beta$  distributed from a GEM( $\gamma$ ), i.e.,  $\beta$  is obtained via a stick-breaking construction [Sethuraman, 1991] with each of its components  $\beta_k = u_k \prod_{l=1}^{k-1} (1 - u_l), u_l \sim \text{Beta}(1, \gamma)$ .

After obtaining their parent’s node  $\beta$ , we can sample our mixed-membership distribution  $\{\pi_i\}$  independently from [Airoldi et al., 2008][Koutsourelakis and Eliassi-Rad, 2008]:

$$\pi_i \sim \begin{cases} \text{Dir}(\alpha \cdot \beta), & \text{fixed } k; \\ \text{DP}(\alpha \cdot \beta), & \text{uncertain } k. \end{cases} \quad \text{For the notational clarity, we concentrate our discussion on the uncertain } k \text{ case without delicately mentioning its finite counterpart, as the finite } k \text{ case can be trivially derived.}$$

### Copula Incorporated Membership Indicator Pair

Our main work of c-MMSB is displayed in phases C3-C4. We consider two cases in this paper for the intra-group correlation modeling: full correlation and partial correlation.

**Full correlation:** i.e., intra-group correlation for all the nodes. We assume each pair of nodes, i.e., all relations of the entire population are using the same Copula function. As we will see in the experimental section that, flexible modelling can still be achieved under this assumption, as parameters of a Copula can vary to support various form of relations.

**Partial correlation:** i.e., intra-group correlation are applied to only a subset of the nodes. With a definition of one subgroup, we use the Copula function on this specific subgroup and the others remain unchanged.

For traditional MMSB, the corresponding membership indicators within one pair  $(s_{ij}, r_{ij})$  are independently sampled from their membership distributions, i.e.,  $s_{ij} \sim \pi_i, r_{ij} \sim \pi_j$ . Using the definition of  $\{\Pi_i^{-1}(\cdot)\}_{i=1}^n$  from Section 3.2, this is equivalently expressed as:

$$\begin{aligned} u_{ij} &\sim U(0, 1), v_{ij} \sim U(0, 1); \\ s_{ij} &= \Pi_i^{-1}(u_{ij}), r_{ij} = \Pi_j^{-1}(v_{ij}). \end{aligned} \quad (3)$$

As discussed in the introduction, we are motivated by examples within social network settings, in which membership indicators from a node may well be correlated with other membership indicators in an intra-group point of view. People’s interactions with each other within the group may more likely (or less likely) belong to the same category, i.e.,  $(s_{ij}, r_{ij})$  has higher (or lower) density in some regions of the discrete space  $(1, 2, \dots, \infty)^2$ , which may not be well described by using only the two independent marginal distributions.

We propose a general framework by employing a Copula function to depict the correlation within the membership indicator pair. This is accomplished by the joint sampling of uniform variables  $(u_{ij}, v_{ij})$  (in Eq. (3).) from the Copula function, instead of from two independent uniform distributions. More precisely, the membership indicator pair is obtained using:

$$\begin{aligned} \forall g_{ij} = 1 : (u_{ij}, v_{ij}) &\sim \text{Copula}(u, v | \theta); \\ s_{ij} &= \Pi_i^{-1}(u_{ij}), r_{ij} = \Pi_j^{-1}(v_{ij}). \end{aligned} \quad (4)$$

Using various Copula priors over the pair  $(u_{ij}, v_{ij})$ , we are able to more appropriately express the way in which the membership indicator pair  $\{s_{ij}, r_{ij}\}$  is distributed, given the different scenarios we are facing. Taking the Gumbel Copula (with larger parameter values) [Nelsen, 2006] as an instance, for certain membership indicator pairs ( $g_{ij} = 1$ ), it generates  $(u_{ij}, v_{ij})$  values that more likely have positive correlation, i.e., within  $[0, 1]^2$  space, which promotes  $s_{ij} = r_{ij}$ . Also, the Gaussian Copula ( $\theta = -1$ ) encourages the  $(s_{ij}, r_{ij})$  pair to be different.

### Binary Observation Modeling

C5-C6 model the binary observation, which directly follows the previous work [Nowicki and Snijders, 2001][Kemp et al., 2006] etc. Due to the beta-bernoulli conjugacy,  $B$  can be marginalized out and the likelihood of binary observation becomes as follows:

$$\Pr(e|z, \lambda_1, \lambda_2) = \prod_{k,l} \frac{\text{beta}(m_{k,l}^1 + \lambda_1, m_{k,l}^0 + \lambda_2)}{\text{beta}(\lambda_1, \lambda_2)} \quad (5)$$

here  $\text{beta}(\lambda_1, \lambda_2)$  denotes the beta function with parameters  $\lambda_1$  and  $\lambda_2$ ,  $m_{k,l}^1$  and  $m_{k,l}^0$  are defined in Table 1.

## 4 Inference & Further Discussion

Let  $K$  be the discovered number of communities, a formal and concise representation of Eq. (4), i.e. the probability of

$(s_{ij}, r_{ij})$ , is:

$$\begin{aligned} \Pr(s_{ij}, r_{ij}) &= \int_{\sum_{d=1}^{K+1} \pi_{jd}=1} \int_{\sum_{d=1}^{K+1} \pi_{id}=1} \int_{(u_{ij}, v_{ij})} \\ &\cdot \mathbf{1}(s_{ij} = \Pi_i^{-1}(u_{ij}), r_{ij} = \Pi_j^{-1}(v_{ij})) \\ &\cdot dC(u_{ij}, v_{ij}) dF(\pi_{i1}, \dots, \pi_{iK+1}) dF(\pi_{j1}, \dots, \pi_{jK+1}) \end{aligned} \quad (6)$$

Unfortunately, we cannot bring  $\Pr(s_{ij}, r_{ij})$  to an analytical form without any integrals present. However, with some mathematical design, we found that, conditioning on the explicit sample of either  $(u_{ij}, v_{ij})$  or  $(\pi_i, \pi_j)$ , it is possible to obtain a marginalised conditional density in which  $s_{ij}, r_{ij}$  is conditioned on either  $(u_{ij}, v_{ij})$  or  $(\pi_i, \pi_j)$ , but not both. Additionally, having a set of variables “collapsed” from the Gibbs sampling, it results in a faster mixing on Markov chains [Liu, 1994]. Therefore, two corresponding inference schemes are needed. To be more concentrated, we present the key parts of these two inference algorithms here (the rest follows the standard procedures as in [Fox et al., 2008]), and name them *Marginal conditional on  $\pi$  only* method and the *Marginal conditional on  $u, v$  only* respectively:

#### 4.1 Marginal Conditional on $\pi$ only: cMMSB $^\pi$

In the *Marginal conditional on  $\pi$  only* (cMMSB $^\pi$  for short) method, the variables of interest include  $\{\pi_i\}, \{s_{ij}, r_{ij}\}, \beta$ . As mentioned before, we describe the formulation using the infinite communities (uncertain  $k$ ) case only, its counterpart in the finite communities (fixed  $k$ ) case can be trivially derived.

##### Sampling $\pi_i$

When a Copula is introduced,  $p(\pi_i)$  and  $\Pr(s_{ij}|\pi_i)$  are no longer a conjugate pair. Therefore, we resort to the use of Metropolis-Hastings (M-H) Sampling in each  $(\tau)$ -th MCMC iteration.

For each node  $i$ ,  $\pi_i$ 's posterior distribution is formed as Eq. (7), where  $p_{ij}^{s_{ij}r_{ij}}(\pi_i, \pi_j)$  is defined in Eq. (4).

$$\begin{aligned} p(\pi_i|\alpha, \beta, \{s_{ij}, r_{ij}\}_{i,j}) \\ \propto \prod_{k=1}^{K+1} \pi_{ik}^{\alpha\beta_k-1} \cdot \prod_{j=1}^n [p_{ij}^{s_{ij}r_{ij}}(\pi_i, \pi_j) p_{ji}^{s_{ji}r_{ji}}(\pi_j, \pi_i)] \end{aligned} \quad (7)$$

The Corresponding proposal distribution of  $\pi_i$  for the above M-H is a posterior Dirichlet distribution in the form of (i.e.,  $\pi_i$ 's posterior distribution under the MMSB framework):

$$q(\pi_i^*|\alpha, \beta, \{s_{ij}, r_{ij}\}_{i,j}) \propto \prod_{k=1}^{K+1} [\pi_{ik}^*]^{\alpha\beta_k + N_{ik} - 1} \quad (8)$$

Then the acceptance ratio becomes:

$$A(\pi_i^*, \pi_i^{(\tau)}) = \min(1, a) \quad (9)$$

$$\begin{aligned} a &= \frac{\prod_{j=1}^n [p_{ij}^{s_{ij}r_{ij}}(\pi_i^*, \pi_j) p_{ji}^{s_{ji}r_{ji}}(\pi_j, \pi_i^*)]}{\prod_{j=1}^n [p_{ij}^{s_{ij}r_{ij}}(\pi_i^{(\tau)}, \pi_j) p_{ji}^{s_{ji}r_{ji}}(\pi_j, \pi_i^{(\tau)})]} \\ &\cdot \frac{\prod_{k=1}^{K+1} [\pi_{ik}^{(\tau)}]^{N_{ik}}}{\prod_{k=1}^{K+1} [\pi_{ik}^*]^{N_{ik}}} \end{aligned} \quad (10)$$

##### Sampling $s_{ij}$ and $r_{ij}$

As  $e_{ij}$  is dependent on both  $\{s_{ij}, r_{ij}\}$ , a joint sampling of  $\{s_{ij}, r_{ij}\}$  is implemented as:

$$\begin{aligned} \Pr(s_{ij}, r_{ij}|e_{ij}, \lambda_1, \lambda_2, \theta, \pi_i, \pi_j, m_{s_{ij}, r_{ij}}^{-e_{ij}}) \\ \propto \Pr(s_{ij}, r_{ij}|\pi_i, \pi_j, \theta) \cdot \Pr(e_{ij}|s_{ij}, r_{ij}, \lambda_1, \lambda_2, m_{s_{ij}, r_{ij}}^{-e_{ij}}) \end{aligned} \quad (11)$$

On the likelihood, we have

$$\Pr(e_{ij}|s_{ij}, r_{ij}, \lambda_1, \lambda_2, m_{s_{ij}, r_{ij}}^{-e_{ij}}) = \begin{cases} m_{s_{ij}, r_{ij}}^{1, -e_{ij}} + \lambda_1, & e_{ij} = 1; \\ m_{s_{ij}, r_{ij}}^{0, -e_{ij}} + \lambda_2, & e_{ij} = 0. \end{cases} \quad (12)$$

where  $m_{k,l} = \sum_{i',j'} \mathbf{1}(s_{i'j'} = k, r_{i'j'} = l)$ ,  $m_{k,l}^1 = \sum_{s_{i'j'}=k, r_{i'j'}=l} e_{i'j'}$ , and  $m_{k,l}^0 = m_{k,l} - m_{k,l}^1$ .

On the first term of the r.h.s. in Eq. (11), we define  $p_{ij}^{kl}(\pi_i, \pi_j) \equiv \Pr(s_{ij} = k, r_{ij} = l|\pi_i, \pi_j, \theta)$ ,  $\forall g_{ij} = 1$ , and let  $C(u_{ij}, v_{ij}|\theta)$  be the chosen Copula cumulative distribution function (c.d.f.) with parameter  $\theta$ . Given the explicit values of  $\pi_i, \pi_j$ , we can integrate over all  $u_{ij}, v_{ij}$  to compute the probability mass of the indicator pair  $(s_{ij} = k, r_{ij} = l)$ ,  $k, l \in \{1, \dots, K+1\}$ :

$$\begin{aligned} p_{ij}^{kl}(\pi_i, \pi_j) &= \int_{\hat{\pi}_i^{k-1}}^{\hat{\pi}_i^k} \int_{\hat{\pi}_j^{l-1}}^{\hat{\pi}_j^l} dC(u, v|\theta) \\ &= C(\hat{\pi}_i^k, \hat{\pi}_j^l) + C(\hat{\pi}_i^{k-1}, \hat{\pi}_j^{l-1}) - C(\hat{\pi}_i^k, \hat{\pi}_j^{l-1}) - C(\hat{\pi}_i^{k-1}, \hat{\pi}_j^l) \end{aligned} \quad (13)$$

$$\text{Here } \hat{\pi}_i^k = \begin{cases} 0, & k = 0; \\ \sum_{q=1}^k \pi_{iq}, & k > 0 \end{cases}$$

Since  $\{\pi_i\}_{i=1}^n$  are piecewise functions, we can easily calculate the probability mass in this “rectangular” area. In other cases of  $\{g_{ij} = 0\}$ , i.e., interaction data  $e_{ij}$  falls outside of the correlated relation group, we have  $p_{ij}^{kl}(\pi_i, \pi_j) = \pi_{ik}\pi_{jl}$ .

It is noted that, using the properties of a Copula function, the marginal distributions of  $\Pr(s_{ij} = k, r_{ij} = l|\pi_i, \pi_j, \theta)$  remain  $\pi_i$  and  $\pi_j$  respectively, which becomes that of:

$$\begin{aligned} \sum_{l=1}^{K+1} \Pr(s_{ij} = k, r_{ij} = l|\pi_i, \pi_j, \theta) &= \pi_{ik}; \\ \sum_{k=1}^{K+1} \Pr(s_{ij} = k, r_{ij} = l|\pi_i, \pi_j, \theta) &= \pi_{jl}. \end{aligned} \quad (14)$$

#### 4.2 Marginal Conditional on $u$ and $v$ only: cMMSB $^{uv}$

In *Marginal conditional on  $u, v$  only* method (cMMSB $^{uv}$  for short), the variables of interest include  $\{u_{ij}, v_{ij}\}, \{s_{ij}, r_{ij}\}, \beta$ , and an auxiliary variable  $\mathbf{m}$ .

##### Sampling $u_{ij}$ and $v_{ij}$

We have used the M-H Sampling for  $(u_{ij}, v_{ij}), \forall i, j \in \{1, \dots, n\}$ , due to the nonconjugacy issue. The Copula function is used as its proposal, and therefore, its corresponding acceptance ratio becomes that of:

$$A((u_{ij}^{(\tau)}, v_{ij}^{(\tau)}), (u_{ij}^*, v_{ij}^*)) = \min(1, a) \quad (15)$$

$$a = \frac{I_{u_{ij}^*}(h_i^{k-1}, \hat{h}_i^{k-1}) - I_{u_{ij}^*}(h_i^k, \hat{h}_i^k)}{I_{u_{ij}^{(\tau)}}(h_i^{k-1}, \hat{h}_i^{k-1}) - I_{u_{ij}^{(\tau)}}(h_i^k, \hat{h}_i^k)} \cdot \frac{I_{v_{ij}^*}(h_j^{l-1}, \hat{h}_j^{l-1}) - I_{v_{ij}^*}(h_j^l, \hat{h}_j^l)}{I_{v_{ij}^{(\tau)}}(h_j^{l-1}, \hat{h}_j^{l-1}) - I_{v_{ij}^{(\tau)}}(h_j^l, \hat{h}_j^l)} \quad (16)$$

Here  $h_i^k, \hat{h}_i^k$ 's definitions are the same as in Eq. (7), assuming  $s_{ij} = k, r_{ij} = l$ .

#### Sampling $s_{ij}$ and $r_{ij}$

An alternative ‘‘collapsed’’ sampling method is to integrate over  $\{\pi_i\}_{i=1}^n$  while we explicitly sample the values of  $\{(u_{ij}, v_{ij})\}_{i,j}$ .

Similar as Eq. (11), we obtain:

$$\begin{aligned} & \Pr(s_{ij} = k, r_{ij} = l) \\ & \propto \Pr(e_{ij}, \lambda_1, \lambda_2, m_{k,l}, u_{ij}, v_{ij}, \{h_i^k\}_k, \{\hat{h}_i^k\}_k, \{h_j^l\}_l, \{\hat{h}_j^l\}_l) \\ & \propto \Pr(s_{ij} = k | u_{ij}, \{h_i^k\}_k, \{\hat{h}_i^k\}_k) \\ & \quad \cdot \Pr(r_{ij} = l | v_{ij}, \{h_j^l\}_l, \{\hat{h}_j^l\}_l) \cdot \Pr(e_{ij} | \lambda_1, \lambda_2, m_{k,l}) \\ & \propto (I_{u_{ij}}(h_i^{k-1}, \hat{h}_i^{k-1}) - I_{u_{ij}}(h_i^k, \hat{h}_i^k)) \\ & \quad \cdot (I_{v_{ij}}(h_j^{l-1}, \hat{h}_j^{l-1}) - I_{v_{ij}}(h_j^l, \hat{h}_j^l)) \cdot \Pr(e_{ij} | \lambda_1, \lambda_2, m_{k,l}) \end{aligned} \quad (17)$$

From Eq. (4), given  $\{(u_{ij}, v_{ij})\}_{i,j}$ 's values, the probabilities  $s_{ij} = k$  and  $r_{ij} = l$  can be computed independently. The Copula function leaves marginal distributions of  $s_{ij}$  and  $r_{ij}$  invariant, which remains the same as the classical MMSB, i.e.,  $\pi_i | \alpha, \beta, \{N_{ik}^{-ij}\}_{k=1}^K \sim \text{Dir}(\alpha\beta_1 + N_{i1}^{-ij}, \dots, \alpha\beta_K + N_{iK}^{-ij}, \alpha\beta_{K+1})$ . Therefore, having the knowledge of  $F(\pi_i | \alpha, \beta, \{N_{ik}^{-ij}\}_{k=1}^K)$ , given  $u_{ij}$ , our calculation of  $\Pr(s_{ij} = k)$  is equal to computing the probability of  $u_{ij}$  falling in  $\pi_i$ 's  $k^{\text{th}}$  interval, i.e.  $\Pr(\sum_{d=1}^{k-1} \pi_{id} \leq u_{ij} < \sum_{d=1}^k \pi_{id})$  (similar case with  $v_{ij} < \pi_{jl}$ ). This can be obtained from the fact that the set  $\{u_{ij} \in [0, 1] | \sum_{d=1}^{k-1} \pi_{id} \leq u_{ij}\}$  can be decomposed into two disjoint sets:

$$\begin{aligned} & \{u_{ij} \in [0, 1] | \sum_{d=1}^{k-1} \pi_{id} \leq u_{ij}\} \\ & = \{u_{ij} \in [0, 1] | \sum_{d=1}^{k-1} \pi_{id} \leq u_{ij} < \sum_{d=1}^k \pi_{id}\} \\ & \cup \{u_{ij} \in [0, 1] | \sum_{d=1}^k \pi_{id} \leq u_{ij}\} \end{aligned} \quad (18)$$

where  $\sum_{d=1}^k \pi_{id} \sim \text{Beta}(\sum_{d=1}^k \alpha\beta_d + N_{id}, \sum_{d=k+1}^{K+1} \alpha\beta_d + N_{id})$ . (A similar result was also found in page 10 of [Teh et al., 2006]). Therefore, we have:

$$\begin{aligned} & \Pr(\sum_{d=1}^{k-1} \pi_{id} \leq u_{ij} < \sum_{d=1}^k \pi_{id}) \\ & = \Pr(\sum_{d=1}^{k-1} \pi_{id} \leq u_{ij}) - \Pr(\sum_{d=1}^k \pi_{id} \leq u_{ij}) \\ & = I_{u_{ij}}(h_i^{k-1}, \hat{h}_i^{k-1}) - I_{u_{ij}}(h_i^k, \hat{h}_i^k) \end{aligned} \quad (19)$$

Here  $h_i^k = \sum_{d=1}^k \alpha\beta_d + N_{id}$ ,  $\hat{h}_i^k = \sum_{d=k+1}^{K+1} \alpha\beta_d + N_{id}$ ;  $I_u(a, b)$  denotes the Beta c.d.f. value with parameter  $a, b$  on  $u$ . The existence and non-negativity of  $I_{u_{ij}}(u_{k-1}, \hat{u}_{k-1}) - I_{u_{ij}}(u_k, \hat{u}_k)$  is guaranteed by the fact that  $\{u_{ij} \in [0, 1] | \sum_{d=1}^{k-1} \pi_{id} \leq u_{ij}\} \subseteq \{u_{ij} \in [0, 1] | \sum_{d=1}^{k-1} \pi_{id} \leq u_{ij}\}$  on the same  $\pi_i$ .

### 4.3 Computational Complexity Analysis

We estimate the computational complexity for each graphical model and present the result in Table 2. Compared to the classical models (especially the MMSB), our cMMSB $^{\pi}$  involves an additional  $\mathcal{O}(Kn)$  term which refers to the sampling of the mixed membership distributions. Note that the computational time varies for different Copulas. cMMSB $^{uv}$  requires an extra  $\mathcal{O}(n^2)$  term for the  $u, v$ 's sampling for each membership indicator. Each operation requires a Beta c.d.f. in a tractable form. In the experimental part, we have observed that our model runs slower (in a linear way) than the original MMSB. The reason might be the additional calculation of the Copula function or the Beta c.d.f.

Table 2: Computational Complexity for Different Models

Models	Computational Complexity
IRM	$\mathcal{O}(K^2n)$ [Palla et al., 2012]
LFRM	$\mathcal{O}(K^2n^2)$ [Palla et al., 2012]
MMSB	$\mathcal{O}(Kn^2)$ [Kim et al., 2012]
cMMSB $^{\pi}$	$\mathcal{O}(Kn^2 + Kn) = \mathcal{O}(Kn^2)$
cMMSB $^{uv}$	$\mathcal{O}(Kn^2 + n^2) = \mathcal{O}(Kn^2)$

## 5 Experiments

Here, our cMMSB's performance is compared with the classical *Mixed-Membership Stochastic Blockmodels* (MMSB)-type methods, including the original MMSB [Airoldi et al., 2008] and the *infinite mixed-membership model* (iMMM) [Koutsourelakis and Eliassi-Rad, 2008]. Additionally, we also compare it with other non-MMSB approaches including the *infinite relational model* (IRM) [Kemp et al., 2006], the *latent feature relational model* (LFRM) [Miller et al., 2009] and the *nonparametric metadata dependent relational model* (NMDR) [Kim et al., 2012].

We independently implement the above benchmark algorithms to the best of our understanding. In order to provide a common ground for all comparisons, we make the following small variations to these algorithms: (1) In iMMM, instead of having an individual  $\alpha_i$  value for each  $\pi_i$  as used in the original work, we use a common  $\alpha$  value for all the mixed-membership distributions  $\{\pi_i\}_{i=1}^n$ ; (2) In LFRM [Miller et al., 2009]'s implementation, we do not incorporate the metadata information into the interaction data's generation, but use only the binary interaction information.

### 5.1 Real-world Datasets for Link Prediction

We analyse three real-world datasets: the NIPS Co-authorship dataset, the MIT Reality Mining dataset [Eagle and (Sandy) Pentland, 2006] and the Lazega-lawfirm dataset [Lazega, 2001].

Table 3: Model Performance (Mean  $\pm$  Standard Deviation) on Real-world Datasets.

Dataset		Train error	Test error	Test log likelihood	AUC
NIPS co-author	IRM	0.0317 $\pm$ 0.0004	0.0423 $\pm$ 0.0014	-135.0467 $\pm$ 7.3816	0.8901 $\pm$ 0.0162
	LFRM	0.0473 $\pm$ 0.0794	0.0540 $\pm$ 0.0735	-105.2166 $\pm$ 179.5505	0.9348 $\pm$ 0.1667
	MMSB	0.0132 $\pm$ 0.0042	0.0301 $\pm$ 0.0064	-86.2134 $\pm$ 10.1258	0.9524 $\pm$ 0.0215
	iMMM	<b>0.0061 <math>\pm</math> 0.0019</b>	0.0253 $\pm$ 0.0035	-83.4264 $\pm$ 9.4293	0.9574 $\pm$ 0.0155
	cMMSB $^{\pi}$	0.0066 $\pm$ 0.0038	<b>0.0231 <math>\pm</math> 0.0043</b>	-83.4261 $\pm$ 9.4280	0.9569 $\pm$ 0.0159
	cMMSB $^{uv}$	0.0097 $\pm$ 0.0047	0.0240 $\pm$ 0.0065	<b>-83.4257 <math>\pm</math> 9.4292</b>	<b>0.9581 <math>\pm</math> 0.0153</b>
MIT reality	IRM	0.0627 $\pm$ 0.0002	0.0665 $\pm$ 0.0004	-133.8037 $\pm$ 1.1269	0.8261 $\pm$ 0.0047
	LFRM	0.0397 $\pm$ 0.0017	0.0629 $\pm$ 0.0037	-143.6067 $\pm$ 10.0592	0.8529 $\pm$ 0.0179
	MMSB	0.0263 $\pm$ 0.0105	0.0716 $\pm$ 0.0043	-129.4354 $\pm$ 7.6549	0.8561 $\pm$ 0.0176
	iMMM	0.0297 $\pm$ 0.0055	0.0625 $\pm$ 0.0015	-126.7876 $\pm$ 3.4774	0.8617 $\pm$ 0.0124
	NMDR	0.0386 $\pm$ 0.0040	0.0668 $\pm$ 0.0013	-139.5227 $\pm$ 2.9371	0.8569 $\pm$ 0.0138
	cMMSB $^{\pi}$	<b>0.0246 <math>\pm</math> 0.0016</b>	0.0489 $\pm$ 0.0016	-125.3876 $\pm$ 3.2689	<b>0.8794 <math>\pm</math> 0.0159</b>
Lazega lawfirm	cMMSB $^{uv}$	0.0283 $\pm$ 0.0035	<b>0.0438 <math>\pm</math> 0.0015</b>	<b>-123.3876 <math>\pm</math> 3.1254</b>	0.8738 $\pm$ 0.0364
	IRM	0.0987 $\pm$ 0.0003	0.1046 $\pm$ 0.0012	-201.7912 $\pm$ 3.3500	0.7056 $\pm$ 0.0167
	LFRM	0.0566 $\pm$ 0.0024	0.1051 $\pm$ 0.0064	-222.5924 $\pm$ 16.1985	0.8170 $\pm$ 0.0197
	MMSB	0.0391 $\pm$ 0.0071	0.0913 $\pm$ 0.0030	-212.1256 $\pm$ 3.2145	0.7989 $\pm$ 0.0102
	iMMM	0.0487 $\pm$ 0.0068	0.1096 $\pm$ 0.0026	-202.7148 $\pm$ 5.3076	0.8074 $\pm$ 0.0141
	NMDR	0.0640 $\pm$ 0.0055	0.1133 $\pm$ 0.0018	-207.7188 $\pm$ 3.4754	0.8285 $\pm$ 0.0114
	cMMSB $^{\pi}$	<b>0.0246 <math>\pm</math> 0.0050</b>	<b>0.1023 <math>\pm</math> 0.0056</b>	<b>-201.0154 <math>\pm</math> 5.2167</b>	<b>0.8273 <math>\pm</math> 0.0148</b>
	cMMSB $^{uv}$	0.0276 $\pm$ 0.0043	0.1143 $\pm$ 0.0019	-204.0289 $\pm$ 9.5460	0.8215 $\pm$ 0.0167

### NIPS Co-authorship Dataset

We use the co-authorship as a relation from the proceeding of the *Neural Information Processing Systems* (NIPS) conference for the years 2000-2012. Due to the sparse nature of the co-authorships, we observe the authors' activities in all the 13 years (i.e. regardless of the time factor) and set the relational data to 1 if the two corresponding authors have co-authored for no less than 2 papers, which remove some of the "by chance" co-authorships. Further, the author with less than 4 relationships with others are considered "inactive" and hence have been manually removed. Thus, a  $92 \times 92$  symmetric and binary matrix is obtained.

On this dataset, no pre-defined group information is obtained in advance. Thus, we consider it as full-correlation case and use one Gumbel Copula function to model all the interactions.

### MIT Reality Dataset

From the MIT Reality Mining [Eagle and (Sandy) Pentland, 2006], we use the subjects' proximity dataset, where weighted links indicate the average proximity from one subject to another at work. We then "binarize" the data, in which we set the proximity value larger than 10 minutes per day as 1, and 0 otherwise. Therefore, a  $94 \times 94$  asymmetric and binary matrix is obtained.

The dataset are roughly divided into four groups: Sloan Business School students (Sloan), lab faculty, senior students with more than 1 year in the lab and junior students. In our experiment, we only apply the Gumbel Copula function to the Sloan portion of the students to encourage similar mixture membership indicators.

### Lazega Law Dataset

The lazega-lawfirm dataset [Lazega, 2001] is obtained from a social network study of corporate located in the north-eastern

part of U.S. in 1988 - 1991. The dataset contains three different types of relations: co-work network, basic advice network and friendship network, among the 71 attorneys, of which the element are labeled as 1 (exist) or 0 (absent).

Since no group information is obtained in this dataset, we use the same setting as in NIPS co-authorship dataset as one Gumbel Copula function is used for all the interactions.

### General Performance

From these reported statistics as shown in Table 3, we can see that our methods (cMMSB $^{\pi}$ , cMMSB $^{uv}$ ) obtain the best performance in these 3 datasets, amongst all other models. Although iMMM can achieve smallest train error in the NIPS co-author dataset, our cMMSB's predictability is better than iMMM and the others. On the MIT reality and Lazega-lawfirm datasets, our cMMSB can achieve at least 1% improvement on the AUC score. On the performance comparison of our two different sampling schemes cMMSB $^{\pi}$  and cMMSB $^{uv}$ , we find they achieve similar results, which is within our expectation. Our cMMSB $^{\pi}$ , cMMSB $^{uv}$  beat both MMSB-like models and non-MMSB models since a hidden intra-group correlation has been adaptively utilized here.

## 6 Conclusions

The principal contribution of our proposed model is the introduction of the Copula function into MMSB, which represents the correlation between the pair of membership indicators, while keeping the membership indicators' marginal distribution invariant. The results show that, using both synthetic and real data, our Copula-incorporated MMSB, i.e., cMMSB, is effective in learning the community structure and predicting the missing links.

## References

- [Airoldi et al., 2008] Airoldi, E., Blei, D., Fienberg, S., and Xing, E. (2008). Mixed-Membership Stochastic Block-models. *The Journal of Machine Learning Research*, 9:1981–2014.
- [Eagle and (Sandy) Pentland, 2006] Eagle, N. and (Sandy) Pentland, A. (2006). Reality mining: sensing complex social systems. *Personal Ubiquitous Comput.*, 10(4):255–268.
- [Fan et al., 2015] Fan, X., Cao, L., and Xu, R. Y. (2015). Dynamic Infinite Mixed-Membership Stochastic Block-model. *Neural Networks and Learning Systems, IEEE Transactions on*, 26(9):2072–2085.
- [Fan et al., 2016a] Fan, X., Li, B., Wang, Y., Wang, Y., and Chen, F. (2016a). The Ostomachion Process. In *AAAI Conference on Artificial Intelligence*.
- [Fan et al., 2016b] Fan, X., Xu, R. Y., Cao, L., and Song, Y. (2016b). Learning nonparametric relational models by conjugately incorporating node information in a network. *IEEE Transactions on Cybernetics*, PP(99):1–11.
- [Fortunato, 2010] Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3):75–174.
- [Fox et al., 2008] Fox, E. B., Sudderth, E. B., Jordan, M. I., and Willsky, A. S. (2008). An hdp-hmm for systems with state persistence. In *Proceedings of the 25th International Conference on Machine Learning (ICML)*, pages 312–319, New York, NY, USA. ACM.
- [Girvan and Newman, 2002] Girvan, M. and Newman, M. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826.
- [Kemp et al., 2006] Kemp, C., Tenenbaum, J. B., Griffiths, T. L., Yamada, T., and Ueda, N. (2006). Learning systems of concepts with an infinite relational model. In *Proceedings of the 21st AAAI Conference on Artificial Intelligence (AAAI)*, volume 3, pages 381–388.
- [Kim et al., 2012] Kim, D. I., Hughes, M., and Sudderth, E. B. (2012). The nonparametric metadata dependent relational model. In Langford, J. and Pineau, J., editors, *Proceedings of the 29th International Conference on Machine Learning (ICML)*, pages 1559–1566, New York, NY, USA. ACM.
- [Koutsourelakis and Eliassi-Rad, 2008] Koutsourelakis, P. and Eliassi-Rad, T. (2008). Finding mixed-memberships in social networks. In *Proceedings of the 2008 AAAI spring symposium on social information processing*, pages 48–53.
- [Lazega, 2001] Lazega, E. (2001). *The collegial phenomenon: The social mechanisms of cooperation among peers in a corporate law partnership*. Oxford University Press on Demand.
- [Li et al., 2009] Li, B., Yang, Q., and Xue, X. (2009). Transfer learning for collaborative filtering via a rating-matrix generative model. In *Proceedings of the 26th Annual International Conference on Machine Learning (ICML)*, pages 617–624.
- [Liu, 1994] Liu, J. S. (1994). The collapsed gibbs sampler in bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, 89(427):958–966.
- [McNeil and Nešlehová, 2009] McNeil, A. J. and Nešlehová, J. (2009). Multivariate archimedean copulas, d-monotone functions and  $l_1$ -norm symmetric distributions. *The Annals of Statistics*, pages 3059–3097.
- [Miller et al., 2009] Miller, K., Griffiths, T., and Jordan, M. (2009). Nonparametric latent feature models for link prediction. *Advances in Neural Information Processing Systems*, 22:1276–1284.
- [Nelsen, 2006] Nelsen, R. (2006). *An introduction to copulas*. Springer.
- [Nowicki and Snijders, 2001] Nowicki, K. and Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455):1077–1087.
- [Palla et al., 2012] Palla, K., Ghahramani, Z., and Knowles, D. A. (2012). An infinite latent attribute model for network data. In Langford, J. and Pineau, J., editors, *Proceedings of the 29th International Conference on Machine Learning (ICML)*, pages 1607–1614. ACM, New York, NY, USA.
- [Schwarz, 1978] Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.
- [Sethuraman, 1991] Sethuraman, J. (1991). A constructive definition of dirichlet priors. Technical report, DTIC Document.
- [Tang and Liu, 2010] Tang, L. and Liu, H. (2010). Community detection and mining in social media. *Synthesis Lectures on Data Mining and Knowledge Discovery*, 2(1):1–137.
- [Teh et al., 2006] Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2006). Hierarchical dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581.
- [Wang et al., 2015] Wang, Y., Li, B., Wang, Y., and Chen, F. (2015). Metadata Dependent Mondrian processes. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, pages 1339–1347.
- [Xing et al., 2010] Xing, E., Fu, W., and Song, L. (2010). A state-space mixed membership blockmodel for dynamic network tomography. *The Annals of Applied Statistics*, 4(2):535–566.