Particle Mixed Membership Stochastic Block Model

Yan Zhang $^1$, Qixia Jiang $^2$, Maosong Sun $^3$

State Key Laboratory on Intelligent Technology and Systems
Tsinghua National Laboratory for Information Science and Technology
Department of Computer Sci. and Tech., Tsinghua University, Beijing 100084, China
$^1$zhang-y-05@mails.tsinghua.edu.cn  $^2$qixia.jiang@gmail.com  $^3$sms@mail.tsinghua.edu.cn

Abstract—Massive real-world data are network-structured, such as semantic web, social network, relationship between proteins, etc. Modeling a network is an effective way for better understanding the properties of a network, while avoiding the complexity of the full description. This paper proposes a novel hierarchical Bayesian model for relational data, which is an extension of Mixed Membership Stochastic Block model. Unlike previous work assumes edges to be of atomic, our model recognizes that each edge is composed of multiple elementary relationships and the weight on this edge is a includes all the weights of these elementary relationships. This allows our model to incorporate more information about the network and increases its ability of uncertainty tolerance. A fast inference based on variational inference is offered. Empirical results on a synthetic data and three real-world data sets demonstrate the effectiveness and the robustness of our method.

I. INTRODUCTION

Many network-structured data involve massive relational information among objects. This information is quite useful for inferring the latent structure of networks, learning the interaction patterns among objects, analyzing the importance of relations, etc.

Therefore, some work analyzes the the average behavior of networks, such as degree distribution, diameter and so on [1-4]. These characteristics are useful to describe networks but not enough to illustrate their whole structures and properties. Some other work tries to measure the relative importance of each element within a network, such as PageRank [5] and HITS [6]. These approaches assume that networks are without omissions, which is always violated in real applications. On the other hand, Hoff et al. [7], Snijers et al. [8], etc., model the conditional probability, the probability of a relation between objects depending on individual properties and the local information. Such approaches suffer from the inferential degeneracy [9] and scalability [7, 10].

Motivated by these problems, Stochastic Block Model (SBM) [11-13] is proposed. SBM is a kind of probabilistic generative models and tries to assign one object to one block. SBM is recently extended to model the mixed membership of each object [14-17]. In these approaches, each edge is considered to be of atomic (atomic-assumption – one edge represents one kind of relation) and its weight reflects its importance. The flexibility of such approaches leads to their success in many applications, e.g., bio-informatics [18], document analysis [19, 20], image processing [21]. However, the atomic-assumption seldom holds in real-world networks. For example, in a social network, two persons always have various relationships, such as colleague relationship, friend relationship, etc. Also, two terms can occur together in different contexts with different meaning. Thus, the atomic-assumption limits the usage of these models.

In this paper, we propose a novel hierarchical Bayesian model called Particle Mixed Membership Stochastic Block Model (pMMSB) to model relational data. Our model is an extension to the notable model, Mixed Membership Stochastic Model (MMSB) [14, 17]. Unlike MMSB regarding each edge as an atom, our model considers each edge as an union of some multiple elementary relationships called Particle and its weight reflects the overall importance of these elementary relationships. In fact, the key difference between our method and previous work lies in the smallest relation unit of a network. In previous work, the smallest relation unit of a network is edge, which means a network is composed of nodes and edges. While pMMSB considers particles (elementary relationships) as the smallest relation unit, which means a network is composed of nodes and particles. The edges in pMMSB, a higher level relation unit, are treated as the composition of particles. This difference allows our method to involve more information and model the uncertainty of edges more appropriately, effectively, and robust. For efficiency, we develop a nested variational Bayesian algorithm [14, 22] for fast inference. We demonstrate the properties of our model on a synthetic network and three real datasets, gene regulations in A. Thaliana, Lesmis and EAT.

The rest of the paper is organized as follows. Section II briefly introduces some related work. In Section III, pMMSB is presented. The variational inference for pMMSB is given in Section IV. The empirical results are given in Section V. Section VI concludes the paper.

II. RELATED WORK

This section briefly describes some related work. The notion of blocks is introduced by Lorrain and White [23] to model uncertainty within networks. This notion is further developed later and a notable class is Stochastic Block Model (SBM) [12]. The basic SBM assumes that there exist underlying blocks, and assigns vertices to these latent blocks probabilistically. Since the ease of interpretation, SBM has been widely studied [7, 8, 11, 13, 24, 25]. However, the conventional SBM
suffers from a restriction that each object can only belong to one block. Therefore, a general Stochastic Block Models of mixed membership, called Mixed Membership Stochastic Block model (MMSB) is recently proposed [14, 18]. Some related work [15-17, 26] can be considered as the extension of the basic MMSB. Specifically, the Bayesian co-clustering algorithm [26] can be seen as an extension on dyadic data, and the work [16, 17, 20] is the extension on weighted graphs with the assumption that the weight of edges follows a specific distribution.

Another generalized SBM of mixed membership, SSN-LDA [27], is of a little bit difference from MMSB. Specifically, SSN-LDA assumes that each node is a bag of outgoing edges. It models each outgoing set of edges as a mixture over latent variables. Specifically, SSN-LDA does not model the interaction patterns between latent blocks.

III. PARTICLE MIXED MEMBERSHIP STOCHASTIC BLOCK MODEL

We represent a network as a weighted directed graph $G = (\mathcal{V}, \mathcal{E}, \mathcal{H})$. The element $v_i$ in the note set $\mathcal{V}$ represents the i-th object in the network. Each element $e_{ij} \in \mathcal{E}$ indicates that there exists a relation from the object $v_i$ to the object $v_j$, and the weight of $e_{ij}$ is represented by $h(e_{ij}) \in \mathcal{H}$. The size of the sets $\mathcal{V}$, $\mathcal{E}$ and $\mathcal{H}$ are $|\mathcal{V}|$, $|\mathcal{E}|$ and $|\mathcal{H}|$ respectively. As above description, the size of $\mathcal{E}$ equals to the size of $\mathcal{V}$ all the time, i.e., $|\mathcal{E}| = |\mathcal{H}|$.

Note that, an edge between the objects $v_i$ and $v_j$ in an undirected graph can be easily represented in the same form with just adding two directed edges $e_{ij}$ and $e_{ji}$ in the edge set. So, in the following, we only discuss directed graphs.

Before we formally describe our model, we will firstly give the formal definitions of Particle and Weight.

Definition 1 (Particle): Particle $P_{ij} \in \mathbb{R}^+$ is the elementary measurement from the node $v_i$ to the node $v_j$, where $v_i, v_j \in \mathcal{V}$.

Note that the definition of Particle indicates that Particle is the smallest relation unit and each edge is composed of many Particles.

Definition 2 (Weight): Weight of the edge $e_{ij}$ ($h(e_{ij})$) is the overall measurement of it, denoted as $h(e_{ij})$. The value is the statistical average of all the particles between the nodes $v_i$ and $v_j$:

$$h(e_{ij}) = C \cdot \sum_t q_{ij}^t \cdot P_{ij}^t$$

(1)

where the $C$ is a scaling factor and the coefficient $q_{ij}^t$ is the confidence of the particle $P_{ij}^t$.

The confidence is the belief or trust in those particles. In practice, it can be a prior probability of the way particles are generated, or just a real number which illustrates the state of being certain that such a particle is correct. Definition 2 illustrates: 1) each edge is union of particles, and 2) the weight of such an edge is the statistical average of these particles.

According to the maximum entropy principle, it is better to treat all the particles with the same confidence if we are without any prior knowledge about the particles. Furthermore, if we set the scaling factor $C$ to one, the weight of the edge $e_{ij}$ is calculated by just summing the values of all the $N_{ij}$ particles, i.e. $h(e_{ij}) = \sum_{t=1}^{N_{ij}} P_{ij}^t$.

For simplicity, we only consider the binary particles, i.e., $P_{ij}^t \in \{0, 1\}$, for any $v_i, v_j \in \mathcal{V}$. So, the weight is just a positive integer.

A. Model Formulation

To generate a network, firstly, each node $v_i$ instantiates its block membership $\pi_i$. Secondly, we generate $N_{ij}$ the number of particles from the node $v_i$ to the node $v_j$. Thirdly, we assign the specific blocks for $v_i$ and $v_j$ according to their mixed membership $\pi_i$ and $\pi_j$. Finally, the value of each particle is sampled according to a Bernoulli distribution with the specific block assignments fixed. And then the weight of edges is obtained by summing the values of all their related particles. Bring the idea of particles into pMMSB makes it involve more information about edges and increases its ability of uncertainty tolerance. This allows our method to model networks more appropriately and effectively. The complete generative progress is as follows:

- for $g, h \in [1, K]$ - sample $B_{gh} \sim p_{Beta}(\beta)$
- for each node $v_i \in \mathcal{V}$ - sample a $K$ dimensional mixed membership vector $\pi_i \sim p_D(\alpha)$
- for each pair of nodes $(v_i, v_j) \in \mathcal{V} \times \mathcal{V}$ - sample $N_{ij} \sim Poiss(\xi)$
- for $t \in [1, N_{ij}]$ - sample $z_{i \rightarrow j}^t \sim p_M(\pi_i)$
- sample $z_{i \rightarrow j}^{\pi_i} \sim p_M(\pi_j)$
- sample $P_{ij}^t \sim p_B(B_{z_{i \rightarrow j}^t z_{i \rightarrow j}^{\pi_i}})$
- $h(e_{ij}) = \sum_{t=1}^{N_{ij}} P_{ij}^t$

where $Poiss(\cdot)$ is a Poisson distribution, $p_{Beta}$ is a Beta, $p_D$ is a Direchlet, $p_M$ is a multinomial, $p_B$ is a Bernoulli, $K$ is the number of blocks. The indicator variable $z_{i \rightarrow j}^t$ denotes the specific block assignment of the node $v_i$ when it connects to the node $v_j$ with the t-th particle, while $z_{i \rightarrow j}^{\pi_i}$ denotes the specific block assignment of the node $v_j$ when it is connected from the node $v_i$ with the t-th particle. The two sets of latent block indicators $\{z_{i \rightarrow j}^t : v_i, v_j \in \mathcal{V}, t \in [1, N_{ij}]\}$ and $\{z_{i \rightarrow j}^{\pi_i} : v_i, v_j \in \mathcal{V}, t \in [1, N_{ij}]\}$ are denoted as $Z_{\rightarrow}$ and $Z_{\rightarrow}$ respectively. $B$ is a $K \times K$ dimensional matrix of Bernoulli rates, and each element $B_{gh}$ of $B$ represents a specific interaction pattern. This generative progress defines a joint probability distribution:

$$p(G, \pi_i | v_i Z_{\rightarrow}, Z_{\rightarrow}; B|\alpha, \beta) = \prod_{i,j=1}^{N_{ij}} p_B(P_{ij}^t | z_{i \rightarrow j}^t z_{i \rightarrow j}^{\pi_i}, B) \cdot p_M(z_{i \rightarrow j}^{\pi_i} | \pi_i) \cdot \prod_{i=1}^{N_{ij}} p_D(\pi_i | \alpha) \cdot p_{Beta}(B | \beta)$$

(2)
B. Comparison with Previous Work

In this section, we compare pMMSB with previous work and point out the key difference between.

Some other extensions of MMSB also posit there exist $K$ latent blocks and each node is associated with a mixed membership. The key difference between pMMSB and these approaches, as described in Section I, is the smallest relation unit of a network. Such difference leads to the difference generation process of edges. This is the key point how our method involves more information about edges and effectively models the uncertainty.

To better understand this difference, we compare pMMSB with M. Mardiassou and S. Robin’s method [29], a notable method in the same framework as ours. We call M. Mardiassou and S. Robin’s method weightedSBM here and after. The generation procedure of relations of both weightedSBM and pMMSB is listed in Table I.

<table>
<thead>
<tr>
<th>weightedSBM</th>
<th>pMMSB</th>
</tr>
</thead>
<tbody>
<tr>
<td>for each pair of nodes $(v_i, v_j) \in V \times V$</td>
<td>$z_i \sim PM(\pi_i)$</td>
</tr>
<tr>
<td></td>
<td>$z_j \sim PM(\pi_j)$</td>
</tr>
<tr>
<td></td>
<td>$h(e_{ij}) \sim Poiss(\lambda_{z_i, z_j})$</td>
</tr>
<tr>
<td></td>
<td>$N_{ij} \sim Poiss(\xi)$</td>
</tr>
<tr>
<td>$z_i \sim PM(\pi_i)$ for $n \in [1, N_{ij}]$</td>
<td>$z_{n_{ij}} \sim PM(\pi_i)$</td>
</tr>
<tr>
<td>$z_{n_{ij}} \sim PM(\pi_i)$</td>
<td>$z_{n_{ij}} \sim PM(\pi_i)$</td>
</tr>
<tr>
<td>$p_{n_{ij}} \sim Beta(B_{z_{n_{ij}}, z_{n_{ij}}})$</td>
<td>$h(e_{ij}) = \sum_{n_{ij}=1}^{N_{ij}} p_{n_{ij}}$</td>
</tr>
</tbody>
</table>

| TABLE I |
| EDGE GENERATION PROCEDURE. |

As shown in Table I, we can see the smallest relation unit of a network is different. In weightedSBM, edges and their corresponding weights are directly drawn from a specific Poisson distribution with the Poisson parameter $\lambda_{z_i, z_j}$. In this case, all the information about one edge is obtained in one instantiation operation. In other words, each edge is treated as a whole, and the presence and the weight information about it are obtained. In contrast, pMMSB posits that each edge is composed of many particles. Each particle is drawn from a Bernoulli distribution with the parameter $B_{z_{n_{ij}}, z_{n_{ij}}}$. Different values of $z_{n_{ij}}$ lead to different Bernoulli distributions such that the particles between two nodes can model the uncertainty more appropriately and effectively.

Our method can also be considered as an extension of conventional MMSB. Specifically, pMMSB can be reduced to the conventional MMSB when 1) limiting the number of particles between any two objects to one, and 2) the value of the particle follows Bernoulli distributions.

IV. INFERENCE AND ESTIMATION

In this section, we detail the variational expectation maximization (EM) procedure to carry out the approximate estimation and inference.

A. Variational E-Step

To maximize the likelihood for the given observations $G$, we need to integrate over all the hidden variables, i.e.,

$$p(G) = \int_{\Pi, B} \sum_{Z_+, Z_-} p(G, \Pi, Z_+, Z_-, B) d\Pi dB$$  (3)

Since no closed-form solution to above problem exists [14, 19], we have to appeal to variational methods [22] to get a lower bound of the (log) likelihood and approximate the posterior distributions.

In this paper, we prefer the mean-field variational method [22], which utilizes a fully-factorized distribution to approximate the true posterior probability of the hidden variables, $p(\Pi, Z_+, Z_-, B|G)$. Based on Jensen’s inequality, we can get the lower bound of the (log) likelihood as follows:

$$\log p(G | \alpha, \beta) \geq \mathcal{H}(q) + E_q[\log p(G, \Pi, Z_+, Z_-, B)] = \mathcal{L} (\Gamma, \Phi, \eta; \alpha, \beta)$$  (4)

where $\mathcal{H}(q)$ is the entropy of the variational distribution $q(\cdot)$ and the mean-field variational distribution $q(\cdot)$ is:

$$q(\Pi, Z, B | \Gamma, \Phi, \eta) = \prod_{i,j} q_{\alpha}(\pi_{ij}) \cdot q_{\beta} (B_{\alpha \beta}) \cdot \prod_{i,j,h} q_{\eta}(\gamma_{ijh})$$  (5)

where $\{\Gamma, \Phi, \eta\}$ are the set of free variational parameters for tightening the bound.

Tightening the lower bound $\mathcal{L} (\Gamma, \Phi, \eta; \alpha, \beta)$ leads to the following updates for free parameters $\Phi$. For each pair of nodes $v_i$ and $v_j$, for $g, h = 1, \ldots, K$, we have

$$\phi^t_{i \rightarrow j, g} \propto \exp \left\{ \psi(\gamma_{ijg}) - \psi \left( \sum_{g=1}^{K} \gamma_{ijg} \right) \right\} \times \exp \left\{ \sum_{h=1}^{K} \phi^t_{i \rightarrow j, h} \cdot E_q[f(P_{ijh}^t, B_{gh})] \right\}$$  (6)

$$\phi^t_{i \rightarrow j, h} \propto \exp \left\{ \psi(\gamma_{ijk}) - \psi \left( \sum_{h=1}^{K} \gamma_{ijk} \right) \right\} \times \exp \left\{ \sum_{g=1}^{K} \phi^t_{i \rightarrow j, g} \cdot E_q[f(P_{ijg}^t, B_{gh})] \right\}$$  (7)

where $E_q[f(P_{ijg}^t, B_{gh})] = P_{ijg}^t \cdot E_q[\log B_{gh}] + (1 - P_{ijg}^t) \cdot E_q[\log (1 - B_{gh})]$. The updates for free parameters $\Gamma$ are as follows, for each node $v_i$:

$$\gamma_{ik} = \alpha_k + \sum_{j=1}^{\left| \mathcal{V} \right|} \sum_{t=1}^{N_{ij}} \phi^t_{i \rightarrow j, k} + \sum_{j=1}^{\left| \mathcal{V} \right|} \sum_{t=1}^{N_{ij}} \phi^t_{i \rightarrow j, k}$$  (8)
Algorithm 1: Outer layer of nested variational inference

initialize $\gamma^0_k$ randomly;
initialize $\gamma^0_h = \frac{2N}{K}$ for all $i, k$;
repeat
  for $i = 1$ to $|V|$ do
    for $j = 1$ to $|V|$ do
      for $t = 1$ to $N_{ij}$ do
        $\phi_{i\to j}^{t+n+1} = \mathcal{F}(P^{t}_{ij}; \gamma^n_i, \gamma^n_j, \eta^n)$;
        update $\gamma^n_i$, $\gamma^n_j$ as Eq. (6);
        update $\eta^{n+1}$ as Eq. (7);
      end
    end
  end
until convergence;

Algorithm 2: Inner layer of nested variational inference: $\mathcal{F}()$

initialize $\phi^0_{i\to j,g} = \phi^0_{i\to j,h} = \frac{1}{N}$ for all $g, h$;
repeat
  for $g = 1$ to $K$ do
    update $\phi^{t+1}_{i\to j,g}$ as Eq. (6);
  end
  normalize $\phi^{t+1}_{i\to j,g}$ to sum to 1;
  for $h = 1$ to $K$ do
    update $\phi^{t+1}_{i\to j,h}$ as Eq. (7);
  end
  normalize $\phi^{t+1}_{i\to j}$ to sum to 1;
until convergence;

And the updates for free parameters $\eta$ are as follows, for each $g, h = 1, \ldots, K$

$$\eta^*_{gh,+} = \beta_+ + \sum_{i,j} \sum_{t=1}^{N_{ij}} P^t_{ij} \cdot \phi^t_{i\to j,g} \cdot \phi^t_{i\to j,h}$$ (9)

$$\eta^*_{gh,-} = \beta_- + \sum_{i,j} \sum_{t=1}^{N_{ij}} (1 - P^t_{ij}) \phi^t_{i\to j,g} \cdot \phi^t_{i\to j,h}$$ (10)

To speed up convergence, we utilize a nested variational reference employed by [14]. The nested variational reference scheme first optimizes $\Phi$ with all the other free parameters fixed. Then, the optimal solution $\Phi^*$ is involved in the updates of $\Gamma$ and $\eta$. This procedure is repeated until convergence. Details are displayed in Algorithm 1 and Algorithm 2.

B. M-Step

After tightening the free variational parameters, we obtain a lower bound $\mathcal{L}(\Gamma^*, \Phi^*, \eta^*, \alpha, \beta)$ of log-likelihood. In M-step, $\mathcal{L}$ is considered as a function of the hyper-parameters $\{\alpha, \beta\}$ to be maximized. A closed form solution to this optimal problem does not exist yet [28]. Therefore, we use a linear-time Newton-Raphson method, where the gradient and Hessian are:

$$\frac{\partial \mathcal{L}}{\partial \alpha_k} = N \times \left( \psi(\sum_{k=1}^{K} \alpha_k) - \psi(\alpha_k) \right)$$ (11)

$$+ \sum_{i=1}^{|V|} \left( \psi(\gamma_{ik}) - \psi(\sum_{k=1}^{K} \gamma_{ik}) \right)$$ (12)

After several iterations, we will get the approximate optimum $\alpha^*$. The algorithm of estimating hyper-parameters $\beta$ is analogous to $\alpha$.

C. Block Number Identification

A natural problem is how to determine the block number $K$. In pMMSB, we solve this problem via the Bayesian method. Given a weighted graph $G = \{V, E, H\}$, the block number $K$ is determined as $K^* = \arg\max_K p(K|G)$. Based on the Maximum-Entropy principle, we prefer a sufficient weak prior $p(K)$ if we are lack of any knowledge about $K$. This implies that maximizing $p(K|G) \propto p(G|K) \cdot p(K)$ is equivalent to maximizing $p(G|K)$.

D. Predicting Missing Edges

We can utilize pMMSB predict missing relations. When predicting missing edges, the posterior expectation of likely-existence of a particle is calculated as:

$$\tilde{\pi}^T_j \hat{\mathbf{B}} \cdot \tilde{\pi}_j$$ (13)

One relation is regarded to be missing when above value is larger than a threshold, otherwise it is considered to be inexistence.

V. EXPERIMENTS

In this section, we present the experimental results on a synthetic data set and three real datasets including a gene dataset [30], a network of co-appearances in Victor Hugo’s novel “Les Miserables” [31] and EAT. These experimental results demonstrate the benefits of pMMSB on different tasks. To illustrate the properties of pMMSB, we compare pMMSB with the state-of-the-art stochastic block model, weightedSBM [15].

The model weightedSBM is an extension of the basic MMSB on the weighted networks. It assumes that the weight of edges follows a specific distribution. In weightedSBM, a network is summarized by a $K \times K$ parameter matrix $\lambda$, where $K$ is the number of latent blocks, and the node-specific mixed memberships $\pi_i$. The complete generative process of weightedSBM defines a joint probability distribution as follows:

$$p(G, \Pi, Z_{\rightarrow}, Z_{\leftarrow}|\alpha, \lambda)$$ (14)

$$= \prod_{i=1}^{|V|} p_D(\pi_i|\alpha) \prod_{i,j} \left( f(h(e_{ij})|\lambda_{z_{i\rightarrow}, z_{i\leftarrow}}) \right)$$ (15)

$$\times p_M(z_{i\rightarrow}|\pi_i) \cdot p_M(z_{i\leftarrow}|\pi_j)$$ (16)

These provide a clear understanding of the model. The prior distribution for the hyper-parameters $\alpha$ and $\lambda$ are considered to be weak. We can follow the Maximum-Entropy principle to prefer a sufficient weak prior $p(K)$ if we are lack of any information about $K$. This implies that maximizing $p(K|G) \propto p(G|K) \cdot p(K)$ is equivalent to maximizing $p(G|K)$. A natural problem is how to determine the block number $K$. In pMMSB, we solve this problem via the Bayesian method. Given a weighted graph $G = \{V, E, H\}$, the block number $K$ is determined as $K^* = \arg\max_K p(K|G)$. Based on the Maximum-Entropy principle, we prefer a sufficient weak prior $p(K)$ if we are lack of any knowledge about $K$. This implies that maximizing $p(K|G) \propto p(G|K) \cdot p(K)$ is equivalent to maximizing $p(G|K)$.

D. Predicting Missing Edges

We can utilize pMMSB predict missing relations. When predicting missing edges, the posterior expectation of likely-existence of a particle is calculated as:

$$\tilde{\pi}^T_j \hat{\mathbf{B}} \cdot \tilde{\pi}_j$$ (13)

One relation is regarded to be missing when above value is larger than a threshold, otherwise it is considered to be inexistence.

V. EXPERIMENTS

In this section, we present the experimental results on a synthetic data set and three real datasets including a gene dataset [30], a network of co-appearances in Victor Hugo’s novel “Les Miserables” [31] and EAT. These experimental results demonstrate the benefits of pMMSB on different tasks. To illustrate the properties of pMMSB, we compare pMMSB with the state-of-the-art stochastic block model, weightedSBM [15].

The model weightedSBM is an extension of the basic MMSB on the weighted networks. It assumes that the weight of edges follows a specific distribution. In weightedSBM, a network is summarized by a $K \times K$ parameter matrix $\lambda$, where $K$ is the number of latent blocks, and the node-specific mixed memberships $\pi_i$. The complete generative process of weightedSBM defines a joint probability distribution as follows:

$$p(G, \Pi, Z_{\rightarrow}, Z_{\leftarrow}|\alpha, \lambda)$$ (14)

$$= \prod_{i=1}^{|V|} p_D(\pi_i|\alpha) \prod_{i,j} \left( f(h(e_{ij})|\lambda_{z_{i\rightarrow}, z_{i\leftarrow}}) \right)$$ (15)

$$\times p_M(z_{i\rightarrow}|\pi_i) \cdot p_M(z_{i\leftarrow}|\pi_j)$$ (16)

These provide a clear understanding of the model. The prior distribution for the hyper-parameters $\alpha$ and $\lambda$ are considered to be weak. We can follow the Maximum-Entropy principle to prefer a sufficient weak prior $p(K)$ if we are lack of any information about $K$. This implies that maximizing $p(K|G) \propto p(G|K) \cdot p(K)$ is equivalent to maximizing $p(G|K)$.
where the hidden variables \( \{ \Pi_i, Z_{-i}, \Pi_{-i} \} \) are defined the same as in pMMSB and the generation distribution \( f(\cdot) \) can be any distribution.

**A. Synthetic Data**

A 100-node undirected weighted network is generated accordingly to the generative model as described in Section III-A with predefined parameters. We specify such network by its adjacency matrix \( A \) as in Figure 2, where the entry \( A_{ij} \) is the weight of the relation between the node \( v_i \) and the node \( v_j \). Since the weight is an integer, we set the generation distribution \( f(\cdot) \) in weightedSBM to be a Poisson.

The number of communities \( K \) is evaluated by the variational Bayesian (VB) method as described in Section IV-C. We compare the estimated parameters with both the true model parameters used to generate the synthetic data and the parameters estimated by weightedSBM. The estimation results are presented in Figure 3. In Table II, we evaluate the cluster assignments in terms of cluster accuracy. Cluster accuracy is defined as:

\[
CA = \frac{1}{N} \sum_{k=1}^{K} n_k,
\]

where \( N \) is the number of nodes, \( K \) is the number of communities and \( n_k \) is, for the \( k \)-th experimental community, the largest number of nodes falling into the correct community. We assign the community for each node \( v_i \) as its result community with the highest value of the corresponding posterior expectation. The total computation time for \( K = 4 \) on a standard PC with MATLAB is about 2 hours.

The results show: 1) The variational Bayesian method can give an accurate estimate of the community number \( K \); 2) Our method can accurately and robustly recover the mixed membership of individual node, and get a high cluster accuracy.

**B. Gene Regulations in A. Thaliana**

Gene regulations in A. Thaliana is the partial correlations between the expression levels of 800 genes in various conditions [30]. We specify the partial correlations as an 800\times800 adjacency matrix as in Figure 1(a), where each entry is a partial correlation coefficient. A partial correlation coefficient quantifies the correlation between two variables (e.g. gene expression levels) when conditioning on one or several other variables. A white point in Figure 1(a) means a positive relation between two genes (a positive correlated genes), while a blue point represents a negative one. Since the weight reflects the class label or assignment, we set the generation distribution \( f(\cdot) \) in weightedSBM to be a two dimensional Multinomial distribution.

The results are presented in Figure 1(c). Note that Mariadassou M and Robin S also evaluate their weightSBM model on this dataset [32]. They improperly split the block 6 into two separate blocks (they identify seven blocks).

---

**Fig. 1.** Block detection on gene regulations in A. Thaliana. (a) An adjacency matrix represents partial correlations between the expression levels of 800 genes in various conditions. (b) The log-likelihood for each \( K \). The peak indicates the number of latent blocks, \( K^* = 6 \). (c) The blocks detected by our approach.

**Fig. 2.** Adjacency matrix \( A \) of 100-node synthetic data. The gray scale indicates the weight of the relation between corresponding two nodes. Note that there obviously exist four latent communities.

**Table II**

<table>
<thead>
<tr>
<th></th>
<th>weightedSBM</th>
<th>pMMSB</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CA</strong></td>
<td>83%</td>
<td>96%</td>
</tr>
</tbody>
</table>

**Cluster accuracy on synthetic data.**
to obtain variational parameters of pMMSB as Algorithm 1 and 2 till convergence, so as integer, we specify the generation distribution model hyper-parameters. Since the weight within Lesmis subsets are used to train pMMSB and weightedSBM to get the existing-log-likelihood. Specifically, we utilize the trained models to predict the probabilities of likely-existent edges, noted by \( p(e_{ij}^{\text{likely exists}}) \). Note that the likely-existent edges \( \{e_{ij}^{\text{likely}}\} \) do not appear in training sets. In pMMSB, this probability can be approximated as described in Section IV-D. Given the variational distribution \( q(\cdot) \) with optimized free parameters \( \{\Gamma^*, \eta^*\} \), such probabilities can be approximated as follows, for any \( i, j \in [1, |V|] \):

\[
p_{pMMSB}(e_{ij}^{\text{likely exists}}) \approx \hat{\pi}_i^\top \cdot \hat{\mathbf{B}} \cdot \hat{\pi}_j
\]

\[
\hat{\pi}_{ik} = \frac{\gamma_{ik}^*}{\sum_{k=1}^K \gamma_{ik}^*}
\]

\[
\hat{B}_{gh,+} = \frac{\eta_{gh,+}^*}{\eta_{gh,+}^* + \eta_{gh,-}^*},
\]

\[
\hat{B}_{gh,-} = \frac{\eta_{gh,-}^*}{\eta_{gh,+}^* + \eta_{gh,-}^*}
\]

Similarly, poissSBM can also approximate the probability by the corresponding variational distribution as following.

\[
p_{poissSBM} \approx \hat{\pi}_i^\top \cdot (1 - P) \cdot \hat{\pi}_j
\]

where \( I \) is a \( K \times K \) matrix with all the entries being 1, and \( P \) is a \( K \times K \) matrix of which the elements \( P_{gh} \) equals to the probability \( Poiss(0|\lambda_{gh}) \).

Figure 4 presents the holdout log-likelihood for each model on Lesmis for different values of \( K \). We can see that pMMSB performs better than poissSBM with different block number \( K \), especially when the number of blocks is large (e.g. \( K \geq 8 \) here). This illustrates pMMSB has stronger prediction ability than poissSBM, which benefits from the process that pMMSB considers the weight as the composition of particles. In fact, the uncertainty within the network increases with the number of blocks to some extent. Thus the simple assumption made in poissSBM that the weight of edges follows a specific distribution is inappropriate.
The probable number of blocks is 100, i.e., described in Section IV-C. The result of block number identification is presented in Figure 5(a), which indicates the number of latent blocks ($K^* = 100$). (b) The structure discovered by pMMSB. The X axis indicates the obtained blocks and the Y axis indicates words. The gray scale implies the membership scores for one word. Table 1 illustrates examples of words in this set are collected, then these responses are used to obtain further responses. 

$EAT$ (Edinburgh Associative Thesaurus) is a set of word association norms, which are sparse and full of noise and uncertainty. The data are collected by the University of Edinburgh (Kiss et al., 1973), and they are available online in the form of the Edinburgh Word Association Thesaurus (EAT) and form part of the MRC Psycholinguistic Database (Wilson, 1988).

The Edinburgh association norms are collected by growing the network from a nucleus set of words. Responses to the words in this set are collected, then these responses are used to obtain further responses. $EAT$ consists of two parts, a response-stimulus network ($eatRS$) and a stimulus-response network ($eatSR$). In this paper, we only use $eatRS$, which is a directed weighted network with 23,219 vertices and 325,624 arcs (564 loops) and the values on arcs from response words to stimulus words are the times that the stimulus words are associated with the response words.

We detect the most probable number of blocks $K$ as described in Section IV-C. The result of block number identification is presented in Figure 5(a), which indicates the most probable number of blocks is 100, i.e., $K^* = 100$.

Figure 5(b) presents the structure discovered by pMMSB, where each row represents the expected posterior mixed membership scores for one word. Table 1 illustrates examples of 6 blocks (out of 100) inferred by pMMSB. Each block is presented with the top 8 words most likely to be generated conditioned on the block in lexicographic order. Figure 6 is a graphical representation of 3 obtained blocks designated.

The results show that pMMSB can effectively capture the underlying structure within a network. The detected blocks have a common property that words have more response-stimulus relations (semantic relativity) within their blocks than the rest of the network.

VI. CONCLUSION AND FUTURE WORK

We have proposed a novel Stochastic Block Model, called Particle Mixed Membership Stochastic Block Model (pMMSB), for relational data. Our model provides an effective tool for underling block detection, missing-edge prediction, etc. Unlike previous stochastic block models assume edges to be of atomic, pMMSB regards each edge as a union of some elementary relationships called Particle. As a consequence, pMMSB can incorporate more information about edges and model the uncertainty within networks better. A fast inference based on nested variational inference has also been offered.

We demonstrate the benefits of pMMSB on a synthetic data set and three real-world networks. The empirical results show that our method can correctly find the underlying blocks and predict missing edges more effectively than the state-of-the-art stochastic block model.

VII. ACKNOWLEDGEMENTS

We would like to thank the anonymous reviewers for their insightful comments. This work was supported by the National Natural Science Foundation of China (Grant No. 60873174).

REFERENCES

TABLE III
AN ILLUSTRATION OF 6 BLOCKS FROM A 100-BLOCK SOLUTION FOR THE eInsRS DATASET. EACH BLOCK IS ILLUSTRATED WITH THE TOP 8 WORDS (IN LEXICOGRAPHIC ORDER) THAT HAVE THE HIGHEST PROBABILITY CONDITIONED ON THAT BLOCK.

<table>
<thead>
<tr>
<th>ANIMAL</th>
<th>KIN</th>
<th>FOOT</th>
<th>MEMORY</th>
<th>GEOMETRY</th>
<th>DOCTOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>alligator</td>
<td>aunt</td>
<td>boots</td>
<td>amnesia</td>
<td>angle</td>
<td>chemist</td>
</tr>
<tr>
<td>antelope</td>
<td>cousin</td>
<td>foot</td>
<td>forget</td>
<td>arc</td>
<td>doctor</td>
</tr>
<tr>
<td>beetle</td>
<td>daughter</td>
<td>heel</td>
<td>forgetfulness</td>
<td>circle</td>
<td>medic</td>
</tr>
<tr>
<td>bug</td>
<td>grandma</td>
<td>sandal</td>
<td>homesickness</td>
<td>cube</td>
<td>physician</td>
</tr>
<tr>
<td>bull</td>
<td>grandma</td>
<td>shoe</td>
<td>memento</td>
<td>curve</td>
<td>prescription</td>
</tr>
<tr>
<td>butterfly</td>
<td>mother</td>
<td>sock</td>
<td>memories</td>
<td>cylindrical</td>
<td>quack</td>
</tr>
<tr>
<td>camel</td>
<td>mum</td>
<td>slipper</td>
<td>nostalgia</td>
<td>ellipse</td>
<td>scalpel</td>
</tr>
</tbody>
</table>

Fig. 6. A graphical representation of examples of three blocks. Each node represents a word and each arrow corresponds to a response-stimulus relation from a response word to a stimulus word. And the width of arrows indicates the weight of relations. The inferred block assignments on three example blocks are designated by color and shape.