Weighted Infinite Relational Model for Network Data

Xiaojuan Jiang and Wensheng Zhang
Institute of Automation, University of Chinese Academy of Sciences, Beijing 100190, PR China
Email: {xiaojuan.jiang, wensheng.zhang}@ia.ac.cn

Abstract —As the availability and scope of social networks and relational datasets increase, learning latent structure in complex networks has become an important problem for pattern recognition. To contract compact and flexible representations for weighted networks, a Weighted Infinite Relational Model (WIRM) is proposed to learn from both the presence and weight of links in networks. As a Bayesian nonparametric model based on the Dirichlet process prior, a distinctive feature of WIRM is its ability to learn the latent structure underlying weighted networks without specifying the number of clusters. This is particularly important for structure discovery in complex networks, especially for novel domains where we may have little prior knowledge. We develop a mean-field variational algorithm to efficiently approximate the model's posterior distribution over the infinite latent clusters. Experiments on synthetic data set and real-world data sets demonstrate that WIRM can effectively capture the latent structure underlying the complex weighted networks.

Index Terms—Pattern recognition, network modeling; bayesian nonparametric model, dirichlet process, Chinese restaurant process, exponential family, variational inference

I. Introduction

Statistical analysis of complex networks has been an active area of research for decades, and is becoming an increasingly important challenge in pattern recognition as the scope and availability of network datasets increase in various scientific fields [1]. Unlike traditional data collected from individual objects, the observations in network data are no longer independent or exchangeable because vertices are pairwise related. Independence or exchangeability is a key assumption made in machine learning and statistics for traditional attribute data [2]. This intrinsic difference in structure requires special treatments for network data.

Uncovering the latent structure based on the observed pairwise interactions between vertices [1], [3] has been a focus of attention in the network literature. Among all the statistical models proposed for this end, the Stochastic Block Model (SBM) [4] is an elegant generative model of group structure in unweighted networks. The Stochastic Block Model has been successfully used for modeling assortative network structure [5], disassortative structure [1] and bipartite structure [6]. And the SBM also has been generalized for count-valued data, degree correction [7] and categorical values [8]. The Mixed Membership Stochastic Block Model [9] increases the expressiveness of the latent class models by allowing mixed membership, associating each object with a distribution vector over clusters.

A basic assumption in most of these models is that the networks are unweighted, where the interaction presence or absence is represented as a binary variable. However, most real-world networks contain information about link weights. For instance, in social networks the weights represent the strengths of social ties between people [1]. A widely-used technique to analyze on weighted network is to transform the data into the binary framework via thresholding. As a result, the potential loss of valuable information from thresholding may lead to obscuration or distortion in recovering underlying structure [10], [11].

To directly learn the latent structure of weighted networks, an extension of the stochastic block model with Poisson likelihood [7], [12] was considered for count-valued pairwise interactions. Moreover, a generalization of the SBM, called Weighted Stochastic Block Model (WSBM) [10], was introduced to learn from both the link-existence and -weight information.

However, the number of latent clusters (or blocks) in all these models is required to be specified. This may be very difficult to access for real-world networks. Usually, this parameter is set a priori or fixed via a computational expensive model selection procedure [3], [10], [13]. To overcome this problem, the Infinite Relational Model (IRM) [14] and the Infinite Hidden Relational Model [15] use the Dirichlet process prior to define a nonparametric relational model for unweighted networks.

In this paper, we introduce the Weighted Infinite Relational Model (WIRM), a Bayesian nonparametric model that can learn a potentially infinite number of clusters from both the existence and weight of links. We take each weighted link as a draw from a parametric exponential family distribution, which includes as special cases most standard distributions, e.g. the Bernoulli, the Gaussian and their generalizations. With this general distributional form, we can directly use the weight information in recovering latent cluster or block structure. Moreover, the WIRM uses a nonparametric Bayesian approach to simultaneously infer the number of latent clusters and the cluster membership of each object, while at the same time inferring how cluster membership influences the observed weighted link interactions.

There are two models that are closely related to WIRM. The infinite relational model (IRM) [14] previously

Manuscript received February 11, 2015; revised June 24, 2015.
This work was supported by the National Natural Science Foundation of China under Grant No.U1135005.
Corresponding author email: xiaojuan.jiang@ia.ac.cn.
doi:10.12720/jcm.10.6.442-449

adapted the Dirichlet process to define a nonparametric relational model for network modeling. IRM fits only to the link-existence information but ignores link-weight information; however, WIRM can learn from both types of information using a general form of exponential family distribution. On the other hand, WIRM can also be seen as a nonparametric extension of the WSBM proposed in [10], where the number of clusters is chosen before the model can be applied to data. Compared to WSBM, a distinctive feature of WIRM is its ability to infer from the observed data that how many latent clusters there are. This is particularly important when we may have little prior knowledge about the number of clusters, especially for structure discovery in novel domains.

The paper is arranged as follows. We first describe the generative process of our model in Section II. Then a variational inference algorithm is derived for performing approximate posterior inference and parameter estimation in Section III. Section IV compares the performance of the WIRM to alternative methods for two link prediction tasks, and analyzes the results. Section V concludes the paper.

II. Weighted Infinite Relational Model

Let A be a $N \times N$ matrix that contains links information among objects in a directed relational network. Here, we consider the following two types of information in the link observations: information about link-existence (presence or absence of links) and information about link-weights (the weighted values). To specify these two types of information in the networks, we can take the adjacency matrix *A* as a binary-valued matrix or a real-valued (or count-valued) matrix. We want to partition the set of objects into clusters, so that the relationships between objects can be predicted by their cluster assignments. The number of latent clusters present in the network, which is not known a prior, is denoted as *K*, so that the cluster assignment variable of each object $z_i \in \{1, 2, \cdots, \mathrm{K}\}$.

A. Modeling Observed Link Information

Suppose we are given the cluster assignment vector $Z = \{z_1, z_2 \cdots, z_N\}$. For each pair of clusters (*kk*'), we can model the 'bundle' of links from objects in cluster *k* to those in cluster *k*', using an exponential distribution family parameterized by $\theta_{kk'}$. That is, for object *i* with cluster assignment $z_i = k$ and object *j* with $z_j = k'$, the likelihood of observing a link A_{ij} is

$$P(\mathrm{A}_{ij} \mid \mathrm{Z}, \theta) \propto \exp\left[T(\mathrm{A}_{ij}) \cdot \eta(\theta_{z_i z_j})\right] \tag{1}$$

where T is the vector valued function of sufficient statistics, and η is the vector valued function of natural parameter.

Exponential family [16] comprises a set of flexible distributions ranging both continuous and discrete random variables, including the Gaussian, the Bernoulli, the Poisson, the Gamma, the Geometric, the NegBinomial, etc.

Specifically, we can choose the Bernoulli distribution to model binary existence information of the links, setting $T = (\mathrm{x}, 1)$ and $\eta = \left(\log\left[p/(1-\mathrm{p})\right], \log\left[1-\mathrm{p}\right]\right)$. For count-valued existence information of links, we may choose the Poisson distributions with $T = (\mathrm{x}, 1)$ and $\eta = (\log\lambda, -\lambda)$. To model real-valued weight information of links, the normal distribution may be used by setting $T = (\mathrm{x}, \mathrm{x}^2, 1)$ and $\eta = \left(\mu/\sigma^2, -1/(2\sigma^2), -\mu^2/(2\sigma^2)\right)$.

We may also incorporate two types of information into the likelihood function via a simple relative importance parameter $c \in [0,1]$:

$$\begin{aligned} &\log P(\mathrm{A}_{ij} \mid \mathrm{Z}, \theta) \\ &\propto cT_e(\mathrm{A}_{ij}) \cdot \eta_e(\theta^{(\mathrm{e})}_{z_i z_j}) + (1-c) T_w(\mathrm{A}_{ij}) \cdot \eta_w(\theta^{(\mathrm{w})}_{z_i z_j}). \end{aligned} \tag{2}$$

where the pair (T_e, η_e) denotes the family of link-existence distributions and (T_w, η_w) denotes the family of link-weight distributions.

B. Nonparametric Prior on Cluster Assignment

In order to allow flexible representation of the latent structure of data, we use the Dirichlet process prior cluster assignments. The Dirichlet Process, introduced in [17], is the underlying random measure of the Chinese restaurant process (CRP) [2], [18], which is widely used as a nonparametric prior for latent class models [19]. A important characteristic of the prior is that conditioned on data, we can examine the posterior distribution of Z to get a data-dependent distribution of clusters number.

The CRP metaphor gives the intuition. Imagine a restaurant with an infinite number of tables, each with an infinite number of seats. The customers enter the restaurant one after another, and each chooses a table at random. In the CRP with parameter α, each customer chooses an occupied table with probability proportional to the number of occupants, and chooses the next vacant table with probability proportional to α. This process continues until all customers have seats, defining a distribution over allocations of people to tables, and, more generally, objects to classes. One important (and surprising) property of this process is that the joint probability of final assignment is not affected by the order of customers getting into the restaurant, which is called exchangeability [18].

The Chinese restaurant construction of Dirichlet Process directly leads itself to a Gibbs sampler; whereas for the variational inference of Bayesian nonparametric models, we turn to the stick-breaking construction of [20], which provides a concrete set of hidden variables on which to place an approximate posterior [21-[23]. The stick-breaking representation of the cluster assignment $z_i \in \{1, 2, \cdots, \mathrm{K}\}$ is as follows:

$$\begin{aligned} & v_k \sim Beta(1,\alpha), k = 1,2,\ldots \\ & \pi_k(v) = v_k \prod\nolimits_{l=1}^{k-1}(1-v_k), k = 1,2,\cdots \\ & z_i \sim Mult(\pi_k(v)), \mathrm{i} = 1,2,\cdots,\mathrm{N} \end{aligned} \tag{3}$$

C. The Full Bayesian Model

To perform fully-Bayesian inference, we need to specify the prior for link bundle parameters θ. For Bayesian models, the inference analysis with conjugate priors would be considerably simplified as the posterior distributions have the same functional form as the priors. Here, for the given likelihood in exponential family form, the standard conjugate prior on θ [16] is

$$p(\theta) = \frac{1}{Z(\tau)} \exp\left[\tau \cdot \eta(\theta)\right] \tag{4}$$

where τ parameterizes the prior and $Z(\tau)$ is a normalizing factor.

For notational convenience, we let r index the $K \times K$ link-bundles between clusters; hence $\theta = (\theta_1, \cdots, \theta_r)$.

Now, we can summarize the whole generative process of the Weighted Infinite Relational Model as:

- For each object i, assign a cluster membership z_i as in (3).
- For each pair of clusters (kk'), draw a link bundle parameter $\theta_{kk'}$ according to (4).
- For each pair of objects with index i and j, draw the link observation A_{ij} from the exponential family in (1).

To specify the link-existence observation and -weight observation simultaneously, we can take (2) instead of (1) as the likelihood function in the generative process of WIRM.

III. Inference

The WIRM defines a generative probabilistic process of network data with hidden structure. Given network link observations A, we need to recover the underlying structure of the network by inferring the posterior distribution of the latent variables. However, the posterior distribution of the latent variables under a Dirichlet Process Prior is not available in closed form [19], [21], [24].

In this section, we represent a variational algorithm [25] for WIRM with likelihood function defined as in (1). For general case with likelihood as in (2), the inference algorithm follows with minor modifications, and here we omit the redundant details.

A. Truncated Variational Distributions

The hidden variables that we are interested in are the auxiliary stick-breaking variables $V = \{v_1, \cdots, v_K\}$, the cluster assignment $Z = \{z_1, \cdots, z_N\}$, and the link parameters $\theta = \{\theta_1, \cdots, \theta_r\}$. To apply the variational approach, we take the truncated stick-breaking representation for the variational distributions. By setting $q(v_K = 1) = 1$ for a fixed K, we enforce the proportions $\pi_k(V)$ to be zero for $k > K$. It is pointed out in [21], that the model follows a full Dirichlet process prior which is not truncated; only the variational posterior distribution is truncated. The truncation level K is a variational parameter which can be freely set; it is not a part of the prior model specification. And if K is large enough, the fitted approximate posterior will exhibit fewer than K clusters.

We use the following fully factorized variational distribution for mean-field variational inference:

$$q(V, Z, \theta) = \prod_{k=1}^{K-1} q(v_k; \gamma_k) \prod_{i=1}^{N} q(z_i; \phi_i) \prod_r q(\theta_r; \tau_r)$$

where $q(v_k;\gamma_k)$ are beta distributions, $q(z_i;\phi_i)$ are multinomial distributions, and $q(\theta_r;\tau_r)$ are exponential family distributions with natural parameters τ_r and sufficient statistics $\eta(\theta_r)$.

B. Lower Bound on the Marginal Likelihood

Using the standard variational theory, we lower bound the marginal log likelihood of the observed data A using Jensen's inequality:

$$\begin{aligned} \log p(A) &\geq E_q[\log p(A, V, Z, \theta))] - E_q[\log q(V, Z, \theta)] \\ &\triangleq \mathcal{L}(q) \end{aligned} \tag{5}$$

here and elsewhere in the paper we omit the variational parameters when using q as a subscript of an expectation.

Now we expand the lower bound $\mathcal{L}(q)$ in (5) with the approximate posterior q. To simplify notation, let $\langle T \rangle_r$ and $\langle \eta \rangle_r$ be the expected values of the sufficient statistics T_r and natural parameters η_r under the approximation distribution q, that is,

$$\langle T \rangle_r = \sum_{i,j} \sum_{(z_i, z_j) = r} \phi_{i,z_i} \phi_{j,z_j} T(A_{i,j}),$$

$$\langle \eta \rangle_r = \frac{\partial}{\partial \tau} \log Z(\tau) \Big|_{\tau = \tau_r}.$$

By substituting q and the conjugate prior *p* in (5), and evaluating all the expectations, we have:

$$\begin{aligned} \mathcal{L}(q) = & \sum_r \left(\langle T \rangle_r + \tau_0 - \tau_r \right) \cdot \langle \eta \rangle_r + \sum_r \log \frac{Z(\tau_r)}{Z(\tau_0)} \\ & + (K-1)\log\alpha - \sum_{k=1}^{K-1} \log \frac{\Gamma(\gamma_{k,1} + \gamma_{k,2})}{\Gamma(\gamma_{k,1})\Gamma(\gamma_{k,2})} \\ & - \sum_{k=1}^{K-1} \left\{ (\gamma_{k,1} - 1) E_q[\log v_k] + (\gamma_{k,2} - \alpha) E_q[\log(1 - v_k)] \right\} \\ & + \sum_{i=1}^{N} \sum_{k=1}^{K-1} \left\{ \sum_{l=k+1}^{K} \phi_{i,l} E_q[\log(1 - v_k)] + \phi_{i,k} E_q[\log v_k] \right\} \\ & - \sum_{i=1}^{N} \sum_{k=1}^{K} \phi_{i,k} \log \phi_{i,k}, \end{aligned} \tag{6}$$

where

$$E_q[\log v_k] = \Psi(\gamma_{k,1}) - \Psi(\gamma_{k,1} + \gamma_{k,2}),$$
$$E_q[\log(1 - v_k)] = \Psi(\gamma_{k,2}) - \Psi(\gamma_{k,1} + \gamma_{k,2}).$$

The digamma function, denoted by Ψ, arises from the derivative of the log normalization factor in the Beta distribution.

C. Coordinate Ascent Algorithm

Now, we present an explicit coordinate ascent algorithm for optimizing the bound (6). We iteratively optimize the variational lower bound with respect to each factor in turn. Convergence is guaranteed [16] because the bound $\mathcal{L}(q)$ is convex with respect to each of the factors in the variational distribution q.

The details of the iteration are as follows:

Update for the link bundle parameters θ_r **:** The variational distribution for the link bundle parameter θ_r is exponential family with sufficient statistics $\eta(\theta_r)$ and natural parameter τ_r. Coordinate ascent update equation for the variational parameter τ_r is

$$\tau_r = \tau_0 + \langle T \rangle_r.$$

Update for the cluster assignment z_i: The variational parameters for the cluster assignment z_i are $\{\phi_{i,k}\}_k$, and the update equation for $\{\phi_{i,k}\}_k$ is

$$\phi_{i,k} \propto \exp\left\{ E_q[\log v_k] + \sum_{l=1}^{k-1} E_q[\log(1 - v_l)] + \sum_r \frac{\partial \langle T \rangle_r}{\partial \phi_{i,k}} \cdot \langle \eta \rangle_r \right\}.$$

Update for the auxiliary stick-breaking variable v_k **:** The variational distribution for the auxiliary stick-breaking variable v_k is a beta distribution parameterized with the shape parameters $(\gamma_{k,1}, \gamma_{k,2})$. Coordinate ascent update equation for these free variational parameters is

$$\gamma_{k,1} = 1 + \sum_i \phi_{i,k}, \gamma_{k,2} = \alpha + \sum_i \sum_{l=k+1}^{K} \phi_{i,l} \cdot$$

Although the variational inference algorithm yields a bound for any starting values of the variational parameters, poor initialization can lead to local maxima that yield poor bounds [16]. In practice, we run the algorithm multiple times with random initializations and choose the final parameter settings that give the best bound on the marginal likelihood (6). To further improve the performance, we can follow a sequential initialization scheme [16] to initialize the variational distribution by incrementally updating the parameters according to a random permutation of the nodes in the network.

IV. Experiments

In this section, we evaluate the performance of WIRM on a synthetic data and several real-world networks. Experiments were conducted for two purposes. First, we generate synthetic data to explore the ability of our model to infer the number of latent clusters using both the link-existence and -weight information. Second, we compare the performance of our model with a number of state-of-the-art network models on two prediction tasks.

A. Synthetic Data

We consider a simple N=100 synthetic dataset generated with 4 known equal-size clusters, which has been used in [10], see Fig. 1. The weights and existences of each link bundle are normally distributed and binary distributed, respectively, with different bundle-specific parameters. This dataset is interesting, as the bundle-specific parameters are shared in a subtle manner. Specifically, if we only consider the weight information of the network, the nodes can be naturally separated into two equal-size sup-clusters: one is the cluster comprised of nodes indexed by {1,…,50}, the other is comprised of nodes indexed by {51,…,100}; whereas considering the existence information (ignoring the weights) leads to a different cluster assignment: one cluster is comprised of nodes indexed by {1,…,25,51,…,75}, the other one is comprised of nodes indexed by {26,…,50,76,…,100}.

To analyze this network, we set the truncation level to be 20 and fit our model with pure weight information by setting c=0, pure existence information by setting c=1, and mixed information by setting c=0.5, respectively. The posterior cluster assignments over 20 possible clusters learned by WIRM are shown in Fig. 2. Examining the results, we can see that the latent structure learned by WIRM with c=0 exactly recovers the two sup-clusters underlying the link-weight information (Fig. 2(a)), and with c=1 recovers the other two different sup-clusters for the existence information of interactions (Fig. 2(b)). Moreover, Fig. 2(c) demonstrates the ability of fitted model with c=0.5 to capture the four ground-truth clusters, regarding the combination of both types of information.

B. Real-World Networks

We now compare our model to several other network models for predicting the existence or the weight of some unobserved interactions on three real-world networks. The weighted networks used for the comparison are given as follows:

Collaboration [26]. Vertices represent 226 nations on Earth, and each of the 20616 edges is weighted by a normalized count of academic papers whose author lists include that pair of nations.

Airport [27]. This is a network of the 500 busiest commercial airports in the United States, and each of the 5960 directed edges is weighted by the number of passengers traveling from one airport to another.

Forum [28]. The student social network at UC Irvine includes 1899 users that sent or received at least one message, and each of the 20291 directed edges is weighted by the number of messages sent between users.

For each of the two prediction tasks and for each dataset, we evaluate the following variants of our model: the `pure` WIRM (pWIRM), using only weight information (c=0), the `mixed` WSBM (mWIRM), using both edge and weight information (c=0.5), and the `non-`WIRM (nWIRM), using only edge information (c=1). We use normal distribution to model the weight of link interactions, and Bernoulli distribution to model the existence of links. A comparative study with the other typical models, (namely, WSBM [10] and IRM [14]), is also performed.

In both prediction tasks, we treat all networks as directed, and fit each model on 80% of N^2 interactions, and use the remaining 20% as a test set. The truncation level for our model is fixed at 50 for each model and each dataset. For those models that were initially established for unweighted networks (nWIRM and IRM), we take their partitions and compute the sample mean weight for each of the induced link bundles in the weighted network and take this value as predictor for the weight of any missing link in that bundle.

For each model and each dataset, we run 5 repeats, each time with a different 80/20 cross-validation split and using a different random initialization, and then compute the average mean-squared error (MSE) on the particular prediction task. To compare the results across different datasets, we normalized link-weights to the interval [-1, 1] after applying a logarithmic transform.

To demonstrate the efficiency and stability of our approximate inference algorithm, we examine the change of the log marginal probability bound during the iterations for variants of our model. The results on three datasets are shown in Fig. 3-5. We can see that within several iterations, the log marginal probability bound converge to a particular region, and then keep stable during the following iterations.

We now report the prediction results. For both prediction tasks, we use the sequential initialization scheme to further improve the performance. Table I represents the results for predicting link-existences and Table II represents the results for predicting link-weights. It is easily seen that, for the link-existence prediction, nWIRM(c=1) and IRM significantly outperform WSBM by using the Dirichlet process prior. For the for the link-weight prediction, as a model designed to learn only from link weight information, pWIRM(c=0) is the most accurate model for these three datasets. We also notice, mWIRM(c=0.5) produces very competitive results on both task by learning from both existence and weight information. This implies that we can learn both types information without confusing each other.

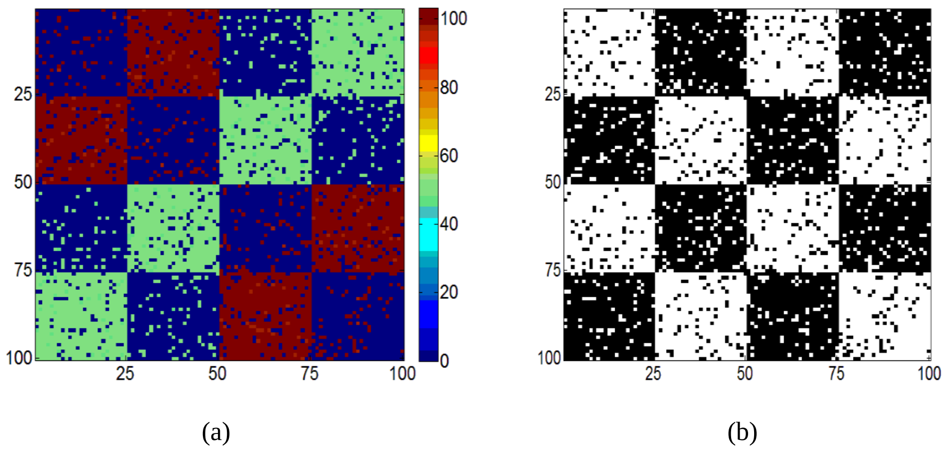


Fig. 1. Observed synthetic data example. (a) Observed synthetic 100×100 link-weight matrix. (b) Observed synthetic 100×100 link-existence matrix. White corresponds to zero, black to one.

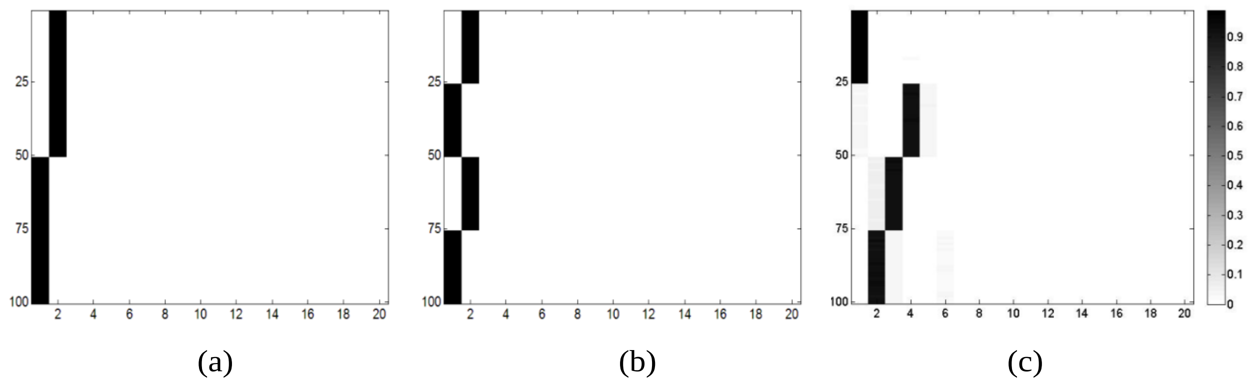


Fig. 2. Results for synthetic data. (a) Posterior cluster assignments learned from link-weight information. (b) Posterior cluster assignments learned from link-existence information. (c) Posterior cluster assignments learned from both type of information

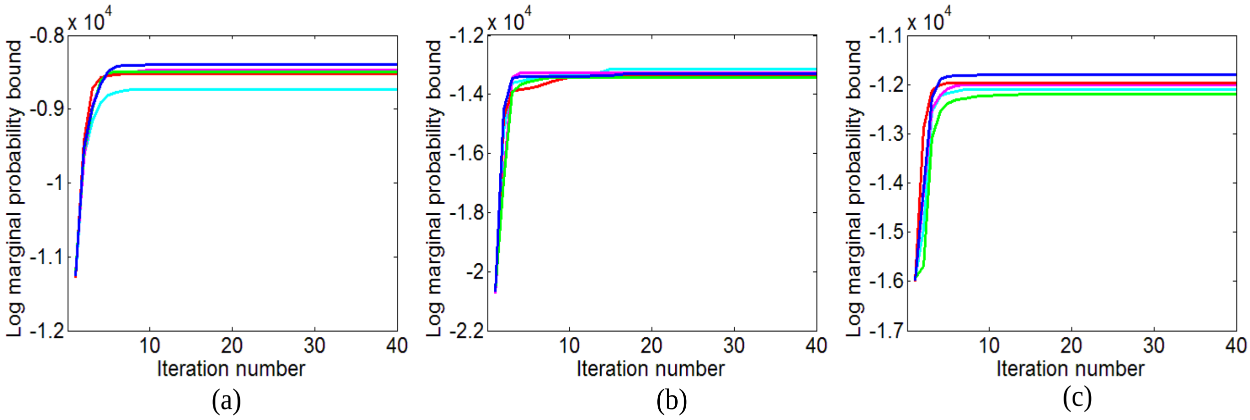


Fig. 3. Log marginal probability bound during iterations for pWIRM in (a), nWIRM in (b) and mWIRM in (c) on Collaboration dataset with 5 randomly initialized runs

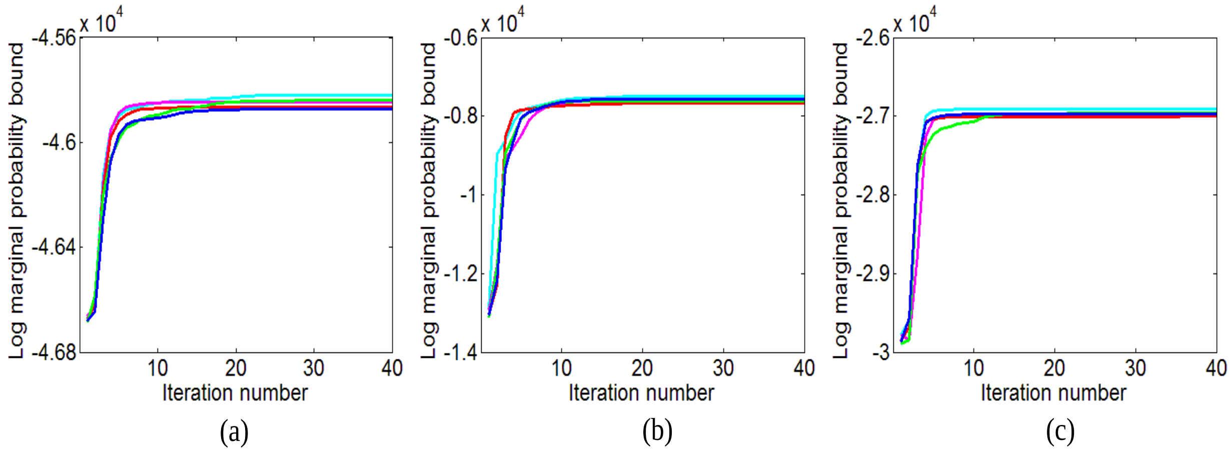


Fig. 4. Log marginal probability bound during iterations for pWIRM in (a), nWIRM in (b) and mWIRM in (c) on Airport dataset with 5 randomly initialized runs

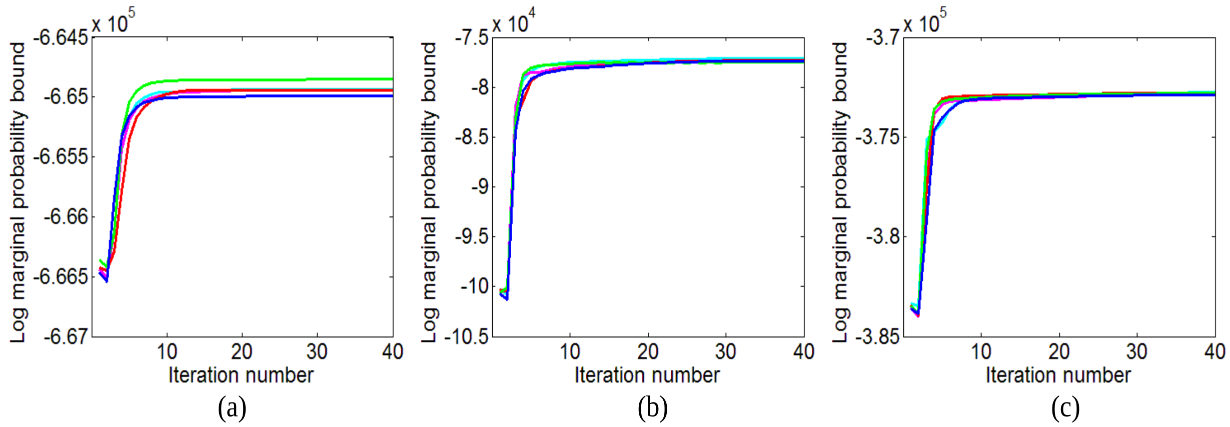


Fig. 5. Log marginal probability bound during iterations for pWIRM in (a), nWIRM in (b) and mWIRM in (c) on Forum dataset with 5 randomly initialized runs

TABLE I: AVERAGE MSE ON LINK EXISTENCE PREDICTION

	pWIRM	mWIRM	nWIRM	WSBM	IRM
Collaboration	0.0738	0.0696	**0.0682**	0.1167	**0.0687**
Airport	0.0103	0.0089	**0.0070**	0.0156	**0.0068**
Forum	0.00548	0.00517	**0.00509**	0.00535	0.00516

TABLE II: AVERAGE MSE ON LINK WEIGHT PREDICTION

	pWIRM	mWIRM	nWIRM	WSBM	IRM
Collaboration	0.0443	0.0465	0.0549	**0.0407**	0.0798
Airport	**0.0158**	0.0180	0.0224	0.0486	0.0227
Forum	**0.0491**	0.0505	0.0519	0.0726	0.0549

V. CONCLUSIONS

In this paper, we propose a novel Bayesian nonparametric model to generalize the classic infinite relation model to the important case of weighted networks. This model follows a Dirichlet process prior, in order to infer the number of latent clusters during the inference procedure. We develop an efficient coordinate ascent algorithm to preform variational inference for our model. The empirical results show that our model can efficiently capture the complex latent structure of weighted networks, and accurately predict the missing interactions and their weights.

REFERENCES

[1] M. E. Newman, *Networks: An Introduction*, Oxford University Press, 2010.

[2] D. J. Aldous, *Exchangeability and Related Topics*, Berlin: Springer, 1985.

[3] J. M. Hofman and C. H. Wiggins, "Bayesian approach to network modularity," *Physical review Letters*, vol. 100, no. 25, 2008

[4] K. Nowicki and T. A. B. Snijders, "Estimation and prediction for stochastic block structures," *Journal of the American Statistical Association*, vol. 96, no. 455, pp. 1077-1087, 2001.

[5] M. E. Newman, "Mixing patterns in networks," *Physical Review E*, vol. 67, no. 2, 2003.

[6] D. B. Larremore, A. Clauset, and A. Z. Jacobs, "Efficiently inferring community structure in bipartite networks," *Physical Review E*, vol. 90, no. 1, 2014.

[7] B. Karrer and M. E. Newman, "Stochastic block models and community structure in networks," *Physical Review E*, vol. 83, no. 1, 2011.

[8] R. Guimerà and M. Sales-Pardo, "A network inference method for large-scale unsupervised identification of novel drug-drug interactions," *PLoS Computational Biology*, vol. 9, no. 12, 2013.

[9] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing, "Mixed membership stochastic block models," in *Advances in Neural Information Processing Systems*, 2009, pp. 33-40.

[10] C. Aicher, A. Z. Jacobs, and A. Clauset, "Learning latent block structure in weighted networks," *Journal of Complex Networks*, 2014.

[11] A. C. Thomas and J. K. Blitzstein, "Valued ties tell fewer lies: Why not to dichotomize network edges with thresholds," arXiv: 1101.0788, 2011.

[12] M. Mariadassou, S. Robin, and C. Vacher, "Uncovering latent structure in valued graphs: a variational approach," *The Annals of Applied Statistics*, vol. 4, no. 2, pp. 715-742, 2010.

[13] T. P. Peixoto, "Parsimonious module inference in large networks," *Physical Review Letters*, vol. 110, no. 14, 2013.

[14] C. Kemp, J. B. Tenenbaum, T. L. Griffiths, T. Yamada, and N. Ueda, "Learning systems of concepts with an infinite relational model," in *Proc. AAAI*, 2006.

[15] Z. Xu, V. Tresp, K. Yu, and H. P. Kriegel, "Infinite hidden relational models," in *Proc. Twenty-Second Conference on Uncertainty in Artificial Intelligence*, 2006.

[16] C. M. Bishop, *Pattern Recognition and Machine Learning*, New York: Springer, 2006.

[17] T. S. Ferguson, "A Bayesian analysis of some nonparametric problems," *The Annals of Statistics*, vol. 1, no. 2, pp. 209-230, 1973.

[18] J. Pitman, "Combinatorial stochastic processes," Technical Report 621, Dept. Statistics, UC Berkeley, 2002.

[19] C. Antoniak, "Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems," *The Annals of Statistics*, vol. 2, no. 6, pp. 1152-1174, 1974.

[20] J. Sethuraman, "A constructive definition of dirichlet priors," *Statistica Sinica*, vol. 4, pp. 639-650, 1994.

[21] D. M. Blei and M. I. Jordan, "Variational inference for Dirichlet process mixture," *Bayesian Analysis*, vol. 1, no. 1, pp. 121-143, 2006.

[22] K. Kurihara, M. Welling, and Y. W. Teh, "Collapsed variational dirichlet process mixture models," in *IJCAI*, 2007, pp. 2796-2801.

[23] K. Kurihara, M. Welling, and N. A. Vlassis, "Accelerated variational dirichlet process mixtures," in *Advances in Neural Information Processing Systems*, 2006, pp. 761-768.

[24] S. Jain and R. M. Neal, "A split-merge markov chain monte carlo procedure for the dirichlet process mixture model," *Journal of Computational and Graphical Statistics*, vol. 13, no. 1, pp. 158-182, 2004.

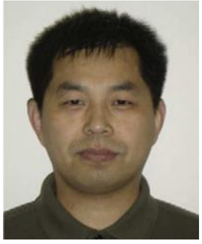
[25] H. Attias, "A variational bayesian framework for graphical models," in *Advances in Neural Information Processing Systems*, 2000, pp. 209-215.

[26] R. K. Pan, K. Kaski, and S. Fortunato, "World citation and collaboration networks: uncovering the role of geography in science," *Scientific Reports*, vol. 2, 2012.

[27] V. Colizza, R. Pastor-Satorras, and A. Vespignani, "Reaction-diffusion processes and metapopulation models in heterogeneous networks," *Nature Physics*, vol. 3, no. 4, pp. 276-282, 2007.

[28] T. Opsahl and P. Panzarasa, "Clustering in weighted networks," *Social Networks*, vol. 31, no. 2, pp. 155-163, 2009.

Xiaojuan Jiang is a Ph.D. candidate at Institute of Automation, Chinese Academy of Science. Her research interests include machine learning, network modeling and probabilistic graphical models.



Wensheng Zhang is a professor and Ph.D. supervisor at Institute of Automation, Chinese Academy of Science. His research interests include pattern recognition and machine learning, Big Data mining, probabilistic graphical model, deep neural networks, 3D numerical simulation and video image processing.