

BAYESIAN LATENT FEATURE MODELING FOR MODELING BIPARTITE NETWORKS WITH OVERLAPPING GROUPS

Philip H. Jørgensen, Morten Mørup, Mikkel N. Schmidt, and Tue Herlau

Technical University of Denmark
Department of Applied Mathematics and Computer Science
Kgs. Lyngby, Denmark

ABSTRACT

Bi-partite networks are commonly modelled using latent class or latent feature models. Whereas the existing latent class models admit marginalization of parameters specifying the strength of interaction between groups, existing latent feature models do not admit analytical marginalization of the parameters accounting for the interaction strength within the feature representation. We propose a new binary latent feature model that admits analytical marginalization of interaction strengths such that model inference reduces to assigning nodes to latent features. We propose a constraint inspired by the notion of community structure such that the edge density within groups is higher than between groups. Our model further assumes that entities can have different propensities of generating links in one of the modes. The proposed framework is contrasted on both synthetic and real bi-partite networks to the infinite relational model and the infinite Bernoulli mixture model. We find that the model provides a new latent feature representation of structure while in link-prediction performing close to existing models. Our current extension of the notion of communities and collapsed inference to binary latent feature representations in bipartite networks provides a new framework for accounting for structure in bi-partite networks using binary latent feature representations providing interpretable representations that well characterize structure as quantified by link prediction.

Index Terms— Latent feature modeling, complex networks, bipartite graphs, relational modeling, link prediction.

1. INTRODUCTION

Many natural and man made systems can well be represented in terms of a network. A network consists of nodes, which represent the entities of the system, and edges between the nodes which describe the relationships between the entities. An important problem in the study of networks is to identify latent structure in these complex systems that can sub-

sequently be used to both facilitate our understanding of the systems and for accounting for unknown relations (i.e., predict unobserved edges) in these complex systems.

Many networks are believed to exhibit community structure in which nodes within the same groups are more likely to be connected compared to nodes which belong to different groups [1]. The group structure can be quantified by for instance modularity [2]. Many have studied *disjoint* communities, i.e. a partition of the nodes into groups with no node belonging to more than one group [1], but recently methods to detect *overlapping* communities have become more widely studied, as the assumption that each node belongs to only one group is often not natural [3]. A set of overlapping communities allows the nodes of the network to belong to any number of groups, instead of only one. Many networks naturally consists of overlapping communities, e.g. a person's social network naturally consist of different groups of people such as family, friends, and co-workers. It has been shown that when analyzing networks that consist of overlapping communities using methods for detecting disjoint communities, the analysis often does not describe the full structure of the network [4], and thus practical methods for identifying overlapping communities can significantly contribute to understanding network structure better. Another property of many networks is that their nodes belongs to two or more domains, where only connections between the domains exist. A network with two domains is known as a bipartite network. This is again significantly different from a unipartite network, and requires specialized approaches to detect communities [5].

In recent years many approaches have also been proposed for the modeling of structure in networks within the machine learning community using Bayesian modeling approaches (see e.g. [6]). These approaches have either been based on latent class or latent feature models. Latent class models such as the infinite relational model [7, 8] assumes the network can be represented in terms of latent groups such that edges are independent given the assignment of nodes to groups. Latent feature models in contrast assume each node can be represented in terms of a latent low-dimensional feature representation. Prominent approaches include low-

MM, MNS, and TH were supported by the Lundbeck Foundation, grant no. R105-9813.

rank approximation procedures such as non-negative matrix factorization [9] and latent feature models based on binary feature representations. The latter includes the multiple-membership stochastic block model [10] and extensions to bi-partite systems [11], as well as the binary matrix factorization [12], the latent feature relational model [13], and the latent attribute model [14]. A benefit of latent class models is that the inference can be reduced to inferring assignment of nodes to groups. Unfortunately, the existing binary latent feature models do not admit analytical marginalization of parameters specifying interactions between the extracted features and have to be treated as part of the inference.

In this paper we propose a new latent binary feature model, the *Bayesian latent feature model* (BLFM), for modeling bipartite networks that allows analytical marginalization of parameters specifying the interactions. The main assumption of the proposed model is that the nodes can belong to several overlapping groups such that if two nodes belong to the same group they will have a higher edge-probability than otherwise. We will furthermore assume the edge-probabilities are modulated with a node-specific parameter to account for some entities being more popular than other entities. The particular parameterizations of the proposed model allows marginalization of most parameters leaving a simple discrete sampling problem. On synthetic and real network we demonstrate the utility of our approach compared to existing marginalized procedures given by the two latent class models, the infinite-relational model (IRM) and the infinite Bernoulli mixture model (IBM) [7]. In particular, we find that our approach admit interpretable feature representations while the model representation predicts edges close to the performance of the IRM and IBM.

2. MODEL

Our model describes the structure of the bipartite network in terms of latent attributes. Suppose the bipartite network is described as a $I \times J$ adjacency matrix \mathbf{A} such that if there is an edge between node i and node j then $A_{ij} = 1$ and otherwise $A_{ij} = 0$. Each node in the network can have zero or more of D latent attributes independently of each other. Say, the network denotes customers and goods in a market-basket setting. Then, if a customer and a good share the same attribute this will increase the edge probability compared to the case where they do not share a latent attribute. Let us denote the event customer i has attribute d by $z_{id} = 1$ (and otherwise $z_{id} = 0$) and similarly the event that good j has attribute d by $q_{jd} = 1$. Then, if $z_{id} = q_{jd} = 1$ for any d , the link probability of customer i and good j will be ϕ_j^+ and otherwise ϕ_j^- . Denoting the probability of a link between customer i and good j by ϕ_{ij} we may write

$$\phi_{ij} = \begin{cases} \phi_j^+ & \exists d. z_{id} = q_{jd} = 1, \\ \phi_j^- & \text{otherwise.} \end{cases} \quad (1)$$

Thus, the probabilities are specific to each of the J goods. If we collect all assignments of latent attributes into matrices $\mathbf{Z} = (z_{id})_{id}$ and $\mathbf{Q} = (q_{jd})_{jd}$ we can write

$$p(A_{ij} = 1 | \mathbf{Q}, \mathbf{Z}) = \phi_{ij} = (\phi_j^+)^{I_{ij}} (\phi_j^-)^{1 - I_{ij}}, \quad (2)$$

where $I_{ij} = I(\sum_{d=1}^D z_{id} q_{jd})$ and $I(\cdot)$ is the indicator function.

To complete our model description we need prior distributions for \mathbf{Q}, \mathbf{Z} . We will simply assume that these depend only on the latent attributes such that

$$p(z_{id} = 1) = \theta_d^z, \quad p(q_{jd} = 1) = \theta_d^q. \quad (3)$$

As these parameters are attribute specific, irrelevant attributes may be pruned. For convenience, we furthermore assume that the probability variables $\phi = (\phi_j^\pm)_j$ and $\theta = (\theta_d^z)_d, (\theta_d^q)_d$ are all independently Beta distributed such that

$$\phi_j^\pm \sim \text{Beta}(a_1^\pm, a_2^\pm), \quad \theta_d^{z,q} \sim \text{Beta}(a_1^{z,q}, a_2^{z,q}), \quad (4)$$

and we collect all eight hyperparameters for the Beta distributions into the vector $\rho = (a_c^+, a_c^-, a_c^z, a_c^q)_{c=1,2}$. To complete the hierarchical description of the model, each entry in ρ is given a Gamma(1, 1) prior distribution. The joint distribution then becomes

$$p(\mathbf{A}, \mathbf{Z}, \mathbf{Q}, \theta^{z,q}, \phi^\pm, \rho) = p(\mathbf{A} | \mathbf{Z}, \mathbf{Q}, \phi^\pm) p(\mathbf{Z} | \theta^z) \times p(\mathbf{Q} | \theta^q) p(\theta^{z,q}, \phi^\pm | \rho) p(\rho). \quad (5)$$

This parameterization of the model is illustrated as a Bayesian network in Fig. 1

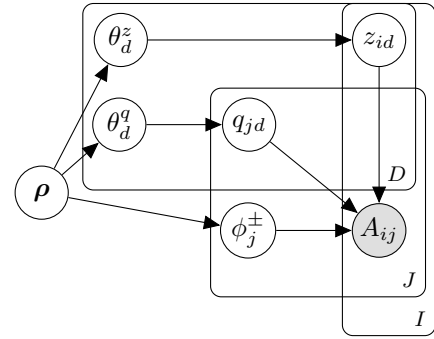


Fig. 1. The Bayesian network of the BLFM.

2.1. Restricted model

The parameters ϕ_j^+ and ϕ_j^- are interpreted as the probability of an edge if i, j share a latent attribute or if they do not, respectively. Thus, it can be sensible to assume that $\phi_j^+ \geq \phi_j^-$ to guarantee that shared features are interpretable as increasing the edge probability, similar to the notion of community structure where edge density is higher within than outside a

group (see also [15]). In addition to the unconstrained case given above we therefore also consider a variant of the model where the constraint $\phi_j^+ \geq \phi_j^-$ is enforced through the prior:

$$p(\phi^+, \phi^-) = \frac{1}{Z} p(\phi^+) p(\phi^-) I(\phi^+ \geq \phi^-) \quad (6)$$

where $p(\phi^\pm)$ is the standard Beta prior (parameters omitted for brevity) and the normalization can be computed as

$$\mathbf{Z} = \frac{\Gamma(a_1^+ + a_1^-) \Gamma(a_2^+ + a_2^-)}{a_2^+ a_1^- \Gamma(a_1^+ + a_2^+ + a_1^- + a_2^- - 1)} \quad (7)$$

$$\times {}_3F_2(1, 1 - a_1^+, 1 - a_2^-; a_2^+ + 1, a_1^- + 1; 1).$$

where $\Gamma(\cdot)$ and ${}_3F_2(\cdot)$ are the gamma and generalized hypergeometric functions respectively.

For efficient inference, the conjugacy of the Beta and Bernoulli distribution allows us to analytically marginalize the parameters ϕ^\pm . In the unconstrained case this yields a ratio of Beta functions, and in the constrained case a ratio of the normalization constants in Eq. (8). In the unconstrained case this leaves a distribution of the form

$$\log p(\mathbf{A}, \mathbf{Z}, \mathbf{Q} | \rho) = \quad (8)$$

$$\sum_j \left[\mathcal{B}(a_1^+, a_2^+, L_{1,j}^+, L_{2,j}^+) + \mathcal{B}(a_1^-, a_2^-, L_{1,j}^-, L_{2,j}^-) \right]$$

$$+ \sum_d \left[\mathcal{B}(a_1^z, a_2^z, C_d^{z+}, C_d^{z-}) + \mathcal{B}(a_1^q, a_2^q, C_{1,d}^q, C_{2,d}^q) \right],$$

with the shorthand $\mathcal{B}(x, y, X, Y) = \log \frac{\mathcal{B}(x+X, y+Y)}{\mathcal{B}(x, y)}$ where $\mathcal{B}(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ is the beta function. For completeness, the count statistics are given as

$$L_{1,j}^+ = \sum_i A_{ij} I_{ij}, \quad L_{2,j}^+ = \sum_i (1 - A_{ij}) I_{ij}, \quad (9)$$

$$L_{1,j}^- = \sum_i A_{ij} (1 - I_{ij}), \quad L_{2,j}^- = \sum_i (1 - A_{ij}) (1 - I_{ij}),$$

$$C_{1,d}^z = \sum_i z_{id}, \quad C_{2,d}^z = \sum_i (1 - z_{id}),$$

and similarly for $C_{1,d}^q, C_{2,d}^q$. The constrained model is identical except the two first \mathcal{B} functions are replaced by the log ratio of the normalization constants in Eq. (8) computed over the same count statistics.

2.2. Inference

Inference of the likelihood Eq. (9) is carried out by Gibbs sampling the binary matrices \mathbf{Z} and \mathbf{Q} and applying a suitable sampling procedure for the eight-dimensional vector ρ . Due to the low dimensionality of ρ the majority of the cost is used sampling \mathbf{Z} and \mathbf{Q} and so we simply opted for random-walk Metropolis Hastings for the parameters in ρ . For each sweep of \mathbf{Z}, \mathbf{Q} we applied 20 sampling steps for the parameters in ρ where a log-transformation $\rho \rightarrow \log \rho$ was applied for numerical stability.

2.2.1. k -flip

Here we propose an alternative sampling approach. Instead of sampling each latent attribute per iteration of the sampler, one could sample a randomly chosen subset of the latent attributes of size k . This requires the evaluation of the log joint probability 2^k times to sample such a subset. If we denote N as the sum of number of nodes in each domain, this is $I + J$, then an iteration of the Gibbs sampler requires $\mathcal{O}(N \times D)$ evaluations of the log joint probability, while the k -flip sampler needs $\mathcal{O}(N \times 2^k)$. By choosing k small, the k -flip sampler can more quickly estimate a new sample, however this new sample will most likely be more similar to the previous sample compared to a new sample from the Gibbs Sampler since an iteration will not consider all latent attributes.

Since the indicator function $I(\cdot)$ only changes when the number of shared latent attributes for two nodes goes from zero to nonzero, or vice versa, only the count statistics affected by such a change have to be updated for evaluating the log joint probability. This property can be used to make the sampler more efficient.

2.3. Predictions

The model will be evaluated in terms of link prediction. This is easily done using the marginalized expression for the likelihood Eq. (9) using S samples of \mathbf{Q}, \mathbf{Z} and using the MCMC approximation for a missing edge A_{ij} :

$$p(A_{ij} = 1 | (\mathbf{Q}^{(s)}, \mathbf{Z}^{(s)})_{s=1}^S)$$

$$\approx \frac{1}{S} \sum_{s=1}^S p(A_{ij} = 1 | \mathbf{A} \setminus A_{ij}, \mathbf{Z}^{(s)}, \mathbf{Q}^{(s)}). \quad (10)$$

It is assumed the sufficient count-statistics is computed for the matrix with the missing edges removed.

3. RESULTS

We will compare the *Bayesian latent feature model* (BLFM) to the infinite-relational model (IRM) and the Bernoulli Mixture model (IBM). For evaluation we will consider both artificially generated data sets from the three models as well as two real data sets, namely the *Animals with Attributes* (AWA) data set originally collected for a psychological experiment [16], but extended and analyzed with the IRM model by *Kemp et al.* [7] and the brain connectivity matrix from the *the Allen Mouse Brain Connectivity Atlas*¹ (MBC) [17]. All the following data processing and computations were carried out with MATLAB². Each sampling trial was evaluated for an equal amount of time between the regular Gibbs sampler and the proposed k -flip sampler.

¹Allen Mouse Brain Atlas [internet]. Seattle (WA): Allen Institute for Brain Science. c 2009. Available from: <http://mouse.brain-map.org>.

²MATLAB Release 2015b, The MathWorks, Inc., Natick, Massachusetts, United States.

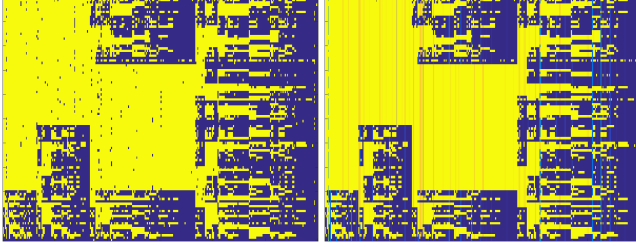


Fig. 2. Recovery of structure by the BLFM when applied to data sets generated from the BLFM model. Left: adjacency matrix \mathbf{A} sorted according to true latent attributes. Right: learned link-probabilities for node (i, j) sorted according to true latent attribute assignments.

3.1. Synthetic Data

In the first set of simulations we will consider data sets generated from each of the three models. In each case we considered ten data sets of dimensions 100×300 . For the BLFM we selected the hyperparameters as $a_1^+ = 10, a_2^+ = \frac{1}{10}, a_1^- = \frac{1}{10}, a_2^- = 10, a_1^z = 1, a_2^z = 5, a_1^q = 1, a_2^q = 5$ and the latent dimension of $D = 20$. To generate data from the IRM and IBM models, the hyperparameters values $a_1 = 1$ and $a_2 = i$ for $i = 1, 2, \dots, 10$ was used for the Beta distributions to create increasingly sparse networks and $\alpha = D/\log(n)$ for the Chinese restaurant process, where n is the number of nodes in the given domain.

3.1.1. Recovery of planted structure

To verify the correctness of the sampler, we will compare the planted structure (i.e. the \mathbf{Z} and \mathbf{Q} matrices) in the generated data sets to those found by the BLFM sampler, where we consider the maximum a posteriori sample. To the left in Fig. 2 such a comparison is visualized by showing \mathbf{A} sorted according to the true matrices \mathbf{Z}, \mathbf{Q} . The estimated probabilities for links when \mathbf{A} is sorted according to the true structure is shown to the right, strongly matching the pattern of links in the true \mathbf{A} . In Fig. 3 a different artificial data set is considered. In the figure, the adjacency matrix \mathbf{A} is sorted according to the MAP estimate of \mathbf{Z}, \mathbf{Q} and each latent group indicated by a color at the top and to the left of \mathbf{A} , i.e. the group assignment of a given link can be found by considering which colored groups are to the left and above it in the matrix. The visualization indicates the larger groups share many links internally as expected.

3.1.2. Pruning Properties

To investigate if the model can recover the true number of groups D we considered experiments where D was selected as 50 in the sampler while the true value of D used to generate

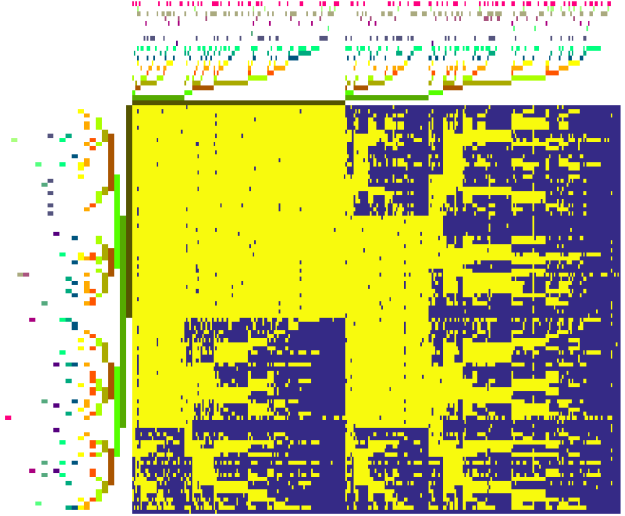


Fig. 3. The adjacency matrix \mathbf{A} and the latent attributes. Each group is represented by a color, and the group assignment of vertices connected by a given edge can be found by considering the colors above or to the left of the edge.

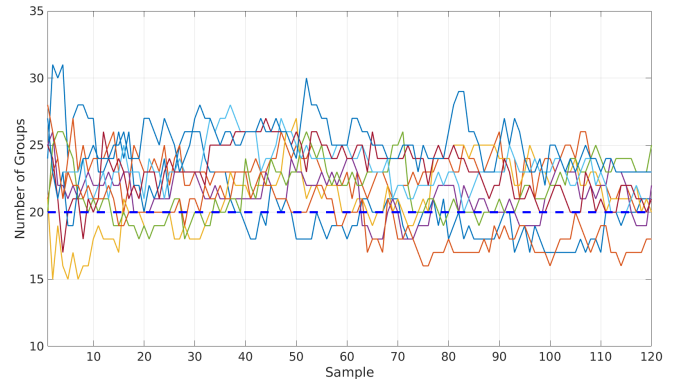


Fig. 4. Number of groups in samples from the BLFM model. The figure shows the samples of ten different runs. All of them end up activating approximately the true number of groups of twenty.

the data was 20 (the other hyperparameter was selected as above). The result can be seen in Fig. 4 for ten repetitions of the BLFM sampler. As seen, the sampler accurately recovers the true number of groups.

3.1.3. Edge prediction

For link prediction, the three models were evaluated on the artificial data generated as described in section 3.1 for each

Table 1. AUC scores for the restricted* Bayesian Latent Feature Model (BLFM), Infinite Relation Model (IRM), and the Bernoulli Mixture Model (IBM) on data generated from each of the three models.

DATA SET	BLFM*	IRM	IBM
BLFM*	0.9961	0.9668	0.9582
IRM	0.6777	0.7388	0.7115
IBM	0.6711	0.7348	0.7347

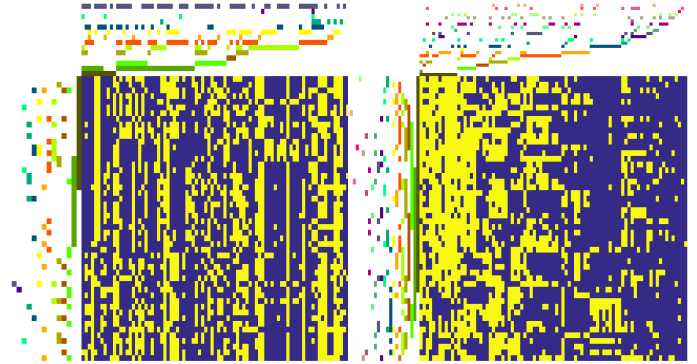
of the models. We evaluated the models for 40000 seconds on each data set with 10% randomly chosen edges marked as missing. This procedure was repeated five times and averages by each trial is reported. The AUC scores (averaged over the five trials) for the results can be seen in table 1. Each model performs best on its own data, except for the IBM data where the IRM performs just as good as the IBM model. The BLFM performs better than the other models on data generated from the BLFM.

3.2. Experimental Data

The models were assessed on two previously mentioned real data sets. The first data set is the *Animals with Attributes* (AWA) data set consisting of a network which represents 50 animals as one domain and 85 traits they have as the second domain. The second data set, *the Allen Mouse Brain Connectivity Atlas* (MBC) represents structural connectivity within 1231 mouse brains estimated through stereotaxic injections for both ipsilateral and contralateral projections resulting in two 213×213 binary matrix. Each model was evaluated for 40000 seconds over 5 trials with 10% data marked as missing for link prediction in every trial. The maximum number of latent groups D for BLFM was set to a higher value than the highest estimate of groups found by the IRM model, such that the model would learn the number of groups by itself. Specifically $D = 50$ for AWA and $D = 170$ and $D = 150$ respectively for MCB Contra/Ipsi.

3.2.1. The Effect of the Restricted Model

Performing inference with and without the restriction on the link probabilities yields the result in table 2. The restricted model has a slightly lower performance compared to the unrestricted model. However the restricted model enforce community structure, which allows for an easier interpretation of the result. An example of the difference between the models can be seen in figure Fig. 5, where the MAP result of respectively the unrestricted and restricted BLFM can be seen. It is clear the restricted model identifies community structures to a much greater extent than the unrestricted.



(a) unrestricted BLFM (b) restricted BLFM

Fig. 5. Comparison of the result on the AWA data of the unrestricted BLFM (a) and the restricted BLFM (b). The unrestricted version does not show a community pattern, while the restricted version indicates some communities in its sorting.

3.2.2. Predictions on real data

Link prediction results for each of the three data sets (as measured by AUC) is shown in table 2. The BLFM performs better than the IBM on the AWA data but slightly worse than the IRM. On the MCB data both the IRM and IBM performs better than the BLFM.

Table 2. AUC scores for the unrestricted/restricted* Bayesian Latent Feature Model (BLFM), infinite-relation Model (IRM), and the infinite Bernoulli mixture model (IBM) on various data sets.

DATA SET	BLFM	BLFM*	IRM	IBM
AWA	0.8991	0.8917	0.9198	0.8749
MBC (CONTRA)	0.8347	0.8257	0.8587	0.8562
MBC (IPSI)	0.8266	0.8249	0.8502	0.8495

3.2.3. Discussion of found group structure

The IRM model found 11 animal groups and 43 attribute groups in the AWA data, while the IBM found 18 attribute groups and the BLFM found 27 overlapping groups. Each model find different groups. An example of a group identified by the BLFM in the AWA data set is:

Animals: *hamster, polar-bear, mole, skunk, rat, otter, persian-cat, giant-panda, dalmatian, wolf, chihuahua, collie, bobcat, raccoon, squirrel, beaver*

Attributes: *agility, furry, quadrapedal, gray, tail, paws, small, claws, pads, solitary, scavenger*

The adjacency submatrix of this group is shown in Fig. 6, and the estimated link probabilities for each attribute within this group is shown in the top of each column.

The link probabilities provides a good picture of how common a given attribute are among the animals in this particular group. Given these, the group makes sense as some of animals and/or attributes not necessarily would be grouped together by a human.

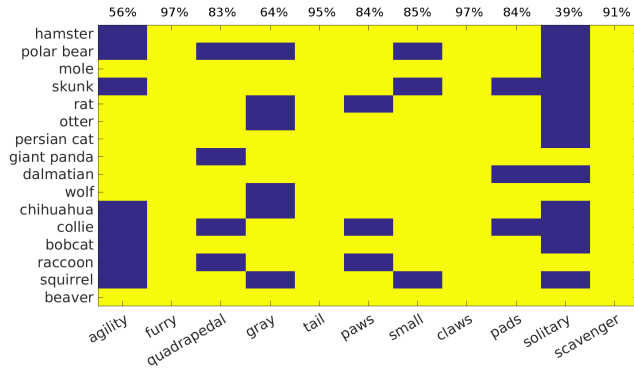


Fig. 6. Adjacency matrix of one of the identified groups by the BLFM MAP estimate. It reveals similar animals and their shared attributes. Estimated link probabilities for each attribute are displayed at the top of each column.

4. DISCUSSION AND CONCLUSION

In this paper, we have proposed a new model for overlapping community structure in bi-partite networks. The main benefits of the proposed model over other approaches to overlapping communities is that it allows analytical marginalization of the continuous parameters thereby reducing Bayesian inference to a simple, discrete sampling problem. We compared the BLFM model against the popular Infinite Relational Model (IRM) and the Infinite Bernoulli Mixture-model (IBM) and found the BLFM to well characterize planted structure and as having performance close to these models in terms of link prediction. We considered two variants of the proposed model, an unconstrained variant where community structure was not enforced and a constrained variant where the community structured was assumed to increase the edge probability. The constrained model has the benefit of easing the interpretability of the latent attributes by ensuring shared attributes increase the edge probability.

5. REFERENCES

[1] Santo Fortunato, “Community detection in graphs,” *Physics Reports*, vol. 486, no. 3-5, pp. 75–174, 2010.

[2] M. E. J. Newman, “Modularity and community structure in networks,” *Proceedings of the National Academy of Sciences*, vol. 103, no. 23, pp. 8577–82, 2006.

[3] Jierui Xie, Stephen Kelley, and Boleslaw K. Szymanski, “Overlapping community detection in networks: the state of the art and comparative study,” *ACM Computing Surveys*, vol. 45, no. 4, pp. 1–35, 2013.

[4] Fergal Reid, Aaron McDaid, and Neil Hurley, “Partitioning breaks communities,” *Proceedings - 2011 International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2011*, pp. 102–109, 2011.

[5] Nan Du, Bai Wang, Bin Wu, and Yi Wang, “Overlapping Community Detection in Bipartite Networks,” *2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, vol. 1, no. 60402011, pp. 176–179, 2008.

[6] Mikkel N. Schmidt and Morten Mørup, “Nonparametric Bayesian modeling of complex networks: An introduction,” *IEEE Signal Processing Magazine*, vol. 30, no. May, pp. 110–128, 2013.

[7] Charles Kemp, Joshua B. Tenenbaum, Thomas L. Griffiths, Takeshi Yamada, and Naonori Ueda, “Learning systems of concepts with an infinite relational model,” *Proceedings of the 21st AAAI conference*, vol. 21, pp. 381–388, 2006.

[8] Zhao Xu, Volker Tresp, Kai Yu, and Hans-Peter Kriegel, “Learning infinite hidden relational models,” *Uncertainty in Artificial Intelligence (UAI2006)*, 2006.

[9] Daniel D. Lee and H. Sebastian Seung, “Algorithms for non-negative matrix factorization,” in *Advances in neural information processing systems*, 2001, pp. 556–562.

[10] Morten Mørup, Mikkel N. Schmidt, and Lars Kai Hansen, “Infinite multiple membership relational modeling for complex networks,” in *Machine Learning for Signal Processing (MLSP), 2011 IEEE International Workshop on*. IEEE, 2011, pp. 1–6.

[11] Christopher Dubois, James Foulds, and Padhraic Smyth, “Latent Set Models for Two-Mode Network Data,” *Fifth International AAAI Conference on Weblogs and Social Media*, pp. 137–144, 2011.

[12] Edward Meeds, Zoubin Ghahramani, Radford M. Neal, and Sam T. Roweis, “Modeling dyadic data with binary latent factors,” in *Advances in neural information processing systems*, 2006, pp. 977–984.

[13] Kurt Miller, Michael I. Jordan, and Thomas L. Griffiths, “Non-parametric latent feature models for link prediction,” in *Advances in neural information processing systems*, 2009, pp. 1276–1284.

[14] Konstantina Palla, David Knowles, and Zoubin Ghahramani, “An infinite latent attribute model for network data,” *arXiv preprint arXiv:1206.6416*, 2012.

[15] Morten Mørup and Mikkel N Schmidt, “Bayesian community detection,” *Neural computation*, vol. 24, no. 9, pp. 2434–2456, 2012.

[16] Daniel N. Osherson, Joshua Stern, Ormond Wilkie, Michael Stob, and Edward E. Smith, “Default Probability,” *Cognitive Science*, vol. 15, no. 2, pp. 251–269, 1991.

[17] Seung Wook Oh, Julie A. Harris, Lydia Ng, et al., “A mesoscale connectome of the mouse brain,” *Nature*, vol. 508, no. 7495, pp. 207–214, 2014.