A Unified Bayesian Model of Community Detection in Attribute Networks with Power-Law Degree Distribution

Shichong Zhang[1], Yinghui Wang[1], Wenjun Wang[1,3], Pengfei Jiao[2], and Lin Pan[4](✉)

[1] College of Intelligence and Computing, Tianjin University, Tianjin, China
{zsc,wangyinghui}@tju.edu.cn

[2] Center for Biosafety Research and Strategy, Law School, Tianjin University, Tianjin, China
pjiao@tju.edu.cn

[3] College of Information Science and Technology, Shihezi University, Shihezi, China
wjwang@tju.edu.cn

[4] School of Marine Science and Technology, Tianjin University, Tianjin, China
linpan@tju.edu.cn

Abstract. Detecting community structure is an important research topic in complex network analysis. How to improve community detection results by using various features in the network is a very challenging problem. The scale-free and attributes of nodes are the two relatively independent aspects of the complex networks in the real world, the former is an inherent structural feature from the global perspective and the later can be used to significantly enhance community detection and community semantics. However, these two aspects are usually modeled and computed independently in previous methods. Based on that, we propose a novel unified Bayesian generative model which combines network topology and node attributes simultaneously to identify community structures via considering to model the scale-free feature. We propose the degree decay variable to preserve the power-law degree characteristic of the network. Specifically, this model composes of two closely correlated parts by a probabilistic transition matrix, one for network topology and the other for nodes attributes. Moreover, we develop a variational EM algorithm to optimize the objective function of the model. Experiments on synthetic and real networks show that our model has a better performance compared with some baselines on community detection in attribute networks.

Keywords: Community detection · Attribute network · Bayesian generative model · Variational EM algorithm · Power-law degree distribution

S. Zhang and L. Pan—Equal contribution.

H. Gao et al. (Eds.): CollaborateCom 2020, LNICST 350, pp. 518–529, 2021.
https://doi.org/10.1007/978-3-030-67540-0_34

1 Introduction

Complex networks are usually denoted and used to represent and analyze a variety of complex systems, such as social systems, biological systems, and ecology systems, etc. Community detection is key to understanding the structure of complex networks, and applications are diverse: from healthcare to regional geography, from human interactions and mobility to economics. Community detection is denoted to divide the nodes of the network into different groups where nodes within a group have more links or similar characteristics. Recent years, many approaches have been proposed to handle the community detection problem, which can be divided into different categories, such as hierarchical clustering algorithms [7], Modularity based optimized approaches, statistical inference [10], etc. A comprehensive and important review can refer to [5].

Based on current research, the real-world complex networks usually demonstrate a variety of features, such as the scale-free property and node attributes [13], which are the key factors that should be considered in the community detection. Many works have shown that making full use of the attribute information of nodes with the topological structure can significantly improve community detection results and it can be used to depict community profiles and semantics [4,16]. However, existing methods only focus on how to improve the community detection results and ignore the maintenance of network structure features. For instance, a generative model for joint identification of network communities and semantics [6] directly used the degree information of nodes, and it did not model the heterogeneity of nodes from the perspective of generation model. Besides node attributes, the power-law degree also an important feature that distributes naturally occurs in real-world networks. Thus how to preserve the heterogeneity of the degree of nodes to model network structures is conducive to community detection. The classical statistical model, such as the stochastic block model (SBM) [10], modeled the cluster structures at a block level and did not take the individuality of nodes into consideration. It treated nodes within a group equally and ignored the heterogeneity. Furthermore, the DC-SBM [1] and the power-law degree SBM [13] were proposed to generate the observed network from the perspective of a single node, i.e. from node-level. However, these methods only focused on the topological structure.

Most previous community detection methods considered only one aspect, and all the above methods ignore combining the scale-free feature and nodes attributes simultaneously for community detection. Actually, considering the heterogeneity of degree and node attributes simultaneously is worth studying in community detection, and faces the following challenges: 1) How to preserve the power-law degree characteristic during modeling the topology, and 2) How to effectively combine the node degree heterogeneity and node attributes to improve community detection results, i.e. jointly encode the varying degree distribution and node multiple attributes to make them complement each other. Only considering one aspect [1,13], modeling network structure or node attributes [16] separately or treating the nodes within a group equally [4] limit the performance for community detection in real-world complex networks.

Keeping the problem in our mind, in this paper, we propose a novel unified and principled Bayesian generated a model to handle the above problem for community detection. When modeling the network structure, we consider the naturally existing scale-free distribution characteristic to reduce the impact of ignoring the property of node degree on the results of community detection. Specifically, we take a global popular parameter for each node to generate the observed network via SBM. We also design a modified topic model for generating the attributes of nodes and semantics of communities and utilize a probabilistic transition matrix to relate network structure and attributes. Thus the model can preserve the power-law degree distributions and fully use node attributes at the same time. We finally propose a variational EM algorithm to optimize the objective function and apply our model on synthetic and real networks respectively. The experimental results show that our model achieves substantial gains compared with existing methods.

2 Related Work

In this section, we study two lines of related work: community detection methods based on attribute networks, and statistical network models especially the works for modeling the heterogeneity of nodes.

Recently, as node attributes have attracted extensive attention in the complex network analysis, some approaches have been proposed combining network topology and node attributes to improve the performance of community detection or community semantics. Ruan et al. [14] proposed a method using content information to determine the strength of the edges between nodes, which can be applied on graph clustering. Yang et al. [16] used a discriminative model combining node attributes and network topology to detect communities. Pool et al. [12] proposed a heuristic method to detect communities by optimizing the community score. Pei et al. [2] developed a model based on the nonnegative matrix tri-factorization method to detect communities via modeling network structure and contents. Chen et al. [4] developed a Bayesian nonparametric attribute (BNPA) model to explore structural regularities in networks. This model combined links and node attributes for community detection via shared hidden variables and assumed network structures and node attributes shared the same community member-ships. These methods failed to give the relevant attributes of each community or semantic descriptions of communities. Jin et al. [9] by distinguishing words from either a background topic or some two-level topics (i.e. general and specialized topics) to help finding communities and provided a semantic community interpretation, but they focused more on words in networked contents. Though He et al. [6] introduced a generative model with two parts, one for communities and the other for semantics to explore network structure and explain the functional modules semantically, the method was only for the network of assortative structures and failed to detect generalized community.

Besides node attributes were considered as important factors in community detection, some research also noticed the characteristics of the network. Most of these methods based on the stochastic block model (SBM), which was a significant benchmark method to detect latent communities according to the observed edge in complex networks. SBM base on the assumption that the probability of links between different nodes only depends on the communities they belong to, so some similarity methods based on the cluster structures in a block-level did not consider the individuality of nodes. In other words, they treated nodes within a group equally. With the degree of heterogeneity becoming a problem worth studying in community detection, several models refined to generate the observed network from the perspective of a single node, i.e. from node-level. DC-SBM [1] introduced a Poisson-valued degree parameter for each node to handle the heterogeneity of node degree, thus networks could be split into heterogeneous. The power-law degree SBM [13] explicitly encoded the power-law feature of networks from the perspective of Bernoulli distribution. Besides, Newman et al. [11] developed a mixture model to explore the network structure, in which the nodes with the same link patterns were divided into the same groups. Thus a broad of structural signatures can be explored without any prior assumptions about the structure of the network. Shen et al. [8] focused on identifying the intrinsic structural rules in networks. In this model, the nodes within the same groups had a similar link preference to other groups.

So we combine the attributes and the power-law of the network to propose a Bayesian probability model for communities detection, let node degree heterogeneity and node attributes can be effectively utilized.

3 Unified Bayesian Model

In this section, we introduce the unified Bayesian model of community detection in attribute networks with power-law degree distribution (PLAC). We give a formal description of the proposed model and introduce the notations, preliminaries related to our model, and the generation process.

3.1 Notations and Preliminaries

Given an undirected attribute network G with N nodes and M attributes, the nodes and attributes are denoted as $V = \big(v_1, v_2 ... v_N\big)$ and $W = \big(\omega_1, \omega_2 ... \omega_M\big)$. The network can be represented by adjacency matrix $A^{N\times N}$ and attributes matrix $X^{N\times M}$, where $a_{ij} = 1$ if there is an edge from node v_i to node v_j and 0 otherwise, $X_{it} = 1$ if node i has the t-th attribute and 0 otherwise. The model contains three types of parameters: 1) Observed quantities: including the number of groups or communities K, the number of nodes N, the number of attributes M, the adjacency matrix A, and the attribute matrix X. 2) Latent quantities: including group labels z where z_i denotes the community membership of node v_i, the content memberships g where g_i denotes the topic labels of the v_i, and the degree decay variable δ_i of each node v_i. 3) Model parameters: including $\pi = (\pi_r)_{1\times K}$, where $\pi_r = p(z_i = r)$ is the probability that node v_i belongs to community r. $B = (b_{rr'})_{K\times K}$, where $b_{rr'} = p(z_i = r, z_j = r')$

is the probability that node in community r links to node in community r'. $\eta = (\eta_{rs})_{K\times K}$, where $\eta_{rs} = p(g_i = s|z_i = r)$ denotes the probability that node v_i is in the s-th topic while it belongs to r-th community. $\theta = (\theta_{st})_{K\times M}$, where $\theta_{st} = p(x_{it} = 1|g_i = s)$ denotes the s-th topic generates attribute ω_t. We try to infer the latent communities and its topics based on the observed networks.

3.2 The Generation Process

We define a unified Bayesian probabilistic generative model to handle network topology and node attributes simultaneously. To model the network structure, we need the link probability of node pairs. In classical stochastic block model, the probability of node v_i links to node v_j is b_{z_i,z_j}. Considering the fact of power-law feature in real-world networks, we define the degree decay variable δ_i to each v_i, and each δ_i is assigned an exponential prior Exp(λ), i.e. $p(\delta_i|\lambda) = \lambda e^{-\lambda\delta_i}$. In order to capture the heterogeneous, we use $b_{z_i,z_j}^{1+\delta_i+\delta_j}$ instead.

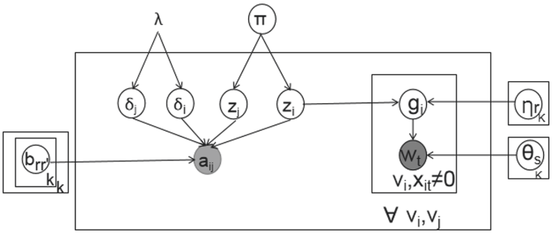

Fig. 1. The graphical representation of PLAC

To jointly model network structure and node attributes, we define the parameter $\eta = (\eta_{rs})_{K\times K}$. In our model, the nodes can be divided into K network communities or K attribute clusters, which called topics in LDA. Meanwhile, attribute topics and network communities may not align. So we use η_r denotes the general topics distribution in community r. When the community label of node v_i is obtained, i.e. $z_i = r$, η_{rs} is the probability that generating the topic of node v_i and $g_i = s$. Thus η provides the transition from communities to topics and correlates two parts closely. When we obtained the topic of node v_i, i.e $g_i = s$, we use θ_s to generate the attributes. Figure 1 is the graphical representation of the model, and the generation process is as follow.

(1) For each new node v_i, $i \in \{1, 2, ..., N\}$:
  (a) Sample a latent group assignment $z_i \sim Multinomial(\pi)$;
  (b) Sample the degree decay variable $\delta_i \sim Exp(\lambda)$;
  (c) Sample the topic assignment $g_i \sim Multinomial(\eta_{z_i})$
(2) For each node-pair $(v_i, v_j) \in \mathcal{N} \times \mathcal{N}$:
  (a) Sample edge $a_{ij} \sim Bernoulli(b_{z_i,z_j}^{1+\delta_i+\delta_j})$;
(3) For each of the t-th attribute ω_t with $x_{it} = 1$:

(a) Sample attribute $\omega_t \sim Multinomial(\theta_{g_i})$.

Then, the probability of the network G is generated by PLAC is:

$$
\begin{aligned}
&p(A, X, z, g, \delta \mid \pi, B, \lambda, \eta, \theta) \\
&\quad = p(A \mid z, B, \delta) p(X \mid g, \theta) p(g \mid z, \eta) p(z \mid \pi) p(\delta \mid \lambda) \\
&\quad = \prod_{a_{ij}=1} b_{z_i,z_j}^{1+\delta_i+\delta_j} \prod_{a_{ij}=0} (1 - b_{z_i,z_j}^{1+\delta_i+\delta_j}) \prod_i (\eta_{z_i g_i}) \prod_{x_{it}=1} (\theta_{g_i t}) \prod_i \lambda e^{-\lambda \delta_i} \pi_{z_i}
\end{aligned}
\tag{1}
$$

subject to $\sum_{r=1}^{K} \pi_r = 1$, $\sum_{s=1}^{K} \eta_{rs} = 1$, and $\sum_{t=1}^{M} \theta_{st} = 1$.

4 Optimization Algorithm

In this section, we use the variational EM algorithm and mean-field method [17] to optimize our model PLAC.

4.1 Variational E-Step

Calculating the observation likelihood $p(A, X \mid \pi, B, \lambda, \eta, \theta)$ is usually analytically intractable because of the integral over all latent variables. So there is no closed-form of joint posterior $p(z, g, \delta \mid A, X, \pi, B, \lambda, \eta, \theta)$. In the variational algorithm, we use a certain distribution family to approximate the real posterior. In our model and set:

$$
q(z, g, \delta) = \prod_i q(z_i) q(g_i) q(\delta_i) \tag{2}
$$

We assume the variables are independent with each other. And we set

$$
q(z_i) = Multi(\phi_i), q(g_i) = Multi(\varphi_i), q(\delta_i) = 1(\bar{\delta_i}) \tag{3}
$$

$q(z_i)$ and $q(g_i)$ are multinomial distributions respectively parameterized with ϕ_i and φ_i. $q(\delta_i)$ is a degenerated distribution with probability 1 at point $\bar{\delta_i}$.

The optimal result can be attained by maximizing the marginal likelihood of observation. It is obtained via minimizing the KL divergence between $q(z, g, \delta)$ and $p(z, g, \delta \mid A, X, \pi, B, \lambda, \eta, \theta)$. We equivalently transform minimization to the maximization of evidence lower bound (ELBO) of the marginal likelihood by using Jensen's inequality and Taylor approximation. The ELBO is:

$$
\begin{aligned}
&\log p(A, X, \mid \pi, B, \lambda, \eta, \theta) \geq \\
&\quad \ell(\phi, \varphi, \delta) = \mathbb{E}_q \log p(A, X, z, g, \delta \mid \pi, B, \lambda, \eta, \theta) - \mathbb{E}_q \log q(z, g, \delta) \\
&\quad \approx \sum_{a_{ij}=1} (1 + \bar{\delta_i} + \bar{\delta_j}) \sum_r \sum_{r'} \phi_{ir} \phi_{jr'} \log b_{rr'} - \sum_{a_{ij}=0} \sum_r \sum_{r'} \phi_{ir} \phi_{jr'} b_{rr'}^{1+\bar{\delta_i}+\bar{\delta_j}} \\
&\quad + \sum_i \sum_r \sum_s \phi_{ir} \varphi_{is} \log \eta_{rs} + \sum_{x_{it}=1} \sum_s \varphi_{is} \log \theta_{st} + \sum_i \sum_r \phi_{ir} \log \pi_r \\
&\quad - \sum_i \sum_r \phi_{ir} \log \phi_{ir} - \sum_i \sum_s \varphi_{is} \log \varphi_{is} + N \log \lambda - \lambda \sum_i \bar{\delta_i}
\end{aligned}
\tag{4}
$$

Verifiably, Eq. (4) only concave to each variable ϕ, φ and $\bar{\delta}_i$, and not concave with $(\phi, \varphi, \bar{\delta}_i)$ simultaneously. So to optimize these variables, we use coordinate gradient ascent. The gradient of $\bar{\delta}_i$ is:

$$\begin{aligned}\frac{\partial \mathcal{O}(\bar{\delta}_i)}{\partial \bar{\delta}_i} &= \sum_r \phi_{ir} \sum_{r'} \log b_{rr'} \sum_{a_{ij}=1} \phi_{jr'} \\ &- \sum_r \phi_{ir} \sum_{a_{ij}=0} \sum_{r'} \phi_{jr'} b_{rr'}^{1+\bar{\delta}_i+\bar{\delta}_j} \ln(b_{rr'}) - \lambda\end{aligned} \tag{5}$$

and the updating of ϕ_i and φ_i is given by:

$$\begin{aligned}\phi_{ir} \propto \pi_r \exp(&\sum_{a_{ij}=1} \sum_{r'} (1 + \bar{\delta}_i + \bar{\delta}_j) \phi_{jr'} \log b_{rr'} \\ - &\sum_{a_{ij}=0} \sum_{r'} \phi_{jr'} b_{rr'}^{1+\bar{\delta}_i+\bar{\delta}_j} + \sum_s \varphi_{is} \log \eta_{rs})\end{aligned} \tag{6}$$

$$\varphi_{is} \propto \prod_{x_{it}=1} \theta_{st} \exp(\sum_r \phi_{ir} \log \eta_{rs}) \tag{7}$$

4.2 M-Step

In this part, we optimize the model parameters (π, λ, B, θ, η). We fix $\lambda = 0.01$ and update π, B, θ and η. The gradient of B is:

$$\frac{\partial \mathcal{O}(b_{rr'})}{\partial b_{rr'}} = \frac{\sum_{a_{ij}=1} (1 + \bar{\delta}_i + \bar{\delta}_j) \phi_{ir} \phi_{jr'}}{b_{rr'}} - \sum_{a_{ij}=0} (1 + \bar{\delta}_i + \bar{\delta}_j) \phi_{ir} \phi_{jr'} b_{rr'}^{\bar{\delta}_i+\bar{\delta}_j} \tag{8}$$

So the optimal π, η, θ are given by:

$$\pi_k \propto \sum_i \phi_{ir}; \qquad \eta_{rs} \propto \sum_i \phi_{ir} \varphi_{is}; \qquad \theta_{st} \propto \sum_{x_{it}=1} \varphi_{is} \tag{9}$$

The predicted community label of each node v_i is obtained by maximizing the variational posterior distribution:

$$z^* = \arg\min_r \phi_{ir} \tag{10}$$

where ϕ_{ir} is the optimal posterior parameter from Eq. (6). The implementation of the above algorithm is shown in Algorithm 1.

5 Experiments and Analysis

We evaluate the performance of PLAC on synthetic and real networks and compared with three other methods. And we use the Normalized Mutual Information (NMI) [3] as the evaluation index.

Algorithm 1. Inference for PLAC

Input: Initialize model parameters λ, B, π, η, θ and parameters ϕ, φ, $\bar{\delta}$; the community number K; stop criterion ε.
Output: Model parameters $\bar{\delta}^*$, π^*, B^*, η^*, θ^* and the community label z^*.

repeat
    Compute variational likelihood ℓ^{new} by Eq. (4).
    $\ell^{old} = \ell^{new}$
    variational E-step
    update $\bar{\delta}$, ϕ, φ respectively by Eq. (5), Eq. (6), Eq. (7).
    M-step
    updte B, π, η, θ respectively by Eq. (8), Eq. (9).
until $|\ell^{new} - \ell^{old}| \leq \varepsilon$
predict the community index z_i of each node by Eq. (10).

5.1 Experiment on Synthetic Networks

The synthetic network is a random network in Newman's model [6]. The network contains 128 nodes divided into 4 disjoint communities with $z_{in} + z_{out} = 16$. As $\rho(= z_{in}/32) > \rho(= z_{out}/96)$, z_{in} (the internal edges) is greater than z_{out} (the external edges). For each v_i, a $4h$-dimensional binary attributes (i.e., x_i) are created to divide 4 attribute clusters in network. h_{in} denotes the number of attributes for every node v_i with $x_{it} = 1$ related to its community and h_{out} (noisy attributes) inversely denotes the number of attributes related to other communities, and $h_{in} + h_{out} = 16$.

We set $h = 50$, network topologies and node attributes share the same membership. Node attributes matrix and the community attributes matrix ($z_{out} = 3$) are shown in the Fig. 2. Firstly, We set $z_{out} = 8$, and h_{out} is changed from 0 to 9 with an increment 1. The larger h_{out} is, the vaguer the structure of node attributes will be.

We adapt PLD-SBM as the baseline method. Other comparison methods are NEMBP [6] and SCI [15]. As shown in Fig. 2(c), our model can fairly well exploit the node attributes information to improve the quality of community detection compare with other models. Then we set $h_{out} = 8$, and change z_{out} from 0 to 9 (the community structure is gradually blurred) with an increment 1 to study the ability to use network structures information. Our method also performs better than NEMBP and SCI. On the whole, Information of attributes and structures can be better used in our model. Figure 3 is the visualization of community detection results. And the results obviously are better compared with other methods.

5.2 Experiment on Real Networks

Five real networks we utilize are shown in Table 1. Texas, Cornell, Washington, and Wisconsin are four sub-networks of the WebKB network. Twitter is the largest sub-network (id629863) in Twitter data.

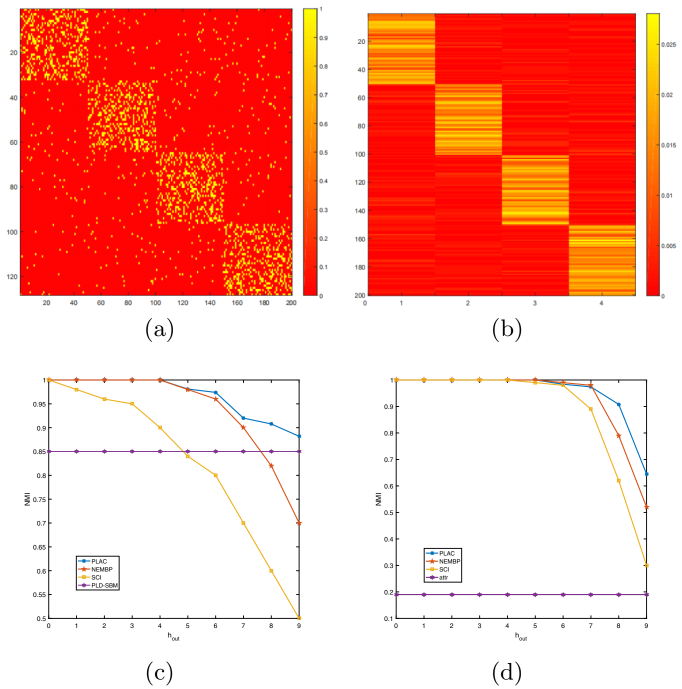

Fig. 2. (a) node attributes matrix. (b) community attributes matrix. The value of NMI of three methods on random networks with the change of (c) h_{out} from 0 to 9 and (d) z_{out} from 0 to 9.

We compare our model PLAC with the methods of three categories: 1) models based on network structures: DCSBM, PLD-SBM. 2) models based on nodes attributes: LDA. 3) models based on both structures and attributes: PCL-DC, NMMA, SCI, NEMBP. Results are shown in Table 2. Our model PLAC outperforms other models on Cornell, Washington and Wisconsin, achieves larger NMIs than most of the models on Texas and Twitter. It is mainly because that the attributes of each community are clear in Texas and the degree of nodes in Twitter does not obey the power-law distribution. Usually, the models based on

Table 1. Statistical characteristics of five real networks.

Datasets	N (nodes)	E (edges)	M (attributes)	K (communities)
Texas	187	328	1703	5
Cornell	195	304	1703	5
Washington	230	446	1703	5
Wisconsin	265	530	1703	5
Twitter	171	796	578	7

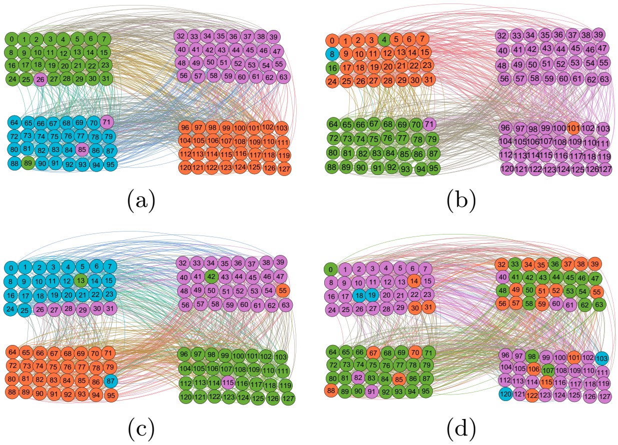

Fig. 3. The four blocks in each diagram indicate the true communities in network. Colors stand for predicted communities by PLAC and NEMBP. Nodes of same colors belong to same communities. The results of community detection: (a) PLAC, $z_{out} = 8$, $h_{out} = 9$. (b) NEMBP, $z_{out} = 8$, $h_{out} = 9$ (c) PLAC, $h_{out} = 8$, $z_{out} = 9$. (d) NEMBP, $h_{out} = 8$, $z_{out} = 9$. (Color figure online)

Table 2. NMI (%) of different models on five real networks

Models	NMI				
	Data				
	Texas	Cornell	Washington	Wisconsin	Twitter
DC-SBM	16.65	9.696	9.87	3.14	57.48
PLD-SBM	25.21	13.92	12.47	17.50	42.20
LDA	34.29	21.09	38.48	46.56	1.67
PCL-DC	10.37	7.23	5.66	5.01	52.64
NMMA	**41.57**	39.95	25.59	4.696	1.67
SCI	17.84	11.44	12.37	17.03	43.00
NEMBP	35.12	18.71	21.24	38.02	**59.73**
PLAC	34.29	**70.16**	**39.17**	**73.40**	44.02

structure and attributes outperform the models with only links or attributes. The results are further improved when considering the power-law degree distributions of each node in networks. Such as NEMBP uses the heterogeneity of node degree to model the network structures, which is associated with DC-SBM. But its optimized equation is dependent on cluster-level degrees rather than node-level rather than encodes degree distribution over each node like our model PLAC. So our method performs better than NEMBP on datasets that have the power-law feature.

5.3 A Case Study

Furthermore, we present a demonstration of community semantic analysis based on the widely used data set on community detection in attribute networks, LASTFM [6] from an online music system Last.fm, which includes 1,892 users and 11,946 attributes, and we set the number of communities is 38 as did in [6]. As represented in Fig. 4, we show the two examples of community interpretation. The first is a community of fans who may like heavy metal rock music, and the users in the second community prefer a more relaxed and relaxing style of pure music. We give that based on the parameters η and θ.

Fig. 4. Examples on community semantic in two different music domain based on the LASTFM data, the word sizes are proportional to the probability they belong to the corresponding topic: (a) topic 2 on 7th community. (b) topic 7 on 21th community

6 Conclusions

In this paper, we proposed a novel Bayesian probability model to detect communities in attribute networks with power-law degree distributions and token a variational EM algorithm to optimize the objective function of the model. We presented a demonstration of community interpretation by applying our model on a specific scene. It is similar to the existing methods that the number of communities needs to be given in advance, and we will focus on how to automatically determine the number of communities for this model in the future.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (61902278), the National Key R&D Program of China (2018YFC0832101), the Tianjin Science and Technology Development Strategic Research Project under Grant (18ZXAQSF00110) and National Social Science Foundation of China (15BGL035).

References

1. Brian, K., Newman, M.E.J.: Stochastic blockmodels and community structure in networks. Phys. Rev. E Stat. Nonlinear Soft Matter Phys. **83**(2), 016107 (2011)

2. Chakraborty, Y.P.N., Sycara, K.: Nonnegative matrix tri-factorization with graph regularization for community detection in social networks. In: Twenty-Fourth International Joint Conference on Artificial Intelligence, pp. 2083–2089. AAAI Press (2015)
3. Chang, Z., Yin, X., Jia, C., Wang, X.: Mixture models with entropy regularization for community detection in networks. Phys. A: Stat. Mech. Appl. **496**, 339–350 (2018)
4. Chen, Y., Wang, X., Bu, J., Tang, B., Xiang, X.: Network structure exploration in networks with node attributes. Phys. A: Stat. Mech. Appl. **449**, 240–253 (2016)
5. Fortunato, S., Hric, D.: Community detection in networks: a user guide. Phys. Rep. **659**, 1–44 (2016)
6. He, D., Feng, Z., Jin, D., Wang, X., Zhang, W.: Joint identification of network communities and semantics via integrative modeling of network topologies and node contents. In: Thirty-First AAAI Conference on Artificial Intelligence (2017)
7. He, D., Yang, X., Feng, Z., Chen, S., Fogelman-Soulié, F.: A network embedding-enhanced approach for generalized community detection. In: Liu, W., Giunchiglia, F., Yang, B. (eds.) KSEM 2018. LNCS (LNAI), vol. 11062, pp. 383–395. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-99247-1_34
8. Hua-Wei, S., Xue-Qi, C., Jia-Feng, G.: Exploring the structural regularities in networks. Phys. Rev. E Stat. Nonlinear Soft Matter Phys. **84**(2), 056111 (2011)
9. Jin, D., Wang, K., Zhang, G., Jiao, P., He, D., Fogelman-Soulie, F., Huang, X.: Detecting communities with multiplex semantics by distinguishing background, general and specialized topics. IEEE Trans. Knowl. Data Eng. **32**, 2144–2158 (2019)
10. Karrer, B., Newman, M.E.: Stochastic blockmodels and community structure in networks. Phys. Rev. E **83**(1), 016107 (2011)
11. Newman, M.E.J., Leicht, E.A.: Mixture models and exploratory analysis in networks. Proc. Natl. Acad. Sci. USA **104**(23), 9564–9569 (2007)
12. Pool, S., Bonchi, F., Leeuwen, M.V.: Description-driven community detection. ACM Trans. Intell. Syst. Technol. (TIST) **5**(2), 28 (2014)
13. Qiao, M., Yu, J., Bian, W., Li, Q., Tao, D.: Adapting stochastic block models to power-law degree distributions. IEEE Trans. Cybern. **49**(2), 626–637 (2019)
14. Ruan, Y., Fuhry, D., Parthasarathy, S.: Efficient community detection in large networks using content and links. In: International Conference on World Wide Web, pp. 1089–1098 (2013)
15. Wang, X., Jin, D., Cao, X., Yang, L., Zhang, W.: Semantic community identification in large attribute networks. In: Thirty-Second AAAI Conference on Artificial Intelligence, pp. 265–271 (2016)
16. Yang, T., Jin, R., Chi, Y., Zhu, S.: Combining link and content for community detection. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 927–936 (2009)
17. Yu, Z., Zhu, X., Wong, H.S., You, J., Zhang, J., Han, G.: Distribution-based cluster structure selection. IEEE Trans. Cybern. **47**(11), 3554–3567 (2017)