Improving Stochastic Block Models
by Incorporating Power-Law Degree Characteristic

Maoying Qiao[1], Jun Yu[1], Wei Bian[2], Qiang Li[2,3], and Dacheng Tao[4]

[1] School of Computer Science, Hangzhou Dianzi University

[2] School of Software, FEIT, The University of Technology Sydney

[3] Department of Computing, The Hong Kong Polytechnic University

[4] UBTECH Sydney Artificial Intelligence Institute, School of IT, FEIT, The University of Sydney

{qiaomy, yujun}@hdu.edu.cn, {brian.weibian, leetsiang.cloud, dacheng.tao}@gmail.com

Abstract

Stochastic block models (SBMs) provide a statistical way modeling network data, especially in representing clusters or community structures. However, most block models do not consider complex characteristics of networks such as scale-free feature, making them incapable of handling degree variation of vertices, which is ubiquitous in real networks. To address this issue, we introduce degree decay variables into SBM, termed power-law degree SBM (PLD-SBM), to model the varying probability of connections between node pairs. The scale-free feature is approximated by a power-law degree characteristic. Such a property allows PLD-SBM to correct the distortion of degree distribution in SBM, and thus improves the performance of cluster prediction. Experiments on both simulated networks and two real-world networks including the Adolescent Health Data and the political blogs network demonstrate the validity of the motivation of PLD-SBM, and its practical superiority.

1 Introduction

In recently years, both theoretical and computational studies on network analysis have been rapidly growing in a wide range of areas, including bioinformatics, academic collaboration and social media [Goldenberg *et al.*, 2010]. Statistical models have a relatively long history under this category in discovering hidden structural knowledge of networks via dealing with uncertainties by using statistical inference [Goldenberg *et al.*, 2010].

An important class of statistical models for modeling latent cluster structures in network data is stochastic block models (SBMs)[Shiga and Mamitsuka, 2012; Bu *et al.*, 2016]. From a generative perspective, a network with a certain cluster structure can be constructed given the parameters defined by its block portions and connectivity, where nodes are chosen from a multinomial distribution while edges between pairs of nodes are generated from a Bernoulli distribution (simple graphs), or a Poisson distribution (multigraphs) [Tang *et al.*, 2012], or exponential family (valued graphs) [Mariadassou *et al.*, 2010]. From an inference perspective, posterior inference

is proposed by [Snijders and Nowicki, 1997] to uncover cluster structures. SBM has been extensively studied. [Kemp *et al.*, 2006] introduced nonparametric Bayesian methods into SBM to adaptively deal with an unknown number of clusters, by imposing a Chinese Restaurant Process (CRP) prior over the node partition. Mixed membership SBM (MMSBM) has been developed [Airoldi *et al.*, 2008] to handle the situation when clusters overlap. Hierarchical SBM [Ho *et al.*, 2011] integrates node attributes into SBM. [Kim *et al.*, 2012] proposed a dynamic infinite extension of MMSBM [Fan *et al.*, 2015], and to improve model scalability, stochastic variational methods [Peng *et al.*, 2015] have been introduced.

However, due to the block property, the traditional SBM is incapable to handle naturally existed scale-free feature of real-world networks, i.e., node degrees characterized by power-law distributions [Barabási and Albert, 1999; Luo *et al.*, 2016]. Take the political blog network [Adamic and Glance, 2005] as a real-world network example. Both the liberal and conservative communities contain different quantities of popular and inactive blogs, so the node degrees in each community are heterogeneously distributed, which can be approximated by power-law distributions. Instead of clustering the nodes into the true communities, SBM tends to group the nodes of degree homogeneity into clusters, namely either high- or low- degree node clusters according to the experimental results in [Karrer and Newman, 2011]. Obviously, such power-law degree distributions within each community cause significant bias when SBM attempts to infer the cluster structures hidden within a network.

To alleviate the homogeneity bias of SBM, different strategies exploring either node degrees or link properties have been investigated. A degree-corrected SBM [Karrer and Newman, 2011] introduces an expected degree parameter for each node into the Poisson-valued SBM to explicitly cope with the degree heterogeneity. However the derived objective function is independent of the node-wise degree parameters, but still only depends on group degrees. Consequently, the inference procedure of DC-SBM does not directly depend on the expected degree parameters of individual nodes. Another typical approach is the Link Density model (LD) [Mørup and Hansen, 2009] built on mixture models of SBM [Airoldi *et al.*, 2008]. It adds another latent layer variables over links, whose probabilities are parameterized with node-specific degrees. Obviously it ignores the power-law characteristic of

real-world networks.

In this paper, we propose a new extension of SBM to address this problem, termed power-law degree stochastic block model (PLD-SBM). It incorporates a power-law prior distribution into the traditional SBM to characterize the scale free feature of real-world networks, and ultimately to correct the degree homogeneity bias exhibited in SBM. Model learning and posterior inference in PLD-SBM are implemented with a Viterbi-type variational EM algorithm. Experiments on simulated networks and two real-world networks including the Adolescent Health Data and the political blogs network are conducted to evaluate the effectiveness of PLD-SBM.

2 Power-Law Degree Stochastic Block Model

We consider networks represented by an undirected binary random graph $\mathcal{G}$, with an $N \times N$ binary adjacency matrix Y. The entry y_{ij} equal to 1 or 0 represents the presence or absence of an edge between nodes i and j. It is assumed that each node i is associated with a latent variable $z_i \in \{1, 2, ..., K\}$, as its cluster index within K clusters or communities. SBM then uses a $K \times K$ matrix B to describe the connection probabilities between clusters, i.e., with probability $0 \leq b_{kk'} \leq 1$ a node from cluster k is linked to a node from cluster k'. Therefore, we have the below single-edge likelihood for each observed edge y_{ij}, given its two end nodes' cluster indexes, and the whole-edge likelihood is easily obtained by multiplying over the whole edge set.

$$p(y_{ij} = 1|z_i, z_j, B) = b_{z_i z_j}. \tag{1}$$

A multinomial distribution $\text{Multi}(\pi)$ is assigned as a prior to z_i. The edges in Y are independent conditioning on cluster indexes $\mathbf{z}$ [Bishop, 2006]. Maximum likelihood estimation (MLE) or Bayesian estimation is applied to learn the model parameters B, then a posterior inference on $\mathbf{z}$ reveals the network's cluster structure, i.e., the cluster indexes of nodes in the network [Snijders and Nowicki, 1997].

2.1 The Proposed Model

As mentioned above, SBM is unable to model the skew feature of degree distributions within communities in a network. This is because SBM (1) treats the degrees of nodes within a cluster, for instance cluster k, the same with one group-level parameter, i.e., b_{kk}, not considering the node-specific factors which are favoured by other clustering processes. To this end, we associate each node i with another latent variable, $\delta_i \geq 0$, to adjust its node degree involving probability of edge generation. Specifically, we assume that

$$p(y_{ij} = 1|z_i, z_j, \delta_i, \delta_j, B) = b_{z_i z_j}^{1+\delta_i+\delta_j}. \tag{2}$$

We call δ_i the degree decay variable. Clearly, large δ_i and/or δ_j will make the connection between nodes i and j unlikely. As a result, heterogeneous node degrees are formed if δ_i varies among nodes. To model the diverse values of δ_i, we assign it an exponential prior $\text{Exp}(\lambda)$, i.e.,

$$p(\delta_i|\lambda) = \lambda e^{-\lambda \delta_i}.$$
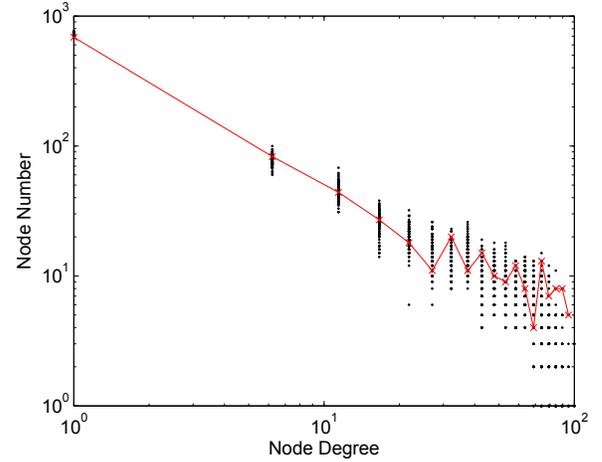
Figure 1: The power-law degree characteristic of PLD-SBM.

Given all hidden variables $\{z_i, \delta_i\}_{i=1}^N$, the edge variables are conditionally independent . Then, the joint probability of observations Y and latent variables $(\mathbf{z}, \delta)$ in PLD-SBM is given by

$$p(Y, \mathbf{z}, \delta|\pi, \lambda, B)$$
$$= \prod_{i,j} p(y_{ij}|b_{z_i, z_j}, \delta_i, \delta_j) \prod_i p(\delta_i|\lambda) p(z_i|\pi)$$
$$= \prod_{y_{ij}=1} b_{z_i, z_j}^{1+\delta_i+\delta_j} \prod_{y_{ij}=0} [1 - b_{z_i, z_j}^{1+\delta_i+\delta_j}] \prod_i \lambda e^{-\lambda \delta_i} \pi_{z_i}.$$

Before developing algorithms for estimation and inference, we first provide some insight into how the introduction of degree decay variables $\delta_i, i = 1, \ldots, N$ helps model power-law degree distributions.

2.2 Degree Characteristic

Denoting $p(d)$ as the degree distribution of a network, its power-law degree characteristic is defined as

$$p(d = k) \propto k^{-\gamma}, \tag{3}$$

where γ is a shape parameter. As (3) is invariant to scale transformation, a network obeying (3) is also regarded as scale-free. It should be noted that we are not proving PLD-SBM to be a "rigorous" scale-free model, rather that it can improve SBM by incorporating it to power-law degree distributions.

Since the contribution to the node degrees is mainly from intra-cluster edges, a basic assumption of clustering and community detection in assortative networks, we consider an intra-cluster or equivalently a single-cluster case to justify the ability of PLD-SBM for power-law degree modeling. Suppose a cluster has n_0 nodes, each associated with a degree decay variable $\delta_i \sim \text{Exp}(\lambda)$, and the connection probability between two nodes is p_0. Then, with PLD-SBM (2), it can be shown that as n_0 increases the normalized degree of node i will converge to a random variable $\bar{d}_i$ that only depends on

δ_i, i.e.,

$$\frac{1}{n_0 - 1} \sum_{j \neq i} y_{ij} \xrightarrow{a.s.} \bar{d}_i(\delta_i) = \frac{\lambda}{\lambda - \ln p_0} p_0^{1+\delta_i}. \quad (4)$$

This is a direct result of the Strong Law of Large Numbers (SLLN). Using the fact that δ_i follows the exponential prior distribution, we can verify that the distribution of $\bar{d}_i$ obeys a power-law distribution

$$p(\bar{d}_i = k) \propto k^{-(1+\frac{\lambda}{\ln p_0})}, \quad (5)$$

with the shape parameter $\gamma = 1 + \lambda/\ln p_0$. We call (5) PLD-SBM's power-law degree characteristic.

Remark 1 The power-law degree characteristic (5) of PLD-SBM is only for the degree distribution of an individual node, not statistic overall as given by (3). However, in a sparse network (valid for most real networks), $\bar{d}_i$, $i = 1, 2, ..., N$, are nearly independent, which makes the overall degree distribution (a statistic over i.i.d. examples) similar to that of an individual one, i.e., (5). Indeed, as demonstrated by the simulation result in Figure 1, the networks generated by PLD-SBM do follow a power-law degree distribution overall. We generated 100 single-cluster networks, where we fixed the node number $n_0 = 1000$, the exponential distribution parameter $\lambda = 0.01$, and the probability for linking any node pairs $p_0 = 0.9$. We calculated the empirical degree distribution, i.e., node number vs node degree, and found that it concentrates approximately along a line in the log-domain, with a slope about -0.892. When using (5), the slope is $-(1 + \lambda/\ln p_0) = -0.905$ - very close to the simulation result.

Remark 2 The maximum value of the shape parameter $\gamma = 1 + \lambda/\ln p_0$ in (5) approaches 1 given a small λ, which is smaller than the typical value for real networks, lying between 2 and 4. However, smaller shape parameters enable PLD-SBM to adapt to much severer heavy-tail cases.

3 Algorithm

In this section, we derive algorithms for the implementation of PLD-SBM. Although a Bayesian approach, with sampling techniques such as MCMC methods [Qiao *et al.*, 2015], could be applied to latent variable model inference, recent studies show that optimization-based approaches using variational methods are more efficient [Airoldi *et al.*, 2008]. Below, we derive a Viterbi-type variational EM algorithm for inference and learning in PLD-SBM.

3.1 Inference

Given edge observations, i.e., the adjacency matrix Y, and model parameters (π, λ, B), we want to infer the posterior distribution of latent variables $(\mathbf{z}, \delta)$. Since the calculation of observation likelihood $p(Y|\pi, \lambda, B)$ is generally intractable, the posterior distribution $p(\mathbf{z}, \delta|Y, \pi, \lambda, B)$ cannot be obtained analytically. Following the mean-field method, we use a fully factorized distribution family:

$$q(\mathbf{z}, \delta) = \prod_i q(z_i) q(\delta_i). \quad (6)$$

The above equation assumes all latent variables are independent from each other. The variational distributions $q(z_i)$ and $q(\delta_i)$ are generally specific to the same family of prior distributions to take advantage of the conjugate property between likelihood and prior to reduce computation complexity. However, this is not the case in our problem. By contrast, we specify

$$q(z_i) = \text{Multi}(\phi_i) \text{ and } q(\delta_i) = \mathbf{1}(\bar{\delta}_i), \quad (7)$$

where we let $q(z_i)$ be a multinomial distribution with the parameters ϕ_i and $q(\delta_i)$, a degenerated distribution with probability 1 at point $\bar{\delta}_i$. The choice for $q(\delta_i)$ is based on whether it should be concentrated somewhere in the positive half real line in the posterior sense. Clearly, the exponential distribution (as the prior) is not suitable; and other options, e.g., truncated Gaussian, are not computationally favorable. Actually, the use of degenerated distribution here is equivalent to using MAP estimation, which is similar to the Viterbi-type EM algorithm used for training Hidden Markov Models (HMMs).

The optimal variational distribution that approximates the true posterior within family (6) is obtained by minimizing the Kullback-Leibler (KL) divergence between the variational posterior distribution $q(\mathbf{z}, \delta)$ and the true posterior distribution $p(\mathbf{z}, \delta|Y, \pi, \lambda, B)$. Often, minimization is equivalently transformed to the maximization of an evidence lower bound (ELBO) of the log likelihood of observations. Specifically, applying Jensen's inequality, the ELBO $\mathcal{L}(\phi, \delta)$ for our problem is given by

$$\log p(Y|\pi, \lambda, B) \geq$$
$$\mathcal{L}(\phi, \delta) = \mathbb{E}_q \log p(Y, \mathbf{z}, \delta|\pi, \lambda, B) - \mathbb{E}_q \log q(\mathbf{z}, \delta).$$

By (7) and the Taylor approximation [Ahmed and Xing, 2007], we get:

$$\begin{aligned}
\mathcal{L}(\phi, \delta) \approx & \sum_{y_{ij}=1} (1 + \bar{\delta}_i + \bar{\delta}_j) \sum_k \sum_{k'} \phi_{ik} \phi_{jk'} \log b_{kk'} \\
& - \sum_{y_{ij}=0} \sum_k \sum_{k'} \phi_{ik} \phi_{jk'} b_{kk'}^{1+\bar{\delta}_i+\bar{\delta}_j} \\
& + \sum_i \sum_k \phi_{ik} \log \pi_k - \sum_i \sum_k \phi_{ik} \log \phi_{ik} \\
& + N \log \lambda - \lambda \sum_i \bar{\delta}_i. \quad (8)
\end{aligned}$$

Note that (8) is not jointly concave w.r.t. $(\phi, \bar{\delta})$. However, it is not hard to verify that it is concave w.r.t. each individual variable ϕ_i and $\bar{\delta}_i$, and thus we can apply coordinate gradient ascend method to optimize these variables iteratively. The optimal $\bar{\delta}_i$ is obtained through gradient ascend method, and the gradient is computed with

$$\begin{aligned}
\frac{\partial \mathcal{O}(\bar{\delta}_i)}{\partial \bar{\delta}_i} = & \sum_k \phi_{ik} \sum_{k'} \log b_{kk'} \sum_{y_{ij}=1} \phi_{jk'} \\
& - \sum_k \phi_{ik} \sum_{y_{ij}=0} \sum_{k'} \phi_{jk'} b_{kk'}^{1+\bar{\delta}_i+\bar{\delta}_j} \ln(b_{kk'}) - \lambda. \quad (9)
\end{aligned}$$

The updating equation for optimal ϕ_i is given by

$$\phi_{ik} \propto \pi_k \exp \left\{ \sum_{y_{ij}=1} \sum_{k'} (1 + \bar{\delta}_i + \bar{\delta}_j) \phi_{jk'} \log b_{kk'} \right.$$
$$\left. - \sum_{y_{ij}=0} \sum_{k'} \phi_{jk'} b_{kk'}^{1+\bar{\delta}_i+\bar{\delta}_j} \right\}. \qquad (10)$$

In (10), the two terms in the $\exp(\cdot)$ operator count the contribution of linked nodes and unlinked nodes respectively when inferring the cluster membership of node i, i.e., ϕ_i. Extra weight $\bar{\delta}_j$ is considered when $y_{ij} = 1$, namely nodes i and j are linked, while less weight is dealt when $y_{ij} = 0$. This is because $\bar{\delta}_j \geq 0$, $b_{kk'} \leq 1$, and $b_{kk'}^{1+\bar{\delta}_i+\bar{\delta}_j}$ decreases for large $\bar{\delta}_j$. This differs PLD-SBM from the traditional SBM which treats the contribution of each node j uniformly, and helps correct the degree homogeneity bias by introducing degree decay variables of power-law degree distributions.

3.2 Learning

By variational EM algorithm, the optimization of model parameters (π, λ, B) also applies ELBO (8) as its objective function. A common problem of MLE, e.g., in Gaussian Mixture Models (GMMs), is overfitting. At first glance, PLD-SBM seems unlikely to suffer from overfitting because the number of model parameters is relatively small and it does not clearly have the component degeneration problem of GMMs. However, an implicit overfitting may come from the latent variables δ_i. Note that every δ_i associates with each node in the network, and therefore has the power to over-explain the edge generation for each node. When optimizing the model parameters with an EM algorithm, the optimal λ attempts to fit the over-inferred variational variables δ_i, and thus, loses the function of controlling model complexity. To overcome this problem, we set a fixed configuration $\lambda = 0.01$ for practical use (We tried different values from $\{0.01, 0.1, 1, 10\}$ for λ, but no performance difference was observed. Therefore, we simply fixed λ as 0.01). Besides, as we have discussed previously, a small λ helps give a more reasonable power law distribution to fit degree distributions of real networks. The optimal B is obtained by a gradient ascend method and the gradient is computed with

$$\frac{\partial \mathcal{O}(b_{kk'})}{\partial b_{kk'}} = \frac{\sum_{y_{ij}=1}(1 + \bar{\delta}_i + \bar{\delta}_j)\phi_{ik}\phi_{jk'}}{b_{kk'}}$$
$$- \sum_{y_{ij}=0}(1 + \bar{\delta}_i + \bar{\delta}_j)\phi_{ik}\phi_{jk'} b_{kk'}^{\bar{\delta}_i+\bar{\delta}_j}. \qquad (11)$$

Finally, the optimal π is given by

$$\pi_k \propto \sum_i \phi_{ik}. \qquad (12)$$

4 Simulations

In this section, we evaluate PLD-SBM on two types of simulated networks - one PLD-SBM biased networks, and the other one SBM biased networks to show how PLD-SBM improves SBM.
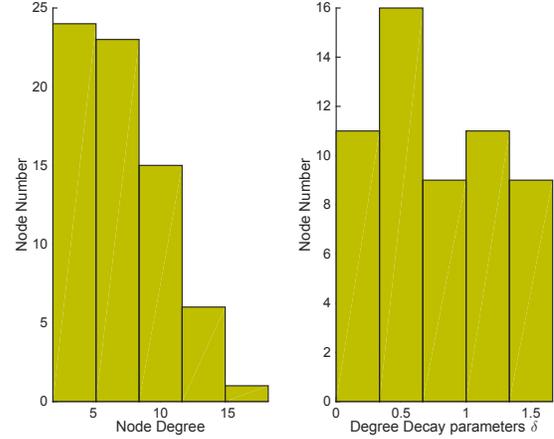
Figure 2: Histograms of Node Degrees and Estimated Degree Decay Parameters $\bar{\delta}_i$ on one Simulated Network.
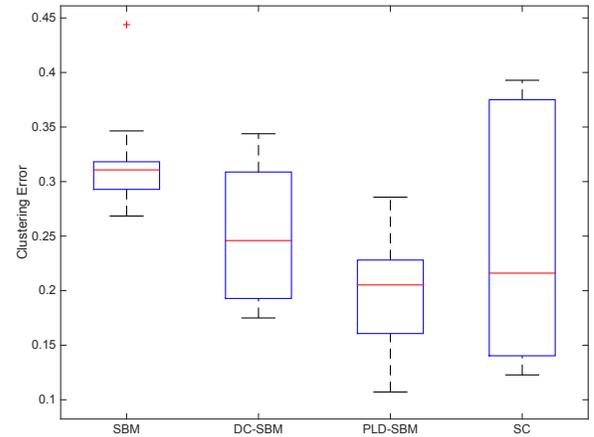
Figure 3: Performance Evaluation on Simulated Networks.

PLD-SBM biased Network Data

We generated 20 networks, each with 3 clusters, and each cluster with around 20 nodes[1]. The edges between nodes are formed in two steps. First, we used the BA model [Barabási and Albert, 1999] to generate intra-cluster edges. Specifically, we started with 5 nodes for each cluster, and iteratively added a new node to the cluster by linking it to 2 old existing nodes. Edge preference, i.e., the probability of linking the new node to an old one, is proportional to the current degree of the latter. Next, we randomly picked 5 node pairs for each different cluster pair to form the inter-cluster edges. Note that the above two-step procedure may differ from that of the direct generation of a scale-free network; however, it is a compromise over the requirement of "true" cluster information for performance evaluation.

We fitted SBM and PLD-SBM to each of the 20 simulated

---

[1]Due to the scale-free network model, we cannot fix node number but only have approximate required node number in the resulting network.

(a) Truth

(b) SBM

(c) PLD-SBM

(d) DC-SBM
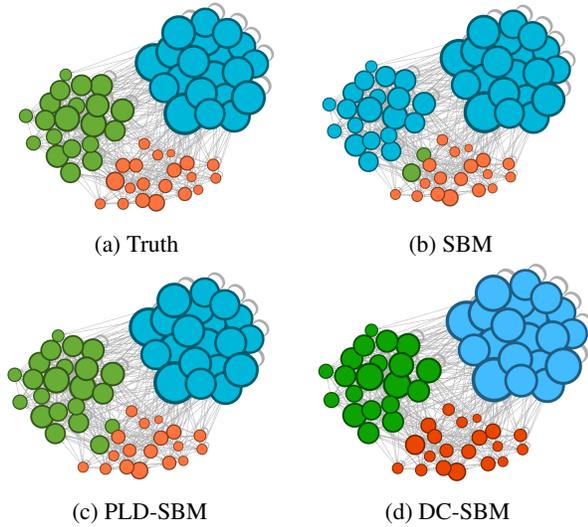
Figure 4: Clustering results on SBM-generated network. Note: Size of nodes indicates

(a) SBM, $K = 6$

(b) PLD-SBM, $K = 6$

(c) SBM, $K = 3$

(d) PLD-SBM, $K = 3$

Figure 5: Prediction on the Adolescent Health Data.

networks. In addition, in order to see if the generated network data is trivially simple for clustering, we also implemented spectral clustering (SC) [Luxburg, 2007], the state-of-the-art similarity based clustering algorithm, for performance comparison. The result of SC in Figure 3 clearly indicates that the clustering task is moderately difficult. PLD-SBM offers a significant improvement over SBM largely due to its ability to correct the bias with the degree decay parameter $\bar{\delta}_i$. Figure 2 shows the histograms of node degree and the estimated degree decay parameter $\bar{\delta}_i$ on one of the simulated networks. The power-law feature of the degree distribution is evident, while the degree decay parameter $\bar{\delta}_i$ varies in a range from 0 to 1.7 to offer automatic adaptation. Though on average, SC seems to outperform SBM, it did not perform as well as PLD-SBM. This might because a well-designed statistical model is better than a non-statistical model, e.g., SC, in dealing with data's probabilistic uncertainties.

SBM biased Network Data

A dataset which is more coherent with SBM is generated. The network contains three clusters with different within link generating probabilities, i.e., 0.3, 0.7, and 0.9. The λ is fixed to 0.1. All inter-cluster link generating probabilities are set to 0.1. Each cluster includes 20 nodes, and all links are generated independently given the cluster labels of the relevant nodes [Holland *et al.*, 1983]. Obviously, the generated clusters exhibit intra-cluster degree homogeneity, namely, the degrees are uniformly distributed amongst the intra-cluster and have no power-law alike distribution.

One snap of the results achieved by different baselines and the proposed PLD-SBM is shown in Figure 4. Even the network is with obvious intra-cluster degree homogeneity which is highly biased to SBM, PLD-SBM still obtains better results than SBM in terms of dividing it into clusters similar to the truth. This is caused by that the two clusters which are separated in the truth but are combined in the SBM output
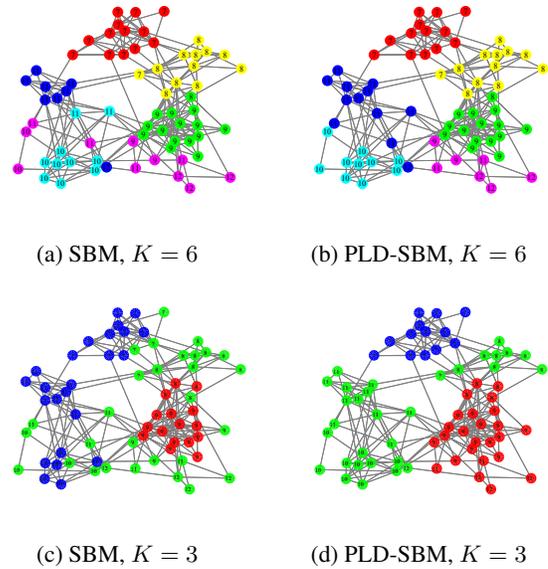
have quite similar degree values, as demonstrated by similar sizes of nodes in the figure. We also draw the result obtained by DC-SBM [Karrer and Newman, 2011]. It is easily seen that DC-SBM and PLD-SBM achieve similar performance in recovering the true cluster structure.

5 Application to Real-world Datasets

Adolescent Health Network [Harris *et al.*, 2003]

Both $K = 6$ and $K = 3$ are studied, which are used to examine clustering performance and the structure differences discovered by the baseline SBM and the proposed PLD-SBM.

Figure 6 presents the degree distribution of the friendship network, which also exhibits a significant power-law feature within each grade. When fitting PLD-SBM to the network, the degree decay parameter $\bar{\delta}_i$ range varied consider-
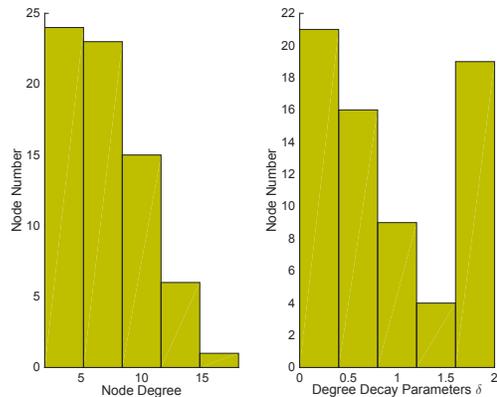
Figure 6: Histograms of Node Degrees and Estimated Degree Decay Parameters $\bar{\delta}_i$ on the Adolescent Health Data.
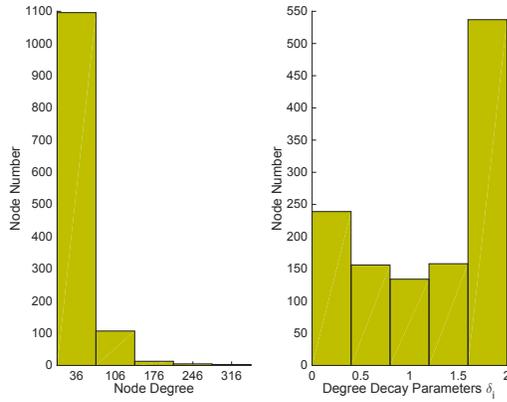
Figure 7: Histograms of Node Degrees and Estimated Degree Decay Parameters $\bar{\delta}_i$ on the political blog network.

ably. The proposed PLD-SBM achieves second best performance in terms of clustering error when comparing to the baseline SBM and two state-of-the-art SBM extensions. , i.e., the mixed membership stochastic model (MMSBM) [Airoldi *et al.*, 2008] and the stochastic block mixture model (SBMM) [Doreian *et al.*, 2007]: SBM has the worst performance with 13 errors; PLD-SBM has the second best performance with 7 errors; two mixture extensions - MMSBM and MSBM - achieve 6 and 10 errors respectively. PLD-SBM was slightly inferior to MMSBM, probably because there are some mixed membership structures in the network, i.e., some students may have best friends from multiple grades. However, the competitive performance between PLD-SBM and SBMM implies that the power-law feature of degree distribution is essential for this network, and raises the question of whether an extension of PLD-SBM to mixed membership situations may work better.

Figure 5 graphs the prediction results of SBM and PLD-SBM using the Fruchterman-Reingold algorithm [Salter-Townshend *et al.*, 2012]. The results with $K = 3$ (the bottom row) help us understand the structure difference discovered by SBM and PLD-SBM. Obviously, PLD-SBM merges small groups of grades into larger ones, which retains the degree skewness within each cluster. By contrast, the traditional SBM, as expected, tends to divide the nodes into groups of high-degree 'hub' nodes such as blue- and red- colored ones and of low-degree 'peripheral' nodes such as green-colored ones.

Political Blog Network [Adamic and Glance, 2005]

Figure 7 presents the overall degree histogram of the political blog network, which obviously exhibits a severely skewed degree distribution. 74% of nodes have less than 36 degrees while 1% have over 176 degrees. When fitting PLD-SBM to this network, the range of learned degree decay parameters $\bar{\delta}_i$, shown in Figure 7, again varied considerably from 0 to 2.

Regarding clustering accuracy, PLD-SBM accomplished a cluster structure with an error rate of 0.0466 while the benchmark SBM obtained a cluster structure with an error rate of 0.4492, which verifies the superiority of PLD-SBM to SBM.

The degree distributions for both political parties are highly skewed. Both the liberal and conservative labels include a few popular blogs with over a hundred links, while hundreds of blogs only have one or two. We therefore anticipate the proposed PLD-SBM will be superior to SBM in detecting cluster structures by addressing the power-law degree distributions. We examined the cases of $K = 2, 3, 4, 5$, and the results are shown in Figure 8, where the cluster structures are visulized by ForceAtlas graph layout algorithm [Jacomy *et al.*, 2014]. It is easily seen that SBM tends to group political blog nodes into high- and low-degree clusters. One typical cluster is the one constituted by low-degree peripheral nodes, as shown by the smallest nodes in the first column. Both the proposed PLD-SBM and DC-SBM aim to add individual degree uniqueness into the traditional SBM, but they behave differently and discover different cluster structures with varying K from 3 to 5. As indicated by the second and third columns in the figure, DC-SBM tends to divide large groups into smaller ones as increasing K, while PLD-SBM tends to retain the two main clusters but splits the peripheral nodes into smaller clusters as increasing K.
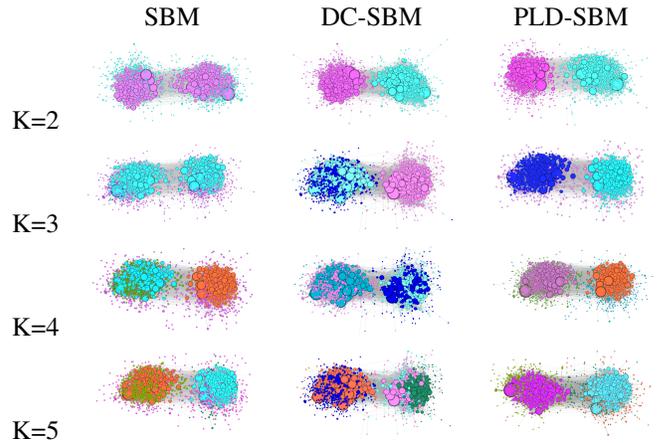
Figure 8: Qualitative clustering results on political blog network with different Ks.

6 Conclusion

In this paper we have proposed a new extension for stochastic block models to address power-law degree distributions. The extended model, PLD-SBM, has a power-law degree characteristic that approximates the scale-free feature of real networks. Such a property allows PLD-SBM to correct the bias of SBM in statistical inference caused by power-law degree distributions. Experiments on simulated networks and a friendship network from the National Adolescent Health data and the political blog network demonstrate the effectiveness of PLD-SBM.

Acknowledgements

References

[Adamic and Glance, 2005] L. A. Adamic and N. Glance. The political blogosphere and the 2004 us election: Divided they blog. In *Proceedings of the 3rd International Workshop on Link Discovery*, pages 36–43. ACM, 2005.

[Ahmed and Xing, 2007] A. Ahmed and E. Xing. On tight approximate inference of the logistic-normal topic admixture model. In *Proceedings of the 11th Tenth International Workshop on Artificial Intelligence and Statistics*, 2007.

[Airoldi *et al.*, 2008] E.M. Airoldi, D.M. Blei, S.E. Fienberg, and E.P. Xing. Mixed membership stochastic blockmodels. *Journal of Machine Learning Research (JMLR)*, 9:1981–2014, 2008.

[Barabási and Albert, 1999] A.L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[Bishop, 2006] C. M. Bishop. *Pattern recognition and machine learning*. springer, 2006.

[Bu *et al.*, 2016] Z. Bu, Z. Wu, J. Cao, and Y. Jiang. Local community mining on distributed and dynamic networks from a multiagent perspective. *IEEE Transactions on Cybernetics*, 46(4):986–999, 2016.

[Doreian *et al.*, 2007] P. Doreian, V. Batagelj, and A. Ferligoj. Discussion of "model-based clustering for social networks". *Journal of the Royal Statistical Society, Series A*, 170:333–334, 2007.

[Fan *et al.*, 2015] X. Fan, L. Cao, and Y. R. Xu. Dynamic infinite mixed-membership stochastic blockmodel. *IEEE Transaction on Neural Networks and Learning Systems (TNNLS)*, 26(9):2072–2085, 2015.

[Goldenberg *et al.*, 2010] A. Goldenberg, A.X. Zheng, S.E. Fienberg, and E.M. Airoldi. A survey of statistical network models. *Foundations and Trends® in Machine Learning*, 2(2):129–233, 2010.

[Harris *et al.*, 2003] K. M. Harris, F. Florey, J. Tabor, P. S. Bearman, J. Jones, and R. J. Udry. The national longitudinal study of adolescent health: Research design. Technical report, 2003.

[Ho *et al.*, 2011] Q. Ho, An. P. Parikh, L. Song, and E. P. Xing. Multiscale community blockmodel for network exploration. *Journal of Machine Learning Research - Proceedings Track*, 15:333–341, 2011.

[Holland *et al.*, 1983] P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

[Jacomy *et al.*, 2014] M. Jacomy, T. Venturini, S. Heymann, and M. Bastian. Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. 2014.

[Karrer and Newman, 2011] B. Karrer and M. E. J. Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1):016107, 2011.

[Kemp *et al.*, 2006] C. Kemp, J.B. Tenenbaum, T.L. Griffiths, T. Yamada, and N. Ueda. Learning systems of concepts with an infinite relational model. In *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI)*, volume 21, page 381, 2006.

[Kim *et al.*, 2012] D.I. Kim, M.C. Hughes, and E.B. Sudderth. The nonparametric metadata dependent relational model. In *Proceedings of the 29th International Conference on Machine Learning (ICML)*, 2012.

[Luo *et al.*, 2016] X. Luo, J. Xuan, J. Lu, and G. Zhang. Measuring the semantic uncertainty of news events for evolution potential estimation. *ACM Transactions on Information Systems*, 34(4):24:1–24:25, June 2016.

[Luxburg, 2007] U. Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, December 2007.

[Mariadassou *et al.*, 2010] M. Mariadassou, S. Robin, and C. Vacher. Uncovering latent structure in valued graphs: a variational approach. *The Annals of Applied Statistics*, pages 715–742, 2010.

[Mørup and Hansen, 2009] M. Mørup and L. K. Hansen. Learning latent structure in complex networks. In *Proceedings of the 23rd NIPS Workshop on Analyzing Networks and Learning with Graphs*, number 2009, 2009.

[Peng *et al.*, 2015] C. Peng, Z. Zhang, K. C. Wong, X. Zhang, and D. E. Keyes. A scalable community detection algorithm for large graphs using stochastic block models. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, 2015.

[Qiao *et al.*, 2015] M. Qiao, R. Xu, W. Bian, and D. Tao. Fast sampling for time-varying determinant point processes. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 2015.

[Salter-Townshend *et al.*, 2012] M. Salter-Townshend, A. White, I. Gollini, and T.B. Murphy. Review of statistical network analysis: Models, algorithms, and software. *Statistical Analysis and Data Mining*, 2012.

[Shiga and Mamitsuka, 2012] M. Shiga and H. Mamitsuka. A variational bayesian framework for clustering with multiple graphs. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 24(4):577–590, 2012.

[Snijders and Nowicki, 1997] T.A.B. Snijders and K. Nowicki. Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification*, 14(1):75–100, 1997.

[Tang *et al.*, 2012] L. Tang, H. Liu, and J. Zhang. Identifying evolving groups in dynamic multimode networks. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 24(1):72–85, 2012.