Directed Network Community Detection: A Popularity and Productivity Link Model

Tianbao Yang[1] Yun Chi[2] Shenghuo Zhu[2] Yihong Gong[2] Rong Jin[1]
[1]Department of Computer Science and Engineering, Michigan State University, MI 48824, USA
[2]NEC Laboratories America, 10080 N. Wolfe Rd, SW3-350, Cupertino, CA 95014, USA
[1]{yangtia1,rongjin}@msu.edu, [2]{ychi,zsh,ygong}@sv.nec-labs.com

Abstract

In this paper, we consider the problem of community detection in directed networks by using probabilistic models. Most existing probabilistic models for community detection are either *symmetric* in which incoming links and outgoing links are treated equally or *conditional* in which only one type (i.e., either incoming or outgoing) of links is modeled. We present a probabilistic model for directed network community detection that aims to model both incoming links and outgoing links *simultaneously* and *differentially*. In particular, we introduce latent variables *node productivity* and *node popularity* to explicitly capture outgoing links and incoming links, respectively. We demonstrate the generality of the proposed framework by showing that both symmetric models and conditional models for community detection can be derived from the proposed framework as special cases, leading to better understanding of the existing models. We derive efficient EM algorithms for computing the maximum likelihood solutions to the proposed models. Extensive empirical studies verify the effectiveness of the new models as well as the insights obtained from the unified framework.

keywords: community detection, popularity, productivity, stochastic block model, directed network

1 Introduction

Community detection is an important topic in analyzing networked data because it reveals underlying structures in a complex network, a key to the network analysis. A community can be intuitively considered as a set of nodes that are densely connected with each other while sparsely connected with other nodes in the network. Based on this intuition, many previous studies (e.g., [7, 13, 15]) focused on defining appropriate metrics to quantify the connection and efficient algorithms to optimize the defined metrics. These approaches usually rely on some heuristics and lack a rigorous mathematical model.

More recently, various probabilistic models have been proposed for community detection. Among them, stochastic block models [10, 14, 17, 3, 16, 2, 12] are probably the most successful ones in terms of capturing meaningful communities, producing good performance, and offering probabilistic interpretations. The basic idea is first to define a generative process where links are generated based on latent community memberships of nodes, and then to infer the community memberships from the links by either maximizing the data likelihood or computing the posterior distribution for community memberships.

Most stochastic block models can be classified into two categories: the *symmetric* approaches [14, 17] that model links by symmetric joint probabilities, and the *conditional* approaches [3, 16] that focus on the conditional probability of receiving links. Neither of these models is satisfying: a symmetric model misses the semantics of link directions, a key factor that distinguishes directed networks from undirected networks; a conditional model only captures one type of links, either incoming links or outgoing links, and therefore is unable to characterize nodes in a full spectrum. As an example, in a blog readership network, there are two types of bloggers: "writers" who generate influential blogs read by many, and "readers" who read a lot but seldom write anything for others to read. Evidently, to characterize these two types of bloggers, it is important to examine both incoming links and outgoing links of the network.

In this work, we propose a novel probabilistic framework for directed network community detection, termed **Popularity and Productivity Link** model or **PPL** for short, that explicitly addresses the shortcomings of the existing stochastic block models. In particular, we model both outgoing links and incoming links by the introduction of the latent variables *productivity* and *popularity*. We demonstrate the generality of the proposed framework by showing that both the symmetric models and the conditional models can be derived from the proposed framework as special cases, leading to the unification of various seemingly different forms for the existing models. We develop efficient EM-algorithms for computing the maximum likelihood solutions to the models proposed in this paper. Extensive empirical studies show the promising performances of the proposed models in several application domains. Further analysis is conducted to investigate the trade-offs of each stochastic block model when data characteristic varies.

The rest of the paper is organized as follows. In the rest of this section, we give a general review of related work. In Section 2, we give background information, including notation we will use and details of several previous approaches. In Section 3, we present the PPL model, several of its variations, and some of their properties. In Section 4, we provide a detailed analysis on the relationship between PPL models and several existing stochastic block models. In Section 5, we describe an efficient estimation algorithm. In Section 6, we show the results of experimental studies. Finally, we conclude in Section 7.

Related Work Link-based approaches for community detection can be roughly put into two categories. The first category is metric-based. For approaches of this category, a metric is first defined to quantify the quality of any potential community structure; and then, procedures are developed to optimize the proposed metric. Some well-known metrics include normalized cut proposed by Shi et al. [15], modularity proposed by Newman et al. [13], betweenness proposed by Gregory [7], etc. Furthermore, some later work introduces simple probabilistic interpretations to some of these metrics and extends the metrics from measuring undirected networks to measuring directed networks [11, 18]. A main weak point among these metric-based approaches is they usually do not have a rigorous generative model. Therefore, on one hand, it is difficult to reach a consensus on the metric; and on the other hand, these approaches lack the generative power to generate new (or unobserved) links, which is necessary in many applications (e.g., link prediction).

The second category of approaches for community detection are based on probabilistic models. For approaches in this category, a generative process is first defined in which links are generated based on latent community memberships, and then, inference algorithms are used to infer the latent community memberships from data. One of the most well-studied probabilistic models is the stochastic block model [10], for which many variations have been recently proposed and analyzed. For example, Cohen et al. [3] extend the HITS algorithm to link analysis and proposed PHITS. Yang et al. [16] propose conditional stochastic block model to fit indegree distribution in a network. Airoldi et al. [2] propose a mixed membership stochastic block model which is later extended by Nallapati et al. [12] to handle directed networks. More recently, Dietz et al. [5], Erosheva et al. [6], and Hofman et al. [8] propose fully Bayesian versions of the stochastic block models by using appropriate priors. The strong points of stochastic block models include rigorous generative processes, well-studied algorithms for efficient inference, and the potentials of fully Bayesian treatment. We will provide the details for several of these approaches in the next section.

2 Background

In this section we first establish some necessary notation for ease of presentation. We then describe the details about three representative existing stochastic block models which are most relevant to our work.

2.1 Notation For a directed network, we denote the nodes by $\mathcal{V} = \{1, \cdots, N\}$, the directed links by $\mathcal{E} = \{(i,j)|s_{ij} \neq 0\}$, where s_{ij} records the value associated with link from node i to node j. s_{ij} can either be binary, to denote whether there is a link from node i to node j, or be non-negative values, to denote the weight of the link. For simplicity, following [16], we assume the "link-in" space (i.e., all possible nodes that can point to a particular node) and "link-out" space(i.e., all possible nodes that can be pointed to by a particular node) of every node to be $\mathcal{V}$, i.e., the complete set of nodes. We use $\mathcal{I}(i) = \{j|s_{ji} \neq 0\}$ to denote the set of all nodes point to node i, and $\mathcal{O}(i) = \{j|s_{ij} \neq 0\}$ to denote the set of all nodes that are pointed to by node i. Let K denote the number of communities, $z_i \in \{1, \cdots, K\}$ denote the community variable of node i, and $\boldsymbol{\gamma}_i = (\gamma_{i1}, \cdots, \gamma_{iK})$ denote the community memberships of node i. In other words, γ_{ik} is the probability for the case $z_i = k$, i.e., node i belongs to community k.

2.2 Existing Models We now review three variants of the well-known stochastic block model [10] that are closely related to the proposed model.

2.2.1 PHITS Model PHITS [3] is a conditional model that focuses on the conditional link probability of $\Pr(j|i)$, i.e., given that node i produces a link, how likely this link will point to node j among all nodes. To compute $\Pr(j|i)$, a community variable z_i is first sampled from a multinomial distribution with parameter $\boldsymbol{\gamma}_i$ that describes the community membership of node i, then for a given z_i, the conditional link probability $\Pr(j|i, z_i)$ is given by $\Pr(j|i, z_i) = \beta_{jz_i}$, where the parameter β_{jk} represents the likelihood for node j to be pointed to by any node in community k. By integrating out z_i, we get the PHITS model

$$\Pr(j|i) = \sum_k \beta_{jk}\gamma_{ik} \tag{2.1}$$

It is well known that PHITS can be considered as an application of PLSA [9] to network data.

2.2.2 Popularity Conditional Link (PCL) Model

PCL [16] is also a conditional model. It models $\Pr(j|i)$ by introducing the latent variable *popularity* for each node that describes how likely a node is to receive a link. Thus, given z_i, i.e., the community assignment for node i, $\Pr(j|i, z_i)$ is given by

$$\Pr(j|i, z_i) = \frac{\gamma_{jz_i} b_j}{\sum_{j'} \gamma_{j'z_i} b_{j'}}$$

where b_j is the popularity of node j. By integrating out z_i we have

$$\Pr(j|i) = \sum_k \frac{\gamma_{jk} b_j}{\sum_{j'} \gamma_{j'k} b_{j'}} \gamma_{ik} \tag{2.2}$$

As shown in [16], PHITS model can be viewed as a relaxed version of PCL model if we introduce the community-dependent popularity. It is important to note that both conditional models only take into account one type of links, and therefore are insufficient for directed network community detection.

2.2.3 Symmetric Joint Link (SJL) Model

SJL [3, 17, 14] is a symmetric model for community detection. It models the link structure by the joint probability $\Pr(i, j)$, i.e., the likelihood of creating a link between node i and j, as follows

$$\begin{aligned}\Pr(i, j) &= \sum_k \Pr(j|k)\Pr(i|k)\Pr(k) \\ &= \sum_k \beta_{jk}\beta_{ik}\pi_k \end{aligned} \tag{2.3}$$

In Equation (2.3), π_k is the prior probability for a link to be produced in community k, and β_{ik} and β_{jk} are the conditional probabilities that nodes i and j are selected as the two ends of the link. Given the symmetric treatment, i.e., $\Pr(i, j) = \Pr(j, i)$, it is evident that SJL may not be suitable for *directed* network community detection.

3 Popularity and Productivity Link (PPL) Model

In this section, we first present our *popularity and productivity link* (PPL) model in its general form, then give three variations of the general PPL model, and finally discuss several properties of the PPL models.

3.1 General Form of PPL

PPL models the joint link probability $\Pr(i, j)$, i.e., how likely there is a directed link from node i to node j. In order to emphasize the different roles played by i and j, we write $\Pr(i, j)$ as $\Pr(i_\rightarrow, j_\leftarrow)$, denoting that node i plays the role of producing the link, and node j plays the role of receiving the link. Following the idea of SJL, we model $\Pr(i_\rightarrow, j_\leftarrow)$ as follows

$$\begin{aligned}\Pr(i_\rightarrow, j_\leftarrow) &= \sum_k \Pr(i_\rightarrow|k)\Pr(j_\leftarrow|k)\Pr(k) \\ &= \sum_k \left(\frac{\gamma_{ik} a_i}{\sum_{i'} \gamma_{i'k} a_{i'}} \frac{\gamma_{jk} b_j}{\sum_{i'} \gamma_{i'k} b_{i'}} \sum_{i'} \gamma_{i'k} c_{i'} \right)\end{aligned} \tag{3.4}$$

where

- γ_{ik}: the probability for node i to belong to community k
- a_i: the *productivity* of node i, i.e., among all the nodes, how likely a link is produced by node i
- b_j: the *popularity* of node j, i.e., among all the nodes, how likely a link is received by node j
- c_i: the weight of node i in terms of deciding the community prior $\Pr(k)$ (which will be elaborated momentarily).

To handle scale invariance, we normalize so that $\sum_i a_i = \sum_j b_j = \sum_i c_i = 1$.

Generative Process We explain Equation (3.4) by the following generative process of PPL:

- Sample a community z according to a prior distribution $\pi_1, \cdots, \pi_K$, where π_k is computed by $\pi_k = \sum_{i=1}^N \gamma_{ik} c_i$.
- Given community z, the conditional link probability is given by

$$\begin{aligned}\Pr(i_\rightarrow, j_\leftarrow|z) &= \Pr(i_\rightarrow|z)\Pr(j_\leftarrow|z) \\ &= \frac{\gamma_{iz} a_i}{\sum_{i'} \gamma_{i'z} a_{i'}} \frac{\gamma_{jz} b_j}{\sum_{i'} \gamma_{i'z} b_{i'}}\end{aligned} \tag{3.5}$$

There are two unique features in the above generative process:

- Prior probability $\pi_k = \sum_i \gamma_{ik} c_i$ is constructed as the weighted sum of node memberships γ_{ik}, where c_i is used to weight node i in the combination. This construction enforces the consistency between node memberships γ_{ik} and community prior $\{\pi_k\}_{k=1}^K$. This specific construction of community priors also simplify relation between the proposed framework and some existing models for community detection.
- In Equation (3.5), the two ends of link $i \rightarrow j$ are treated differently when modeling $\Pr(i_\rightarrow, j_\leftarrow|z)$: besides the dependence on community memberships γ_{ik} and γ_{jk}, $\Pr(i_\rightarrow|z)$ and $\Pr(j_\leftarrow|z)$ are modeled by a_i (i.e., the productivity of node i) and b_j

(i.e., the popularity of node j), respectively, leading to the differentiation of the roles played by the two nodes.

With the joint link probability defined in Equation (3.4), the log-likelihood for links can be written as

$$\mathcal{L}(a,b,c,\gamma) = \tag{3.6}$$
$$\sum_{(i,j)\in\mathcal{E}} s_{ij} \log \sum_{k} \frac{\gamma_{ik}a_i}{\sum_{i'}\gamma_{i'k}a_{i'}} \frac{\gamma_{jk}b_j}{\sum_{i'}\gamma_{i'k}b_{i'}} \sum_{i'}\gamma_{i'k}c_{i'}$$

Note that we use original data s_{ij} in the joint link model rather than normalized data $\hat{s}_{ij} = \frac{s_{ij}}{\sum_j s_{ij}}$ used in conditional link models [3, 16] . Parameters γ, a, b, and c can be inferred by maximizing the log-likelihood $\mathcal{L}(a,b,c,\gamma)$.

3.2 Three Variants of the General PPL Model

In this subsection, we show three variants of PPL model by introducing different restrictions on parameters a, b and c.

Popularity Link (PoL) Model In the first restricted variation, we enforce $c_i = a_i, \forall i$ in Equation (3.4), leading to the following expression for $\Pr(i_{\rightarrow}, j_{\leftarrow})$

$$\Pr(i_{\rightarrow}, j_{\leftarrow}) = \sum_k \frac{\gamma_{jk}b_j}{\sum_{i'}\gamma_{i'k}b_{i'}}\gamma_{ik}a_i \tag{3.7}$$

We refer to this variant as the Popularity Link (PoL) Model. By assuming $c_i = a_i$, we essentially assume that the prior probability of each community (i.e., $\sum_i \gamma_{ik}c_i$) is identical to the prior probability for a link to be produced from that community (i.e., $\sum_i \gamma_{ik}a_i$).

Productivity Link (PrL) Model In the second restricted variation, we enforce c_i to be equal to b_i, leading to the following expression for $\Pr(i_{\rightarrow}, j_{\leftarrow})$,

$$\Pr(i_{\rightarrow}, j_{\leftarrow}) = \sum_k \frac{\gamma_{ik}a_i}{\sum_{i'}\gamma_{i'k}a_{i'}}\gamma_{jk}b_j \tag{3.8}$$

We refer to this variant as the Productivity Link (PrL) Model. By assuming $c_i = b_i$, we essentially assume that the prior probability of each community (i.e., $\sum_i \gamma_{ik}c_i$) is identical to the prior probability for a link to be received by that community (i.e., $\sum_i \gamma_{ik}b_i$).

Regularized PPL (PPL-D) Model In this variation, instead of enforcing the relationship between c_i and a_i or b_i, we learn c_i from data, under certain regularization. In particular, we introduce a Dirichlet prior for parameters $c = (c_1, \ldots, c_N)$, i.e., $\Pr(c) \propto \prod_i c_i^{\alpha}$, where α is the hyper-parameter of Dirichlet distribution. Using the prior $\Pr(c)$ as the regularization, we obtain an MAP estimation of parameters by maximizing the following log-posterior probability

$$\mathcal{L}(a,b,c,\gamma) + \log\Pr(c) \tag{3.9}$$

where $\mathcal{L}(a,b,c,\gamma)$ is given in Equation (3.6). We call this PPL model regularized by the Dirichlet prior the PPL-D model.

3.3 Properties of PPL Models

In this subsection, we show two important properties of the PoL, PrL, and general PPL model.

Equivalence between the Variants of PPL Models The first property is about the relationship between PoL model, PrL model, and general PPL model. Surprisingly, although their formulas are different, the optimal solutions for the three models actually result in identical joint link probability and therefore identical data likelihood. This property is described in the following theorem.

THEOREM 3.1. *Under the optimal solution, the joint link probability* $\Pr(i_{\rightarrow}, j_{\leftarrow})$ *of PoL model, PrL model and general PPL model are the same. That is,* $\Pr^1(i_{\rightarrow}, j_{\leftarrow}|a^1, b^1, \gamma^1) = \Pr^2(i_{\rightarrow}, j_{\leftarrow}|a^2, b^2, \gamma^2) = \Pr^3(i_{\rightarrow}, j_{\leftarrow}|a^3, b^3, c^3, \gamma^3)$, *where* $\Pr^1(i_{\rightarrow}, j_{\leftarrow})$, $\Pr^2(i_{\rightarrow}, j_{\leftarrow})$, $\Pr^3(i_{\rightarrow}, j_{\leftarrow})$ *are the joint link probabilities of PoL model, PrL model, and general PPL model, respectively;* $\{a^1, b^1, \gamma^1\}$, $\{a^2, b^2, \gamma^2\}$ *and* $\{a^3, b^3, c^3, \gamma^3\}$ *are the optimal solutions to maximizing the log-likelihood of PoL model, PrL model and general PPL model, respectively. In particular, denoting the log-likelihood of PoL, PrL and general PPL model by* $\mathcal{L}_1(a,b,\gamma), \mathcal{L}_2(a,b,\gamma), \mathcal{L}_3(a,b,c,\gamma)$ *respectively, we have* $\mathcal{L}_1(a^1,b^1,\gamma^1) = \mathcal{L}_2(a^2,b^2,\gamma^2) = \mathcal{L}_3(a^3,b^3,c^3,\gamma^3)$.

The proof for Theorem 3.1 is given in the appendix. One implication of this theorem is that the space of the optimal solution to the general PPL model is not a unique fixed point. As a consequence, if in addition to the joint link probability, we also care about the exact solution to the community membership γ, then we should not directly apply PPL in its general form. Instead, we should either choose PoL and PrL if the MLE solution is needed, or choose PPL-D if the MAP solution is needed.

Perfect Fitting of the Distributions of Indegree and Outdegree The second property of the PPL model is about degree fitting. It turns out that the optimal solutions to PoL model, PrL model, and general

PPL model all fit exactly the degree distributions (both indegree and outdegree) in the network data. This is described in the following theorem, whose proof is given in the appendix.

THEOREM 3.2. *The model outdegree distribution* $\Pr(i_\rightarrow)$ *and model indegree distribution* $\Pr(j_\leftarrow)$ *of PoL model, PrL model and general PPL model fit exactly the actual indegree and outdegree distributions of the network data.*

This property of degree fitting is a consequence of the concepts of *productivity* and *popularity*. We argue that degree fitting is a very important property for a generative model. This is because in real world, most networks have heavy-tailed (or power-law) degree distribution. So far, no existing stochastic block models can guarantee to generate degree distributions fitting both indegree and outdegree distributions of real-world networks.

4 Relationship with Existing Models

In this section, we describe the relationship between PPL and several existing models, including conditional link models, namely PCL [16] and PHITS [3], and symmetric joint link model (SJL) [14, 17]. It turns out that these existing models all can be considered as special cases of PPL, with different constraints. Such a connection demonstrates that PPL provides a consistent framework to unify the existing models.

4.1 Relationship with Conditional Link Models

We show that the Popularity Conditional Link(PCL) model in [16] is a *conditional* version of the Popularity Link(PoL) model described in Section 3.2. Starting from the joint probability given in Equation (3.7), we can express the conditional probability of the PoL model as

$$\Pr(j_\leftarrow|i_\rightarrow) = \frac{\Pr(i_\rightarrow, j_\leftarrow)}{\Pr(i_\rightarrow)} = \sum_k \frac{\gamma_{jk} b_j}{\sum_{i'} \gamma_{i'k} b_{i'}} \gamma_{ik} \tag{4.10}$$

Note that in the above derivation, we use the fact that $\Pr(i_\rightarrow) = a_i$, which is obtained in the proof of Theorem 3.2. Equation (4.10) is exactly the same as the Popularity Conditional Link(PCL) model proposed in [16]. Because of this connection, in the following discussion, we also refer to the PCL model described in Equation (4.10) (and in [16]) as PoCL model.

Following a similar idea, from Productivity Link(PrL) model we can derive a Productivity Conditional Link model by computing the conditional probability $\Pr(i_\rightarrow|j_\leftarrow)$ from Equation (3.8) as the following

$$\Pr(i_\rightarrow|j_\leftarrow) = \frac{\Pr(i_\rightarrow, j_\leftarrow)}{\Pr(j_\leftarrow)} = \sum_k \frac{\gamma_{ik} a_i}{\sum_{i'} \gamma_{i'k} a_{i'}} \gamma_{jk} \tag{4.11}$$

In the above derivation, we use the fact that $\Pr(j_\leftarrow) = b_j$, which is obtained in the proof of Theorem 3.2. Because of its connection to PrL, we refer to this new conditional model as PrCL.

As we can see, PoCL and PrCL capture the conditional link probability in different directions. PoCL depends on the *popularity* of the receiving node j while PrCL depends on the *productivity* of the producing node i. In addition, both PoCL and PrCL can be naturally derived from the PPL models, i.e., PoL and PrL.

Because as we have discussed before, PHITS is a relaxed version of PoCL, obviously it can also be derived from PPL.

4.2 Relationship with Symmetric Joint Link Model To show its relationship with SJL, we enforce that $c_i = a_i = b_i, \forall i$ in the general PPL model. From a probabilistic view point this restricts that for each node, the probability for producing links is equal to that for receiving links. With this restriction, Equation (3.4) is reduced to

$$\Pr(i_\rightarrow, j_\leftarrow) = \sum_k \frac{\gamma_{ik} c_i}{\sum_{i'} \gamma_{i'k} c_{i'}} \frac{\gamma_{jk} c_j}{\sum_{j'} \gamma_{j'k} c_{j'}} \sum_{i'} \gamma_{i'k} c_{i'}$$

The following theorem, whose proof is given in the appendix, shows that this restricted version of PPL is exactly the SJL model.

THEOREM 4.1. *Under the constraint that* $a_i = b_i = c_i, \forall i$, *the general PPL model is equivalent to the SJL model.*

The relationship revealed by Theorem 4.1 shows that SJL is a special PPL with the constraint that nodes having the same probability in terms of producing and receiving links, which is appropriate only for modeling *undirected* networks.

4.3 Putting It All Together In Table 1, we summarize all the models discussed in this paper. Models that are newly developed in this paper are print in bold. We believe such a unified picture, offered through the PPL model, will be very helpful for understanding and further studying different stochastic block models for community detection.

5 Estimation Algorithm

In this section, we present efficient EM algorithms for computing the MLE solutions to PoL and PrL and the MAP solution to PPL-D. Because the derivation of the algorithms is rather lengthy, here we only present the final form of the algorithms as well as offer several observations, and we provide the detailed derivation in the appendix.

Table 1: Taxonomy of the models discussed in this paper, categorized by (1) if a conditional, joint, or symmetric probability is modeled and (2) if popularity, productivity, or both are considered. Names of the models newly introduced in this paper are in bold.

	popularity	productivity	both
conditional	PHITS, PoCL	**PrCL**	
joint	**PoL**	**PrL**	**PPL,** **PPL-D**
symmetric			SJL

Theorem 5.1. *The following EM algorithms converge to the MLE solutions to PoL and PrL, and the MAP solution to PPL-D.*
E-step:

$$q_{ijk} \propto \Pr{}^{t-1}(i,j,k)$$

where t-1 indicates the result in the previous iteration
M-step:

$$PoL: \gamma_{ik} = \frac{n_{ik}}{m^{\tau}_k b_i + n_{out}(i)},$$

$$b_i = \frac{n_{in}(i)}{\sum_k m^{\tau}_k \gamma_{ik}}, \quad a_i = \frac{n_{out}(i)}{\sum_i n_{out}(i)}$$

$$PrL: \gamma_{ik} = \frac{n_{ik}}{m^{\eta}_k a_i + n_{in}(i)},$$

$$a_i = \frac{n_{out}(i)}{\sum_k m^{\eta}_k \gamma_{ik}}, \quad b_i = \frac{n_{in}(i)}{\sum_i n_{in}(i)}$$

$$PPL\text{-}D: \gamma_{ik} = \frac{n_{ik} + m^{\zeta}_{ik}}{m^{\eta}_k a_i + m^{\tau}_k b_i + m^{\zeta}_i}, c_i = \frac{m^{\zeta}_i + e\alpha}{\sum_i (m^{\zeta}_i + e\alpha)}$$

$$a_i = \frac{n_{out}(i)}{\sum_k m^{\eta}_k \gamma_{ik}}, \quad b_i = \frac{n_{in}(i)}{\sum_k m^{\tau}_k \gamma_{ik}}$$

where e is the summation of all s_{ij} and the rest variables are defined as:

$$\eta_k = \sum_{i'} \gamma^{t-1}_{i'k} a^{t-1}_{i'}, \tau_k = \sum_{j'} \gamma^{t-1}_{j'k} b^{t-1}_{j'}, \zeta_{ik} = \frac{\gamma^{t-1}_{ik} c^{t-1}_i}{\sum_{i'} \gamma^{t-1}_{i'k} c^{t-1}_{i'}}$$

$$n_{in}(i,k) = \sum_{j \in \mathcal{I}(i)} s_{ji} q_{jik}, \quad n_{out}(i,k) = \sum_{j \in \mathcal{O}(i)} s_{ij} q_{ijk}$$

$$n_{in}(i) = \sum_k n_{in}(i,k), \quad n_{out}(i) = \sum_k n_{out}(i,k)$$

$$n_{ik} = n_{in}(i,k) + n_{out}(i,k), \quad m_k = \sum_{(i \to j) \in \mathcal{E}} s_{ij} q_{ijk}$$

$$m^{\tau}_k = \frac{\sum_{(i \to j) \in \mathcal{E}} s_{ij} q_{ijk}}{\tau_k}, \quad m^{\eta}_k = \frac{\sum_{(i \to j) \in \mathcal{E}} s_{ij} q_{ijk}}{\eta_k}$$

$$m^{\zeta}_{ik} = \zeta_{ik} \sum_{(i \to j) \in \mathcal{E}} s_{ij} q_{ijk}, \quad m^{\zeta}_i = \sum_k m^{\zeta}_{ik}.$$

It can be observed from the EM algorithm that in every iteration (and therefore in the final solutions) for each node i, its productivity a_i is proportional to its outdegree and its popularity b_i is proportional to its indegree. This is consistent with our intentions that the productivity of a node reflects how likely it produces a link and the popularity of a node reflects how likely it receives a link.

In addition, it is worth mentioning that in the real implementation, we avoid to explicitly compute all q_{ijk}'s (whose number is N^2K, which can be extremely large). Instead, q_{ijk}'s are computed in an "on-demand" fashion. We can show that the complexity (per iteration) of our EM algorithms is linear in the number of links in the network. Therefore, the algorithm is very efficient because in most real applications, networks are sparse and so the number of links is usually manageable.

6 Experiments

In this section, we show experiment results. We evaluate a variety of models (variations of PPL and existing models) on two tasks: community detection and link prediction. In addition, we also investigate the issue of degree fitting. We start by describing the data sets used in the experiments.

6.1 Data Sets In the following experiments, we use a blog network and two paper citation networks.

Political Blog Network This is a directed network of hyperlinks between a set of weblogs about US politics, recorded by Adamic and Glance [1]. In this network, there are totally 1,490 nodes and 19,090 links. Each node is labeled as either conservative or liberal.

Paper Citation Networks We use the Cora paper citation network and the Citeseer paper citation network processed by Getoor et al.[1]. There are totally 2,708 nodes and 5,429 links in Cora network, and 3,312 nodes and 4,732 links in Citeseer network. Each paper in Cora network is categorized into one of 7 classes (e.g., Genetic Algorithms, Neural Networks, etc.), and each paper in Citeseer network is labeled as one of 6 classes.

6.2 Community Detection In the first task, communities are to be detected from the networks. In this task, the real class labels in the data sets are used as the ground truth to evaluate the communities detected by different models. More specifically, we use the following evaluation metrics.

[1] http://www.cs.umd.edu/projects/linqs/projects/lbc/

Evaluation Metrics for Community Detection We use three commonly used metrics for evaluating the performance of community detection, i.e. normalized mutual information (NMI), pairwise F measure (PWF), and modularity (Modu). We first give detailed description about the three metrics.

Normalized mutual information (NMI) is defined as follows: given the true community structure $C = \{C_1, \cdots, C_K\}$, where C_k denote the set of nodes in the k-th community, and the community structure $C' = \{C'_1, \cdots, C'_K\}$ obtained from a model, the mutual information is computed as

$$MI(C, C') = \sum_{C_k, C'_l} p(C_k, C'_l) \log \frac{p(C_k, C'_l)}{p(C_k)p(C'_l)}$$

where $p(C_k)$ denotes the probability that a randomly selected node belongs to C_k, and $p(C_k, C'_l)$ denotes the joint probability that a randomly selected node belongs to C_k and C'_l. The normalized mutual information is defined as

$$NMI(C, C') = \frac{MI(C, C')}{\max(H(C), H(C'))}$$

where $H(C) = \sum_k p(C_k) \log \frac{1}{p(C_k)}$ is the entropy of partition C.

Pairwise F measure (PWF) is another commonly used measure for evaluating clustering algorithms. Assume T is the set of node pairs (i, j) where nodes i and j belong to the same community in the ground truth, and S is the set of node pairs that belong to the same community in the outcome of a specific model. Then the pairwise F measure is computed from pairwise precision and recall as

$$precision = |S \bigcap T|/|S| \quad recall = |S \bigcap T|/|T|$$

$$PWF = \frac{2 \times precision \times recall}{precision + recall}$$

where $|\cdot|$ indicates the cardinality of a set.

Note that to compute the normalized mutual information and pairwise F measure, the ground truth must be used. However, in some cases, the ground truth does not necessarily faithfully reflect the link structure. Therefore, we also use another measure called directed modularity (Modu), which is proposed by Leicht et al. [11] for measuring community partitions in directed networks without using ground truth. The definition of the directed modularity is given by

$$Modu = \frac{1}{e} \sum_{ij} \left(s_{ij} - \frac{d_{out}(i) d_{in}(j)}{e} \right) \delta(c_i, c_j)$$

where $d_{in}(i)$ and $d_{out}(i)$ are the indegree and outdegree of node i in the network, e is the number of directed links in the network, and c_i denotes the community of node i assigned by a model, and $\delta(\cdot, \cdot)$ is the Kronecker delta function.

For all the three metrics, i.e., NMI, PWF, and Modu, larger values correspond to better performances.

Performance on Community Detection The community detection performances for different models on the three data sets are given in Tables 2, 3, and 4. Among the models, PHITS, PoCL, and PrCL are conditional link models. PHITS [4] represents the model described in Equation (2.1); PoCL represents the Popularity Conditional Link model [16] described in Equation (2.2); PrCL represents the Productivity Conditional Link model described in Equation (4.11). SJL represents symmetric link model as described in [14]. PoL, PrL, and PPL-D are the joint link models proposed in this work.

All the EM algorithms for MLE and MAP are run with 100 iterations, which according to our observation is more than enough for convergence. In order to alleviate the problem of local minimum of EM algorithms, for each test we conduct 10 trials with different random initializations, and choose the one giving the largest likelihood. The prior α for the parameter c in PPL-D is set to 1. Actually, we found the performance not sensitive to α—we tested different values for α ranging from 0.01 to 1, and the results are almost the same.

From the performance results, we can make the following comparisons and observations.

Joint link models vs. conditional link models Joint link models clearly outperform conditional link models. These can be seen from that our joint link models, i.e., PoL, PrL, and PPL-D always have the top performances and clearly outperform their conditional counterparts PoCL and PrCL. Even the symmetric joint link model SJL outperforms its conditional counterpart PHITS in most of the cases. This result verifies our assumption that modeling both behavior in receiving links (popularity) and that in producing links (productivity) is better than modeling just one behavior or none at all.

Non-symmetric vs. symmetric joint link models Comparing the performances of non-symmetric link models, i.e., PoL, PrL, and PPL-D, with that of traditional SJL model, which is symmetric, we can see that the non-symmetric models consistently outperform SJL and the improvement is quite significant in many cases. This verifies the benefit of separating the behavior of nodes in receiving links and that in producing links over simply ignoring the direction of links.

PPL models without vs. with restrictions Comparing PPL-D, which does not restrict c other than providing a weak prior, with PoL, PrL and SJL, which enforce $c = a$, $c = b$, and $c = a = b$, respectively, we can see that PPL-D has the best performance. However, as shown in Section 3.3 we can always derived PoL and PrL from PPL-D that give the identical data likelihood, and so the above result suggests that PPL-D tends to find better solutions for community memberships.

Popularity vs. productivity If we can only choose one feature between popularity and productivity for community detection in our data sets, it seems that popularity has a small edge over productivity. This can be observed both in joint link models (i.e., PoL over PrL) and conditional models (i.e., PoCL over PrCL). Such a result suggests that to determine the community membership of a node i in these three data sets, those nodes point to i may be more important than those nodes pointed to by i.

Table 2: Community detection performance on the Political Blog data set, where the best performances are in bold.

Algo.	NMI	PWF	Modu
PHITS	0.3829	0.7152	0.4200
PoCL	0.4905	0.7947	0.4270
PrCL	0.4569	0.7776	0.4243
SJL	0.4409	0.7425	0.4323
PoL	0.5156	0.8072	**0.4324**
PrL	0.5178	0.8091	**0.4324**
PPL-D	**0.5365**	**0.8167**	**0.4324**

Table 3: Community detection performance on the Cora data set, where the best performances are in bold.

Algo.	NMI	PWF	Modu
PHITS	0.0591	0.1862	0.3594
PoCL	0.0797	0.1982	0.5982
PrCL	0.0211	0.1666	0.4959
SJL	0.0602	0.1840	0.6091
PoL	0.0886	0.2014	0.6310
PrL	0.0870	0.1993	0.6307
PPL-D	**0.0972**	**0.2085**	**0.6381**

Table 4: Community detection performance on the Citeseer data set, where the best performances are in bold.

Algo.	NMI	PWF	Modu
PHITS	0.0117	0.1788	0.4374
PoCL	0.0292	0.1909	0.6214
PrCL	0.0131	0.1805	0.5954
SJL	0.0236	0.1896	0.6348
PoL	0.0292	0.1921	0.6648
PrL	0.0263	0.1904	0.6612
PPL-D	**0.0317**	**0.1948**	**0.6687**

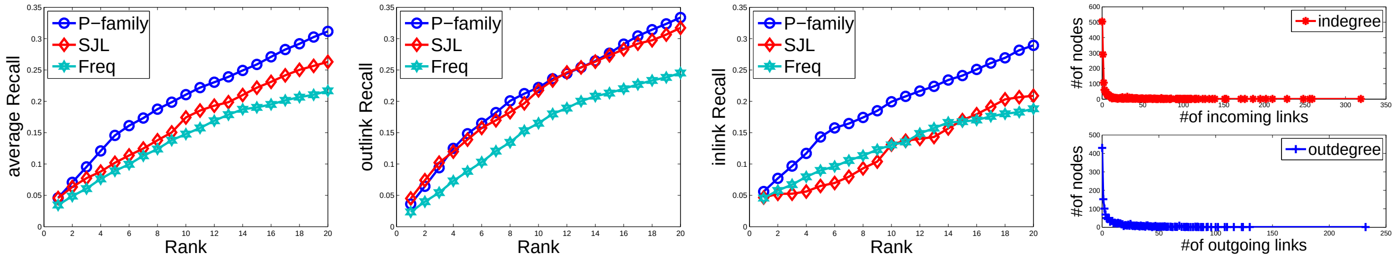
6.3 Link Prediction In this task, we study the performance of the *joint link* models on predicting the links (both incoming links and outgoing links). Specifically, for each node in the network we randomly hide one of its incoming links and one of its outgoing links and ask each model to recover the missing links. Such a task has practical values in applications such as friend recommendation in social networks and citation suggestion in citation networks.

Evaluation Metric for Link Prediction We measure the performance of link prediction by Recall measure. Two types of recall are presented, namely *outlink recall* and *inlink recall*. The outlink recall measures the ability of a model to predict nodes pointed to by a given node. The inlink recall measures the ability of a model to predict the nodes point to a given node. To compute outlink recall for node i, we first compute the outlink probabilities $\Pr(j_{\leftarrow}|i_{\rightarrow})$ for node i to all other nodes by $\Pr(j_{\leftarrow}|i_{\rightarrow}) = \dfrac{\Pr(i_{\rightarrow}, j_{\leftarrow})}{\sum_j \Pr(i_{\rightarrow}, j_{\leftarrow})}$. The resulting probabilities assign an outlink rank to each node j. The outlink recall at rank position K is defined as the fraction of nodes whose top-K ranked predictions contain the true missing link. Inlink recall is defined similarly based on $\Pr(j_{\rightarrow}|i_{\leftarrow})$. In addition, we also report the average of the inlink and outlink recalls.

Performance on Link Prediction The recalls at top-1 through top-20 on the three data sets are given in Figures 1, 2, and 3. All the results are averaged over 10 trials with different randomly selected missing links. Because we have that PoL, PrL and general PPL model have equal link probabilities, and because we also found that PPL-D achieves almost the same performance as PoL and PrL, we will only report one result for these models which are denoted by P-family models. We also report the results of a naive baseline, the Frequency-based model, where the outgoing link probabilities are proportional to the indegree of nodes, i.e., $\Pr(j_{\leftarrow}|i_{\rightarrow}) \propto d_{in}(j)$, and the incoming link probabilities are proportional to the outdegree of nodes, i.e., $\Pr(j_{\rightarrow}|i_{\leftarrow}) \propto d_{out}(j)$.

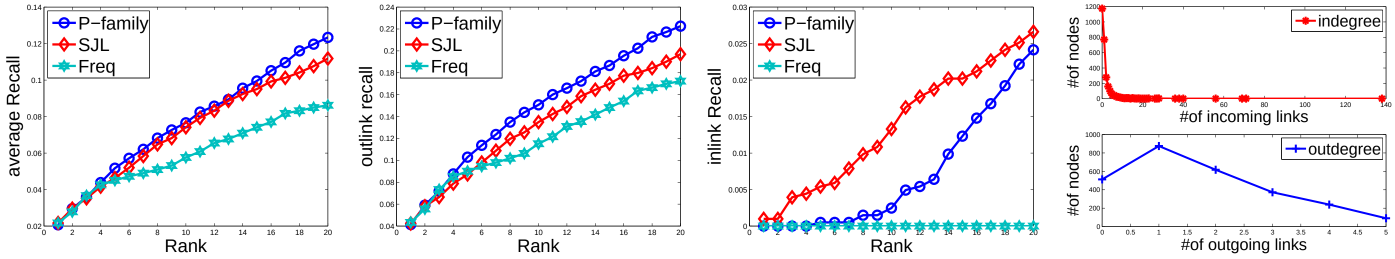
As can be seen from the figures, compared to SJL and the Frequency-based baseline, P-family models perform the best in all the cases except the *inlink* recall for Cora network. This result illustrates that most of the time, it is beneficial to use productivity and popularity to model indegree and outdegree distributions separately in a directed network.

However, the inlink recall for Cora network is an abnormal case, where SJL performs the best, P-family models perform worse, and the Frequency-based model



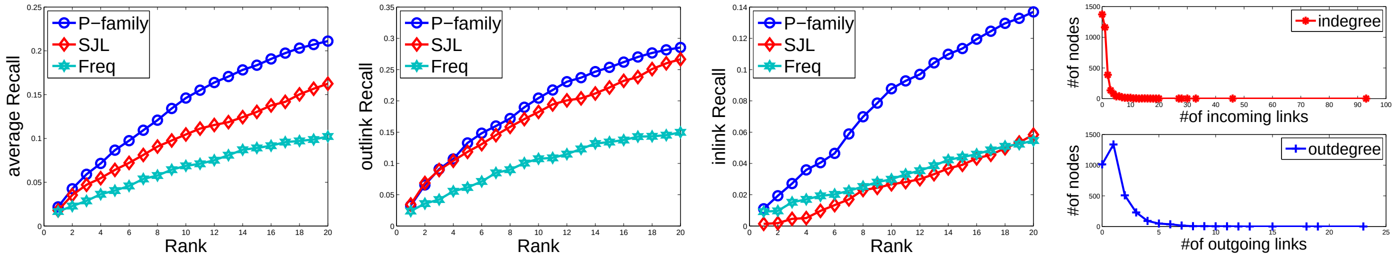
(a) average Recall on Political Blog (b) *outlink* Recall on Political Blog (c) *inlink* Recall on Political Blog (d) degree distribution

Figure 1: Average (a), outlink (b), and inlink (c) recalls at ranks 1 through 20 for different models on Political Blog data. The histograms of degree distribution of the network are shown in (d).



(a) average Recall on Cora (b) *outlink* Recall on Cora (c) *inlink* Recall on Cora (d) degree distribution

Figure 2: Average (a), outlink (b), and inlink (c) recalls at ranks 1 through 20 for different models on Cora data. The histograms of degree distribution of the network are shown in (d).

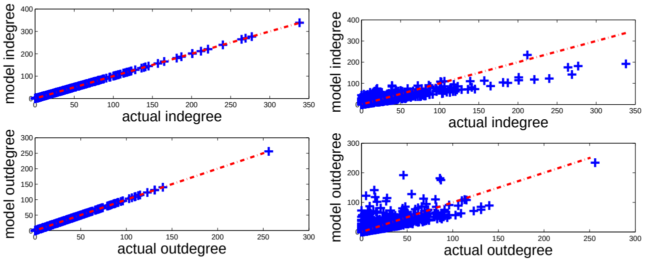


(a) average Recall on Citeseer (b) *outlink* Recall on Citeseer (c) *inlink* Recall on Citeseer (d) degree distribution

Figure 3: Average (a), outlink (b), and inlink (c) recalls at ranks 1 through 20 for different models on Citeseer data. The histograms of degree distribution of the network are shown in (d).

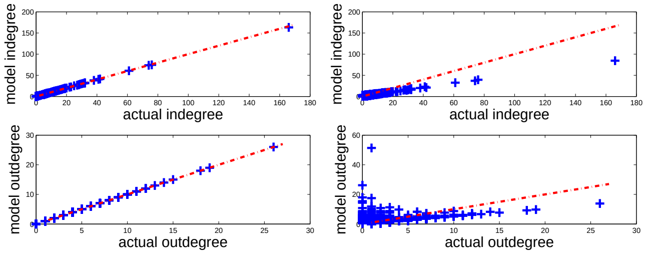
has extremely poor performance (almost constantly zero). To see why this case is special, we show the degree distributions of the three networks in the rightmost panels of Figures 1, 2, and 3. All the degree distributions follow a power-law distribution except the outdegree distribution in Cora network. The outdegree in Cora follows a rather uniform distribution with outdegree no lager than 5. (We suspect such a distribution is due to the small scale of the Cora data which leads to many references, and therefore outlinks, to be outside the data set.) Because of such a uniform distribution, the outdegrees of nodes are not informative, which explains the extremely poor performance of the Frequency-based model. The P-family models treat indegree and outdegree equally importantly and therefore also suffer from the uninformative outdegree distribution. SJL, in comparison, ignores the link direction and as a result makes the more informative indegree distribution dominate the uninformative outdegree distribution and therefore suffers the least. This special case actually reveals some trade-offs made by different models.

6.4 Degree Fitting Finally, we verify the degree fitting properties of PPL models. Figure 4(a) shows the scatter plots for the indegree and outdegree fitting of PPL models on the Political Blog data set. Note that PoL, PrL and PPL-D again give almost the same result is this experiment and so we refer to them together as PPL. Each point in the plot represents a node, where its position on the horizontal axis is determined by its actual degree (indegree or outdegree) and its position on the vertical axis is determined by the degree predicted by the model. Therefore, a point fell on the diagonal line (the red lines in the plots) indicates a perfect degree match. As can be seen from the figure, all the points fall on the red line, which indicates that PPL captures the degree distributions for each node exactly. In comparison, as shown in Figure 4(b), SJL has very poor performance in terms of degree fitting. Similar results are obtained for the paper citation data sets, where in Figures 5(a) and 5(b) we show some of the results. These empirical studies clearly validate the degree fitting property of the PPL models that we previously stated in Section 3.3.



(a) degree fitting by PPL (b) degree fitting by SJL

Figure 4: Scatter plots for degree fitting on the Political Blog data for (a) PPL and (b) SJL.



(a) degree fitting by PPL (b) degree fitting by SJL

Figure 5: Scatter plots for degree fitting on the indegree of Cora data (upper panels) and outdegree of Citeseer data (lower panels) for (a) PPL and (b) SJL.

7 Conclusion and Future Work

Stochastic block model is a promising probabilistic model for community detection. In this paper, we present a new stochastic block model, PPL, for community detection in *directed* networks. On one hand, our model is *complete*, in that it captures the roles of each node both as a link producer and as a link receiver whereas a consistent community membership serves both the roles; on the other hand, our model is *unified*, in that it offers a unified framework to connect and to understand existing models. We believe such a complete and unified model provides a solid foundation for further studies in stochastic block models for community detection. For future work, we are in the process of incorporating information other than links, such as the content information, into the model to obtain an even more general framework.

References

[1] L. Adamic and N. Glance. The political blogosphere and the 2004 U.S. election: divided they blog. In *Proc. of the 3rd International Workshop on Link Discovery*, 2005.

[2] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing. Mixed membership stochastic block models for relational data with application to protein-protein interactions. In *Proc. of the International Biometrics Society Annual Meeting*, 2006.

[3] D. Cohn and H. Chang. Learning to probabilistically identify authoritative documents. In *Proc. of the 17th International Conference on Machine Learning*, 2000.

[4] D. Cohn and T. Hofmann. The missing link - a probabilistic model of document content and hypertext connectivity. In *Proc. of the 13th Neural Information Processing Systems*, 2001.

[5] L. Dietz, S. Bickel, and T. Scheffer. Unsupervised prediction of citation influences. In *Proc. of the 24th International Conference on Machine Learning*, 2007.

[6] E. Erosheva, S. Fienberg, and J. Lafferty. Mixed membership models of scientific publications. In *Proceedings of the National Academy of Sciences*, 2004.

[7] S. Gregory. An algorithm to find overlapping community structure in networks. In *Proceedings of the 11th European Conference on Principles and Practice of Knowledge Discovery in Databases*, 2007.

[8] J. M. Hofman and C. H. Wiggins. A Bayesian approach to network modularity. *Phy. Rev. Letters*, 100, 2008.

[9] T. Hofmann. Probabilistic latent semantic indexing. In *Proc. of the 22nd International Conference on Research and Development in Information Retrieval*, 1999.

[10] P. Holland and S. Leinhardt. Local structure in social networks. *Socialogical Methodology*, 1976.

[11] E. A. Leicht and M. E. J. Newman. Community structure in directed networks. *Physical Review Letters*, 100, 2008.

[12] R. M. Nallapati, A. Ahmed, E. P. Xing, and W. W. Cohen. Joint latent topic models for text and citations. In *Proc. of the 14th ACM International Conference on Knowledge Discovery and Data Mining*, 2008.

[13] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phy. Rev. E*, 69, 2003.

[14] W. Ren, G. Yan, X. Liao, and L. Xiao. Simple probabilistic algorithm for detecting community structure. *Physical Review E*, 79, 2009.

[15] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(8), 2000.

[16] T. Yang, R. Jin, Y. Chi, and S. Zhu. Combining link and content for community detection: A discriminative approach. In *Proc. of the 15th ACM International Conference on Knowledge Discovery and Data Mining*, 2009.

[17] K. Yu, S. Yu, and V. Tresp. Soft clustering on graphs. In *Proc. of 19th Advances in Neural Information Processing Systems*, 2005.

[18] D. Zhou, J. Huang, and B. Schölkopf. Learning from labeled and unlabeled data on a directed graph. In *Proc. of the 22nd International Conference on Machine Learning*, 2005.

Appendix

Proof for Theorem 3.1 In order to prove Theorem 3.1, we first state the following lemma about the optimal solution to the PPL model given in Equation (3.4).

Lemma 7.1. *Given that $(a^3, b^3, c^3, \gamma^3)$ is the optimal solution to maximizing the log-likelihood of PPL model, we define $\pi_k = \sum_i \gamma_{ik}^3 c_i^3$. Then we can obtain one set of parameters (a^1, b^1, γ^1) such that $\sum_i \gamma_{ik}^1 a_i^1 = \pi_k$ and (a^1, b^1, γ^1) is the optimal solution to maximizing the log-likelihood of PoL model. Similarly, we can obtain another set of parameters (a^2, b^2, γ^2) such that $\sum_j \gamma_{jk}^2 b_j^2 = \pi_k$ and (a^2, b^2, γ^2) is the optimal solution to maximizing the log-likelihood of PrL model.*

Proof. we first show how to construct such (a^1, b^1, γ^1) and (a^2, b^2, γ^2).
Given $(a^3, b^3, c^3, \gamma^3)$ and $\pi_k = \sum_i \gamma_{ik}^3 c_i^3$, we can define $\hat{q}$ such that $\sum_i \gamma_{ik}^3 a_i^3 \hat{q}_k = \pi_k$. We then construct $\gamma_{ik}^1 = \frac{\gamma_{ik}^3 \hat{q}_k}{\sum_k \gamma_{ik}^3 \hat{q}_k}$, $a_i^1 = a_i^3 \sum_k \gamma_{ik}^3 \hat{q}_k$, $b_j^1 = \frac{b_j^3 \sum_k \gamma_{jk}^3 \hat{q}_k}{\sum_{j'} b_{j'}^3 \sum_k \gamma_{j'k}^3 \hat{q}_k}$, and we can show that

$$\sum_i \gamma_{ik}^1 a_i^1 = \pi_k$$

We can also define $\tilde{q}$ such that $\sum_j \gamma_{jk}^3 b_j^3 \tilde{q}_k = \pi_k$. We then construct $\gamma_{ik}^2 = \frac{\gamma_{ik}^3 \tilde{q}_k}{\sum_k \gamma_{ik}^3 \tilde{q}_k}$, $a_i^2 = \frac{a_i^3 \sum_k \gamma_{ik}^3 \tilde{q}_k}{\sum_{i'} a_{i'}^3 \sum_k \gamma_{i'k}^3 \tilde{q}_k}$ and $b_j^2 = b_j^3 \sum_k \gamma_{jk}^3 \tilde{q}_k$, and we can show that

$$\sum_j \gamma_{jk}^2 b_j^2 = \pi_k$$

With constructed (a^1, b^1, γ^1) and (a^2, b^2, γ^2) we can show that

$$\mathcal{L}_3(a^3, b^3, c^3, \gamma^3) = \mathcal{L}_1(a^1, b^1, \gamma^1) = \mathcal{L}_2(a^2, b^2, \gamma^2)$$

Next, we need to show that (a^1, b^1, γ^1) is the optimal solution to PoL model, (a^2, b^2, γ^2) is the optimal solution to PrL model. We prove this by contradiction. Assume their exists another set of parameters (a^*, b^*, γ^*) such that $\mathcal{L}_1(a^*, b^*, \gamma^*) > \mathcal{L}_1(a^1, b^1, \gamma^1) = \mathcal{L}_3(a^3, b^3, c^3, \gamma^3)$, then

$$\mathcal{L}_3(a^*, b^*, a^*, \gamma^*) = \mathcal{L}_1(a^*, b^*, \gamma^*) > \mathcal{L}_3(a^3, b^3, c^3, \gamma^3)$$

which contradicts that $(a^3, b^3, c^3, \gamma^3)$ is the optimal solution to PPL model. Similarly, we can show (a^2, b^2, γ^2) is the optimal solution to PrL model. Thus, we complete the proof.

Following the above lemma, we can easily prove Theorem 3.1.

Proof for Theorem 3.2

Proof. We can easily show that the optimal solution to a_i in PoL model is equal to the normalized outdegree of node i, i.e., $a_i^1 = \frac{\sum_j s_{ij}}{\sum_{ij} s_{ij}}$; and the optimal solution to b_j in PrL model is equal to the normalized indegree of node j, i.e., $b_j^2 = \frac{\sum_i s_{ij}}{\sum_{ij} s_{ij}}$. From the model formulation in Equation (3.7) for PoL model, we have

$$\Pr{}^1(i_\rightarrow | a^1, b^1, \gamma^1) = \sum_j \Pr{}^1(i_\rightarrow, j_\leftarrow | a^1, b^1, \gamma^1) = a_i^1$$

So the model outdegree distribution of PoL model fits exactly the actual outdegree distribution of the network.

From the model formulation in Equation (3.8) for PrL model, we have

$$\Pr{}^2(j_\leftarrow | a^2, b^2, \gamma^2) = \sum_i \Pr(i_\rightarrow, j_\leftarrow | a^2, b^2, \gamma^2) = b_j^2$$

So the model indegree distribution of PrL model fits exactly the actual indegree distribution of the network. Following Theorem 3.1 we have

$$\begin{aligned} \Pr{}^3(i_\rightarrow | a^3, b^3, c^3, \gamma^3) &= \Pr{}^2(i_\rightarrow | a^2, b^2, \gamma^2) \\ &= \Pr{}^1(i_\rightarrow | a^1, b^1 \gamma^1) = a_i^1 \end{aligned}$$

and

$$\begin{aligned} \Pr{}^3(j_\leftarrow | a^3, b^3, c^3, \gamma^3) &= \Pr{}^1(j_\leftarrow | a^1, b^1, \gamma^1) \\ &= \Pr{}^2(j_\leftarrow | a^2, b^2, \gamma^2) = b_j^2 \end{aligned}$$

We conclude that the model indegree and outdegree distributions estimated from PoL model, PrL model and PPL model fit exactly the actual indegree and outdegree distributions of the network.

Proof for Theorem 4.1

Proof. The joint link probability of SJL model is given in Equation (2.3), i.e.,

$$\Pr(i_{\rightarrow}, j_{\leftarrow}) = \sum_k \beta_{ik}\beta_{jk}\pi_k$$

The community membership of SJL model is defined as[17, 14]

$$\gamma^f_{ik} = \frac{\beta_{jk}\pi_k}{\sum_k \beta_{ik}\pi_k} \tag{7.12}$$

We can also define c^f_i as

$$c^f_i = \sum_k \beta_{ik}\pi_k \tag{7.13}$$

Similarly, given solution (γ, c) to PPL model with $a = b = c$, we can define

$$\pi^p_k = \sum_i \gamma_{ik}c_i, \quad \beta^p_{ik} = \frac{\gamma_{ik}c_i}{\sum_i \gamma_{i'k}c_{i'}} \tag{7.14}$$

All we need to show is that given that (β, π) is the solution to SJL model, (γ^f, c^f) defined as in Equations (7.12,7.13) is the solution to PPL model under the restriction of $a = b = c$; and given that (γ, c) is the optimal solution to PPL model under the restriction of $a = b = c$, (π^p, β^p) defined in Equation (7.14) is the optimal solution to SJL model. First note that

$$\begin{aligned} \mathcal{L}_0(\beta, \pi) &= \mathcal{L}_3(\gamma^f, c^f) \\ \mathcal{L}_3(\gamma, c) &= \mathcal{L}_0(\beta^p, \pi^p) \end{aligned}$$

where $\mathcal{L}_0$ and $\mathcal{L}_3$ are the log-likelihood of SJL model and PPL model respectively. Given that (β, π) is the optimal solution to SJL model, if there exists (γ^*, c^*) such that $\mathcal{L}_3(\gamma^*, c^*) > \mathcal{L}_3(\gamma^f, c^f) = \mathcal{L}_0(\beta, \pi)$, then we can construct (β^*, π^*) as in Equation (7.14) such that $\mathcal{L}_0(\beta^*, \pi^*) = \mathcal{L}_3(\gamma^*, c^*) > \mathcal{L}_0(\beta, \pi)$, which contradicts the assumption that (β, π) is the optimal solution to SJL model. Similarly, given that (γ, c) is the optimal solution to PPL model, if there exists (π^*, β^*) such that $\mathcal{L}_0(\pi^*, \beta^*) > \mathcal{L}_0(\pi^p, \beta^p) = \mathcal{L}_3(\gamma, c)$, then we can construct (γ^*, c^*) as in Equations (7.12,7.13) such that $\mathcal{L}_3(\gamma^*, c^*) = \mathcal{L}_0(\pi^*, \beta^*) > \mathcal{L}_3(\gamma, c)$, which contradicts the assumption that (γ, c) is the optimal solution to PPL model. Therefore, we prove that PPL model under the restriction of $a = b = c$ is equivalent to SJL model.

Proof for Theorem 5.1 In the E-step, we would bound the log-likelihood from below. The key point is to apply the Jensen inequality $\log \sum_k p_k \ge \sum_k q_k \log p_k / q_k$, where $\sum_k q_k = 1$, to the log-sum-term in the log-likelihood and to apply the inequality $-\log x \ge 1 - \dfrac{x}{y} - \log y$ to the summation term in the denominator of the log-sum-term in the log-likelihood. In particular, at the t-th iteration the log-sum-term is lower bounded as

$$\log \sum_k \Pr(i, j, k) \ge \sum_k q_{ijk} \log \Pr(i, j, k)/q_{ijk}$$

with q_{ijk} computed as $q_{ijk} \propto \Pr^{t-1}(i, j, k)$ where superscript $t-1$ means the probability is computed under the values of the parameters in the $(t-1)$-th iteration. Then the denominator term in $\Pr(i, j, k)$ would be lower bounded as

$$\begin{aligned} -\log \sum_{i'} \gamma_{i'k}a_{i'} &\ge 1 - \frac{\sum_{i'} \gamma_{i'k}a_{i'}}{\eta_k} - \log \eta_k \\ -\log \sum_{i'} \gamma_{i'k}b_{i'} &\ge 1 - \frac{\sum_{i'} \gamma_{i'k}b_{i'}}{\tau_k} - \log \tau_k \end{aligned}$$

with η_k, τ_k computed as

$$\eta_k = \sum_{i'} \gamma^{t-1}_{i'k} a^{t-1}_{i'} \quad \tau_k = \sum_{j'} \gamma^{t-1}_{j'k} b^{t-1}_{j'}$$

and the summation term $\sum_{i'} \gamma_{i'k}c_{i'}$ in PPL model is lower bounded as

$$\log \sum_{i'} \gamma_{i'k}c_{i'} \ge \sum_{i'} \zeta_{i'k} \log \gamma_{i'k}c_{i'}/\zeta_{i'k}$$

with ζ computed as $\zeta_{ik} = \dfrac{\gamma^{t-1}_{ik}c^{t-1}_i}{\sum_{i'} \gamma^{t-1}_{i'k}c^{t-1}_{i'}}$. Due to the limit of space, we omit the details about deriving the lower bound of the three log-likelihoods.

In the M-step, we will maximize the corresponding lower bound over the corresponding parameters as follows:

$$\text{PoL} : \sum_{(i,j)\in\mathcal{E}} s_{ij} \sum_k q_{ijk} \left(\log \gamma_{ik}\gamma_{jk}b_j - \sum_{j'} \frac{\gamma_{j'k}b_{j'}}{\tau_k} \right)$$

$$\text{PrL} : \sum_{(i,j)\in\mathcal{E}} s_{ij} \sum_k q_{ijk} \left(\log \gamma_{jk}\gamma_{ik}a_i - \sum_{i'} \frac{\gamma_{i'k}a_{i'}}{\eta_k} \right)$$

$$\begin{aligned} \text{PPL-D} : &\sum_{(i,j)\in\mathcal{E}} s_{ij} \sum_k q_{ijk} \Bigg(\log \gamma_{ik}a_i\gamma_{jk}b_j - \sum_{i'} \frac{\gamma_{i'k}a_{i'}}{\eta_k} \\ &- \sum_{i'} \frac{\gamma_{i'k}b_{i'}}{\tau_k} + \sum_{i'} (\zeta_{i'k} + \alpha) \log c_i + \sum_{i'} \zeta_{i'k} \log \gamma_{i'k} \Bigg) \end{aligned}$$

By taking the derivatives of the expressions and setting them to zero, we can obtain the corresponding formulas in the M-step.