Statistica Sinica **29** (2019), 1277-1299
doi:https://doi.org/10.5705/ss.202017.0103

A POPULARITY-SCALED LATENT SPACE MODEL FOR LARGE-SCALE DIRECTED SOCIAL NETWORK

Xiangyu Chang, Danyang Huang and Hansheng Wang

Xi'an Jiaotong University, Renmin University of China and Peking University

Abstract: Large-scale directed social network data often include degree heterogeneity, reciprocity, and transitivity properties. Thus, a sensible network-generating model should consider these features. To this end, we propose a popularity-scaled latent space model for large-scale directed network structure formulations. This model assumes each node occupies a position in a hypothetically assumed latent space. Then, the nodes close to (far away from) each other should have a higher (lower) probability of being connected. Thus, reciprocity and transitivity can be derived analytically. In addition, we assume a popularity parameter for each node. Nodes with larger (smaller) popularity are more (less) likely to be followed. By assuming different distributions for the popularity parameters, we model various types of degree heterogeneity. Based on the proposed model, we construct a comprehensive probabilistic index for link prediction. We demonstrate the performance of the proposed model using simulation studies and a Sina Weibo data set. The results show that the performance of the model is competitive.

Key words and phrases: Degree heterogeneity, large-scale social network, latent space model, link prediction, reciprocity, transitivity.

1. Introduction

We consider a network with n nodes (indexed by $i = 1, \ldots, n$) and a set of directed edges (denoted by $a_{ij} \in \{0, 1\}$, with $1 \le i, j \le n$). Here, $a_{ij} = 1$ if a relationship exists from node i to node j, and $a_{ij} = 0$ otherwise. For example, in Twitter, $a_{ij} = 1$ implies that node i follows node j. Throughout the rest of this paper, we do not allow any node to directly follow itself, that is, $a_{ii} = 0$ for any $1 \le i \le n$. This leads to a high-dimensional binary matrix $A = (a_{ij}) \in \mathbb{R}^{n\times n}$, called an adjacency matrix. An adjacency matrix is an important tool used to describe a network structure. For example, it can be used to describe a large-scale social network (e.g., Facebook or Twitter), where different users are nodes and their relationships are edges. It can also be used to describe connections between websites, where individual websites are nodes and hyperlinks are edges Adamic and Huberman (2000). Note that network data, which can be represented by an adjacency matrix, are common in various scientific disciplines, including anthropology, economics, education, marketing, psychology, physics, and sociology. For a good summary, refer to Holland and Leinhardt (1981), Wasserman and Faust (1994), Knoke and Yang (2008), and Newman (2010). As a result, related problems are becoming increasingly relevant.

As an important research direction, a large body of literature studies how network structures (i.e., the adjacency matrix A) are generated. Accordingly, various probabilistic models have been proposed. The simplest of these is the so-called Erdös-Rényi model (Erdős and Rényi (1960)), in which edges (i.e., a_{ij}) are assumed to be independently generated from a Bernoulli distribution. To allow for reciprocal dependence, Holland and Leinhardt (1981) proposed the well-known p_1 model. Later, Wang and Wong (1987) and Nowicki and Snijders (2001) extended the model to accommodate stochastic block structures. See also Airoldi et al. (2008), Bickel and Chen (2009), Choi, Wolfe and Airoldi (2012), and Bickel et al. (2013) for relevant discussions. Because these models are all based on certain independence assumptions about either edges or dyads, they cannot describe more complicated higher-order dependence structures. Consider, for example, three arbitrary nodes (denoted by i, j, and k). Here, we expect the marginal probability $P(a_{ij} = 1)$ to be very small and close to zero, because large-scale social networks are typically extremely sparse (Huang et al. (2015)). However, given $a_{ik} = 1$ and $a_{kj} = 1$, the conditional probability $P(a_{ij} = 1|a_{ik}a_{kj} = 1)$ should be nontrivial and clearly above zero. This interesting phenomenon is referred to as the *transitivity* property (Hoff, Raftery and Handcock (2002); Faust (1988)) and has been observed extensively in empirical networks. Unfortunately, it cannot be described by any of the aforementioned network models.

In another related literature stream, Frank and Strauss (1986) studied an exponential random graph model (ERGM), which has a more flexible theoretical framework. According to the ERGM, the random behavior of A is fully determined by a p-dimensional sufficient statistic through an exponential transformation. Based on the choice of sufficient statistics, the ERGM might contain p_1 and stochastic block models as special cases. The ERGM might also have the transitivity property if the sufficient statistic is selected appropriately. However, the likelihood function involves a normalizing constant, the computation of which requires ultrahigh-dimensional integration. This makes the ERGM difficult to estimate for a large-scale social network. To solve this problem, the Markov chain Monte Carlo (MCMC)-type algorithms and accompanying R packages have been developed (Hunter et al. (2008)). From a practical perspective, a pseudo-likelihood method has been proposed (Van Duijn, Gile and Handcock (2009)). However, MCMC algorithms are not feasible for large-scale networks because they involves sampling networks in each iteration, which leads to a prohibitive computational cost when n is large (Hunter and Handcock (2006)).

Another research direction is the latent space model (LSM) (Hoff, Raftery and Handcock (2002)). In the LSM, nodes are assumed to be embedded in a hypothetically assumed latent space. Thus, the probability of two nodes being connected is assumed to be negatively correlated with their Euclidean distance in the latent space. Hoff, Raftery and Handcock (2002), employed a logit-type link function and considered a set of observed covariates. As a result of the symmetry of the Euclidean distance, the LSM cannot describe degree heterogeneity (Hoff (2003); Krivitsky et al. (2009)), which is observed extensively in the empirical network data sets. To fix this problem, Hoff (2003) and Krivitsky et al. (2009) introduced nodal random effects and Austin, Linkletter and Wu (2013) regressed the latent positions on nodal attributes. The aforementioned latent space methods are conceptually attractive, but their estimations require computationally intensive MCMC algorithms. Most recently, graphon-based methods have been proposed for network modeling (Wolfe and Olhede (2013); Olhede and Wolfe (2014); Gao, Lu and Zhou. (2015)). A graphon is a nonnegative symmetric function f, measurable and bounded, used to model the probability of two nodes being connected, such as $P(a_{ij} = 1) = f(Z_i, Z_j)$, where Z_i and Z_j are latent variables. In contrast to the classical LSM, which typically assumes a multivariable normal distribution for the latent variables, graphon-based models suppose Z_i and Z_j are drawn from a uniform distribution between zero and one, which leads to elaborate theoretical properties.

In this work, we propose a new latent space model, which we refer to as a popularity-scaled latent space model (PSLSM) for large-scale social network. Similar to the LSM (Hoff, Raftery and Handcock (2002)), the new model assumes each node i occupies a position $Z_i \in \mathbb{R}^1$ in a hypothetically assumed latent space. Then, for two arbitrary nodes i and j, the PSLSM assumes that the likelihood for $a_{ij} = 1$ is negatively correlated with the scaled inter-node distance $(Z_i - Z_j)^2/\gamma_j^2$. Here, $(Z_i - Z_j)^2$ is the inter-node Euclidean distance, and $\gamma_j > 0$ is a positive scale parameter for j. Intuitively, γ_j measures the popularity of the ending node j. For example, if j is popular, then its popularity parameter γ_j should be large, making the scaled inter-node distance $(Z_i - Z_j)^2/\gamma_j^2$ small.

Consequently, the probability of i following j should be large. Accordingly, j's in-degree, defined as $d_j = \sum_i a_{ij}$, should be high. However, the probability of j following i remains small, because it is determined by $(Z_i - Z_j)^2/\gamma_i^2$, and γ_i should be small for a usual node. Accordingly, i's in-degree should be small. As a result, degree heterogeneity is accommodated. Popularity parameters have been considered in several works. For example, Karrer and Newman (2011) proposed a degree-corrected blockmodel for detecting communities of networks with hubs. Sarkar and Moore (2005), Krivitsky et al. (2009), and Daniel and Chen (2015) have tried a similar idea of including popularity parameters in a latent space model. However, although this improved the flexibility of the model, the computational complexity increased, primarily because they use logit links in the models. Compared with the above models, the proposed PSLSM is a type of LSM. Therefore, higher-order network properties (e.g., transitivity) can be easily accommodated. However, unlike the the other methods, the PSLSM is analytically tractable and thus, computationally simple owing to the probit link function in (2.1). For instance, a comprehensive probabilistic index can be analytically derived for link prediction. Its competitive performance is demonstrated using simulation studies and a Sina Weibo (a Chinese social network similar to Twitter) data set.

The rest of this paper is organized as follows. Section 2 presents the model and analytically discusses its properties. The corresponding probability index for link prediction is constructed in Section 3. Here, we also demonstrate the finite-sample performance of the proposed model using a simulation and a Sina Weibo data set. A brief discussion of our findings is given in Section 4.

2. A Popularity-Scaled Latent Space Model

2.1. Model and notation

Consider a network with n nodes, indexed by $1 \leq i \leq n$. Next, define a binary variable $a_{ij} = 1$ if i follows j, and $a_{ij} = 0$ otherwise. Following the standard literature, we require $a_{ii} = 0$ for $1 \leq i \leq n$. Next, in a given latent space (Hoff, Raftery and Handcock (2002)), we assume each node occupies a position $Z_i \in \mathbb{R}^d, d \geq 1$. These positions are collected by $\mathbb{Z} = \{Z_i : 1 \leq i \leq n\}$. Then, PSLSM is defined as,

$$P(a_{ij} = 1|\mathbb{Z}, \mathcal{P}) = P(a_{ij} = 1|Z_i, Z_j, \gamma_j) = \exp\left\{-\frac{\|Z_i - Z_j\|^2}{2\gamma_j^2}\right\}, \tag{2.1}$$

where $\gamma_j > 0$ is j's popularity parameter, and all popularity parameters are collected by $\mathcal{P} = \{\gamma_i : 1 \le i \le n\}$. Note that covariates could be involved in the model. For simplicity, we consider the model with no covariates in this paper. Conditional on $\mathbb{Z}$ and $\mathcal{P}$, the variables a_{ij} are assumed to be independent. This never implies that they are marginally independent. In fact, edges could be marginally well correlated, if they share a common node. Moreover, we can easily verify that γ_j, Z_i, and Z_j are not jointly identifiable. To demonstrate this, let c be an arbitrary nonzero constant. We can then redefine $\gamma_j' := c\gamma_j$, $Z_i' := cZ_i$, and $Z_j' := cZ_j$. Then, model (2.1) remains valid. To fix the problem, we assume that Z_i is independently generated from a standard normal distribution. As a result, $\gamma_j > 0$ can be identified (Refer to subsection 2.2 for further detail). Theoretically, many distributions can be assumed for Z_i. We adopt a normal distribution only because it works well with the model formulation (2.1). Furthermore, if we assume that $d = 1$, this leads to elegant and analytically tractable derivations (see the next four subsections). Therefore, throughout this paper, we assume $d = 1$, and that Z_i is drawn from a standard normal distribution.

2.2. Degree distribution

We first study the degree distribution implied by (2.1). We consider both in- and out-degrees (Wasserman and Faust (1994); Barabási and Albert (1999); Clauset, Shalizi and Newman (2009); Zhang and Chen (2013)) given by, respectively, $d_i = \sum_j a_{ji}$ and $d_i^* = \sum_j a_{ij}$. The out-degree d_i^* counts the number of followees. Intuitively, active nodes are likely to have many followees. Thus, the out-degree d_i^* can be viewed as a rough measure of how active i is. In contrast, the in-degree d_i denotes the number of followers. Intuitively, popular nodes are likely to have many followers. Thus, d_i should have a close relationship with i's popularity parameter γ_i. We consider the in-degree first. For convenience, let $T_{ij} = Z_i - Z_j$. Then, the random variable T_{ij} follows a normal distribution with mean 0 and variance 2. Thus, we have

$$\begin{aligned} E(d_i|\gamma_i) &= (n-1)E(a_{ji}|\gamma_i) = (n-1)P(a_{ji} = 1|\gamma_i) \\ &= (n-1)\int_{\mathbb{R}^2} P(a_{ji} = 1|\gamma_i, Z_i, Z_j)dF_Z(z_i)dF_Z(z_j) \\ &= (n-1)\int_{\mathbb{R}} P(a_{ji} = 1|\gamma_i, T_{ij})dF_T(t) \\ &= \frac{(n-1)}{2\sqrt{\pi}}\int_{\mathbb{R}} \exp\left\{-\frac{t^2}{2}\left(\frac{1}{\gamma_j^2} + \frac{1}{2}\right)\right\} dF_T(t) \end{aligned}$$

$$= \frac{(n-1)\gamma_i}{\sqrt{2+\gamma_i^2}}. \tag{2.2}$$

Note that large-scale networks are typically extremely sparse (Watts and Strogatz (1998); Schweinberger and Handcock (2015); Huang et al. (2015)). Empirically, the observed in-degree for each node is bounded, on average. For example, for Sina Weibo, each node is allowed to have no more than 2,000 followees. Thus, the average out-degree is bounded. Because the total number of links is fixed, the average in-degree and average out-degree should be the same. This makes the average in-degree bounded. Mathematically, this requires that $E(d_i|\gamma_i) = O_p(1)$. However, by (2.2), if γ_i is fixed, d_i should diverge towards infinity as the network size $n \to \infty$. To satisfy $E(d_i|\gamma_i) = O_p(1)$, we must have $\gamma_i = O_p(1/n)$. Therefore, we redefine $\sigma_i = n\gamma_i$ for the remainder of this paper, leading to $\sigma_i = O_p(1)$. For convenience, we also refer to σ_i as i's *popularity*. Then, model (2.1) becomes,

$$P(a_{ij} = 1|\mathbb{Z}, \mathcal{P}) = \exp\left\{-\frac{n^2(Z_i - Z_j)^2}{2\sigma_j^2}\right\}, \tag{2.3}$$

and the expected degree (2.2) can be re-expressed as

$$E(d_i|\gamma_i) = \frac{\sigma_i}{\sqrt{2+\sigma_i^2/(n-1)^2}} = \frac{\sigma_i}{\sqrt{2}} + O_p\left(\frac{1}{n^2}\right). \tag{2.4}$$

By (2.4) and for a given node i, the expected in-degree is determined mainly by its popularity σ_i. Correspondingly, any type of degree heterogeneity can be expressed approximately, as long as an appropriate distribution is assumed for σ_i. For example, if a power-law or log-normal type distribution is assumed for σ_i, we expect the observed degree sequence to be very heavy-tailed (Barabási and Albert (1999); Liben-Nowell et al. (2005); Kim and Leskovec (2012)). Similarly, for the out-degree,

$$E(d_i^*|\gamma_i) = \frac{1}{n}\sum_{i\neq j}\frac{\sigma_j}{\sqrt{2}} + O_p\left(\frac{1}{n^2}\right) = \frac{E(\sigma_i)}{\sqrt{2}} + O_p\left(\frac{1}{n^{1/2}}\right),$$

which is approximately constant and not siginificantly affected by i's popularity σ_i. Therefore, we expect that the variability exhibited by d_i^* should be substantially smaller than that of d_i. This is confirmed by many empirical data sets (e.g., the Sina Weibo data set in Section 3.3).

2.3. Reciprocity

The reciprocity property is explored in this section under model (2.1). Consider two arbitrary nodes i and j. $P(a_{ij} = 1)$ is expected to be close to zero, because large-scale social networks are typically extremely sparse. However, given $a_{ij} = 1$, the conditional probability $P(a_{ji} = 1|a_{ij} = 1)$ should be well above zero, which is referred to as *reciprocity* (Faust (1988); Hoff, Raftery and Handcock (2002)). We evaluate this probability for the PSLSM as follows:

$$P(a_{ji} = 1|a_{ij} = 1, \gamma_i, \gamma_j) = \frac{P(a_{ji}a_{ij} = 1|\gamma_i, \gamma_j)}{P(a_{ij} = 1|\gamma_i, \gamma_j)}.$$

Note that the numerator is given by

$$\begin{aligned} P(a_{ji}a_{ij} = 1|\gamma_i, \gamma_j) &= \int_{\mathbb{R}^2} P(a_{ji}a_{ij} = 1|\gamma_i, \gamma_j, Z_i, Z_j) dF_Z(z_i) dF_Z(z_j) \\ &= \left(\frac{\sigma_i^2 \sigma_j^2}{2n^2\sigma_i^2 + 2n^2\sigma_j^2 + \sigma_i^2\sigma_j^2} \right)^{1/2}. \end{aligned}$$

Combining this result with (2.1), we obtain

$$\begin{aligned} P(a_{ji} = 1|a_{ij} = 1, \gamma_i, \gamma_j) &= \left(\frac{2n^2\sigma_i^2 + \sigma_i^2\sigma_j^2}{2n^2\sigma_i^2 + 2n^2\sigma_j^2 + \sigma_i^2\sigma_j^2} \right)^{1/2} \\ &= \left(\frac{\sigma_i^2}{\sigma_i^2 + \sigma_j^2} \right)^{1/2} + O_p\left(\frac{1}{n^2} \right). \end{aligned} \tag{2.5}$$

Note that (2.5) confirms the reciprocity property, that is, $P(a_{ji} = 1|a_{ij} = 1, \gamma_i, \gamma_j)$ is approximately a fixed positive number $\sigma_i^2/(\sigma_i^2 + \sigma_j^2)$. However, this changes if j has many followers. In that case, we should have a large σ_j^2 and an extremely small reciprocal probability. Practically, this means that many normal users like to follow celebrities. However, celebrities seldom follow them back. Thus, the reciprocity property fails.

2.4. Transitivity

Transitivity is another stylized network phenomenon that needs to be investigated for the PSLSM (Hoff, Raftery and Handcock (2002); Krivitsky et al. (2009)). For three arbitrary nodes, denoted by i, j, and k, transitivity refers to the fact that $P(a_{ji} = 1)$ should be clearly above zero, if both i and j are connected to a common node k in some way. Otherwise, the likelihood for $a_{ji} = 1$ should be extremely low, because large-scale social networks are extremely sparse. This amounts to evaluating the probability

$$P(a_{ji} = 1|i \text{ and } j \text{ are connected with common node } k). \tag{2.6}$$

Based on how i and j are connected to node k, the above probability can be grouped into nine different types; see Figure 1 for an intuitive illustration. In this

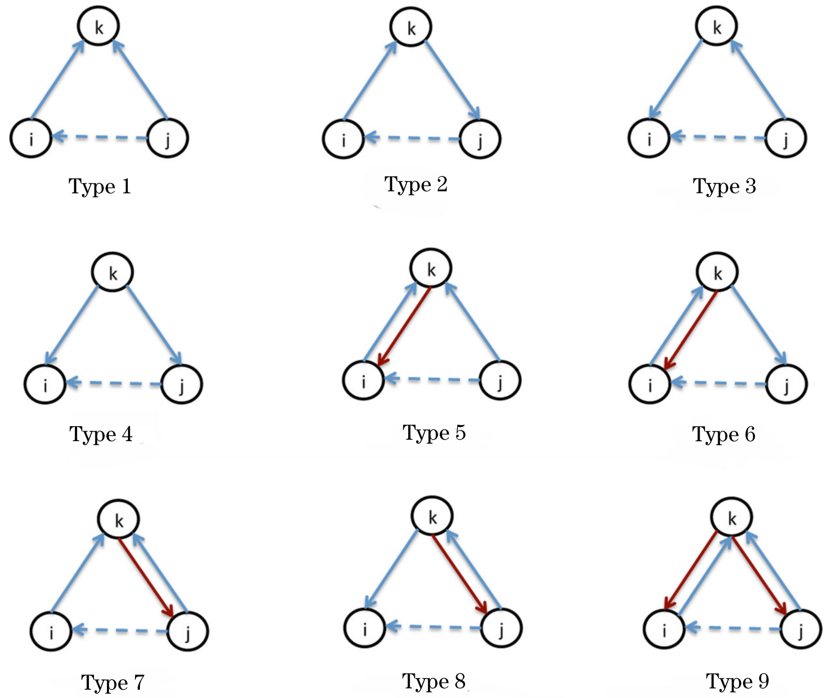

Figure 1. We consider nine typical transitivity types. j is a starting node, and i is an ending node. An arrow implies a follower-followee relationship. For example, $k \to j$ means $a_{kj} = 1$.

paper, the nine structures in Figure 1 are referred to as a generalized transitivity structure. The detailed analytical results are given below. See Appendix A for technical details. The probabilities are,

TYPE 1. $P(a_{ji} = 1 | a_{ik}a_{jk} = 1, \mathcal{P}) = \left(\dfrac{\sigma_i^2}{\sigma_i^2 + 2\sigma_k^2}\right)^{1/2} + O_p\left(\dfrac{1}{n^2}\right)$,

TYPE 2. $P(a_{ji} = 1 | a_{ik}a_{kj} = 1, \mathcal{P}) = \left(\dfrac{\sigma_i^2}{\sigma_i^2 + \sigma_j^2 + \sigma_k^2}\right)^{1/2} + O_p\left(\dfrac{1}{n^2}\right)$,

TYPE 3. $P(a_{ji} = 1 | a_{ki}a_{jk} = 1, \mathcal{P}) = \left(\dfrac{\sigma_i^2}{2\sigma_i^2 + \sigma_k^2}\right)^{1/2} + O_p\left(\dfrac{1}{n^2}\right)$,

TYPE 4. $P(a_{ji} = 1 | a_{ki}a_{kj} = 1, \mathcal{P}) = \left(\dfrac{\sigma_i^2}{\sigma_i^2 + 2\sigma_j^2}\right)^{1/2} + O_p\left(\dfrac{1}{n^2}\right)$,

TYPE 5. $P(a_{ji} = 1 | a_{ik}a_{ki}a_{jk} = 1, \mathcal{P}) = \left(\dfrac{\sigma_i^4 + \sigma_k^2\sigma_i^2}{\sigma_i^4 + 3\sigma_k^2\sigma_i^2 + \sigma_k^4}\right)^{1/2} + O_p\left(\dfrac{1}{n^2}\right)$, TYPE 6. $P(a_{ji}=1|a_{ik}a_{ki}a_{kj}=1,\mathcal{P})$

$$=\left(\frac{\sigma_i^4+\sigma_i^2\sigma_k^2}{\sigma_i^4+\sigma_i^2\sigma_j^2+\sigma_j^2\sigma_k^2+2\sigma_i^2\sigma_k^2}\right)^{1/2}+O_p\left(\frac{1}{n^2}\right),$$

TYPE 7. $P(a_{ji}=1|a_{ik}a_{kj}a_{jk}=1,\mathcal{P})$

$$=\left(\frac{\sigma_i^2\sigma_j^2+\sigma_j^2\sigma_k^2}{\sigma_i^2\sigma_j^2+\sigma_i^2\sigma_k^2+2\sigma_j^2\sigma_k^2+\sigma_k^4}\right)^{1/2}+O_p\left(\frac{1}{n^2}\right),$$

TYPE 8. $P(a_{ji}=1|a_{ki}a_{jk}a_{kj}=1,\mathcal{P})$

$$=\left(\frac{\sigma_i^2\sigma_j^2+\sigma_i^2\sigma_k^2}{2\sigma_i^2\sigma_k^2+\sigma_j^2\sigma_k^2+2\sigma_i^2\sigma_j^2}\right)^{1/2}+O_p\left(\frac{1}{n^2}\right),$$

TYPE 9. $P(a_{ji}=1|a_{ik}a_{ki}a_{jk}a_{kj}=1,\mathcal{P})$

$$=\left(\frac{\sigma_i^4\sigma_j^2+\sigma_i^4\sigma_k^2+\sigma_i^2\sigma_k^4+\sigma_i^2\sigma_j^2\sigma_k^2}{\sigma_i^4\sigma_j^2+\sigma_i^4\sigma_k^2+2\sigma_i^2\sigma_k^4+3\sigma_i^2\sigma_j^2\sigma_k^2+\sigma_k^4\sigma_j^2}\right)^{1/2}+O_p\left(\frac{1}{n^2}\right).$$

The above results share the following common properties. Holding σ_i^2, σ_j^2, and σ_k^2 fixed and letting $n\to\infty$, all transitivity probabilities given above converge to a nonzero constant. However, if the common node k has many followers in an extreme case $\sigma_k^2\to\infty$, the transitivity probabilities of Types 1, 2, 3, 5, and 7 converge to zero. This is because, for these five cases, the super popular common node k fails to follow i and j simultaneously. This provides no evidence that the latent positions of i and j should be sufficiently close to each other. However, as long as k follows both i and j simultaneously (i.e., Types 4, 6, 8, and 9), the transitivity property remains valid, even if k has many followers. This suggests that the probability (2.6) is greatly affected by both the transitivity type (i.e., one of the above nine types) and the popularity level of the common node (i.e, σ_k^2).

2.5. Common neighbors

The previous subsection presented explicit expressions of nine types of generalized transitivity structures, among which, we find the Type 3 is particularly important. This is because, for many empirical directed networks, this type of transitivity is especially useful for link predictions (Lü and Zhou (2011); Yu and Wang (2014)); see the next section for further detail. The Sina Weibo data set in Section 3.3 provides empirical evidence. Thus, we are motivated to explore this particular type of transitivity property in greater depth. Specifically, we focus on the type of network structure given in Figure 2, which can be viewed as a

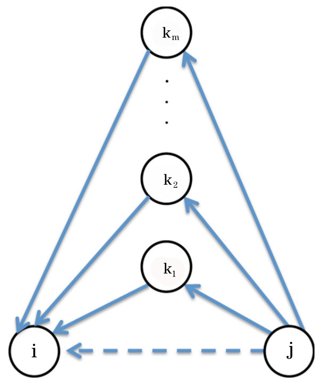

Figure 2. Common neighbors. j is a starting node and i is an ending node. i and j share m common neighbors, denoted by $k_1, k_2, \ldots, k_m$.

generalization of the Type 3 transitivity in Figure 1. By Figure 2, there exist m common neighbors (of Type 3) of i and j. Thus, the conditional likelihood of j following i is a problem of interest. Theoretically, this amounts to evaluating the transitivity probability in (2.6), but with multiple common neighbors, which is

$$P\left(a_{ji}=1 \middle| \prod_{l=1}^{m} a_{jk_l}a_{k_li}=1,\mathcal{P}\right)=\left\{\frac{\sum_{l=1}^{m}(\sigma_{k_l}^2+\sigma_i^2)^{-1}}{\sigma_i^{-2}+\sum_{l=1}^{m}(\sigma_{k_l}^2+\sigma_i^2)^{-1}}\right\}^{1/2}+O_p\left(\frac{1}{n^2}\right). \tag{2.7}$$

See Appendix B for the technical details. Once again, the transitivity probability in (2.7) can be arbitrarily close to 0, if the common neighbors all have many followers. Note that if i and j share only $m-1$ common neighbors (i.e., $k_1, \ldots, k_{m-1}$), the corresponding transitivity probability becomes

$$P\left(a_{ji}=1 \middle| \prod_{l=1}^{m-1} a_{jk_l}a_{k_li}=1,\mathcal{P}\right)=\left\{\frac{\sum_{l=1}^{m-1}(\sigma_{k_l}^2+\sigma_i^2)^{-1}}{\sigma_i^{-2}+\sum_{l=1}^{m-1}(\sigma_{k_l}^2+\sigma_i^2)^{-1}}\right\}^{1/2}+O_p\left(\frac{1}{n^2}\right). \tag{2.8}$$

We can easily verify that

$$\left(\frac{\sum_{l=1}^{m}(\sigma_{k_l}^2+\sigma_i^2)^{-1}}{\sigma_i^{-2}+\sum_{l=1}^{m}(\sigma_{k_l}^2+\sigma_i^2)^{-1}}\right)^{1/2} \geq \left(\frac{\sum_{l=1}^{m-1}(\sigma_{k_l}^2+\sigma_i^2)^{-1}}{\sigma_i^{-2}+\sum_{l=1}^{m-1}(\sigma_{k_l}^2+\sigma_i^2)^{-1}}\right)^{1/2}.$$

The interpretation confirms our intuition. That is, the more common neighbors two nodes share, the more likely it is that they are connected. In an extreme situation, as the number of shared common neighbors tends to infinity, the probability of the nodes being connected tends to one.

Note that similar conclusions were also obtained in recent studies. For example, Sarkar, Chakrabarti and Moore (2011) provide upper and lower bounds of the conditional link probability $P(a_{ji} = 1 | \prod_{l=1}^{m} a_{ik_l} a_{jk_l} = 1, \mathcal{P})$ for the classical LSM Hoff, Raftery and Handcock (2002). They proved that the probability tends to one as the number of common neighbors tends to infinity. Sarkar, Chakrabarti and Bickel (2015) established the same results for stochastic block models. However, the proofs in both studies are technically challenging because a number of sophisticated high-level inequalities need to be used. In contrast, (2.8) explicitly presents the conditional link probability, which is fairly intuitive and straightforward.

3. Link Prediction

3.1. Probability index

We next demonstrate how the proposed PSLSM can be used in a large-scale social network. Here, we discuss one particular type of application, namely link prediction. Statistically, the problem is one of estimating the likelihood of $a_{ji} = 1$, based on the observed network information. This is of fundamental importance in industry applications (Lü and Zhou (2011)). As a result, many methods have been developed. Define $\Gamma_i^{in} = \{j : a_{ji} = 1\}$ as the set of nodes following i, and $\Gamma_i^{out} = \{j : a_{ij} = 1\}$ as the set of nodes that i follows. Then, for two arbitrary nodes i and j, the set $\Gamma_i^{in} \bigcap \Gamma_j^{out}$ contains the Type 3 common neighbors between i and j. Its total number is given by $|\Gamma_i^{in} \bigcap \Gamma_j^{out}|$, which is referred to as the common neighbor index (CNI) in directed networks (Yu and Wang (2014)). Intuitively, two nodes i and j are more likely to be connected, if they share a great number of common connected nodes. Thus, the simplest way to conduct a link prediction for a_{ji} is to check whether CNI is sufficiently large (Kossinets (2006); Yu and Wang (2014); Lü and Zhou (2011)). To further improve the CNI, a number of variants have been developed. For example, Chowdhury (2010) penalized $|\Gamma_i^{in} \bigcap \Gamma_j^{out}|$ using the degrees of each node, leading to the Salton Index (SI). See Lü and Zhou (2011) for a comprehensive discussion and Table 1 for a succinct summary of similar indices.

The indices listed in Table 1 are constructed based on two sources of information: (1) the number of nodes in the neighborhood set $\Gamma_i^{in} \bigcap \Gamma_j^{out}$; and (2) the degrees d_i, d_j, or d_k^* ($k \in \Gamma_i^{in} \bigcap \Gamma_j^{out}$). Past experience suggests that the indices given in Table 1 are closely related to whether $a_{ji} = 1$. However, none of the indices give a direct estimate of the conditional link probability (Sarkar, Chakrabarti and Moore (2011)). In contrast, a comprehensive estimate can be

Table 1. Local similarity indices for a symmetric network.

Method	Index
Common Neighbors Index (CNI)	$\lvert\Gamma_i^{in}\bigcap\Gamma_j^{out}\rvert$
Salton Index (SI)	$\lvert\Gamma_i^{in}\bigcap\Gamma_j^{out}\rvert/\sqrt{d_i\times d_j}$
Sørensen Index (SOI)	$2\lvert\Gamma_i^{in}\bigcap\Gamma_j^{out}\rvert/(d_i+d_j)$
Hub Promoted Index (HPI)	$\lvert\Gamma_i^{in}\bigcap\Gamma_j^{out}\rvert/\min\{d_i,d_j\}$
Hub Depressed Index (HDI)	$\lvert\Gamma_i^{in}\bigcap\Gamma_j^{out}\rvert/\max\{d_i,d_j\}$
Leicht-Holme-Newman Index (LHNI)	$\lvert\Gamma_i^{in}\bigcap\Gamma_j^{out}\rvert/(d_i\times d_j)$
Adamic-Adar Index (AAI)	$\sum_{k\in\Gamma_i^{in}\bigcap\Gamma_j^{out}}(\log d_k^*)^{-1}$
Resource Allocation Index (RAI)	$\sum_{k\in\Gamma_i^{in}\bigcap\Gamma_j^{out}} d_k^{*-1}$

obtained under the PSLSM using (2.7). More specifically, a probability index (PI) can be constructed as

$$\mathrm{PI}(j,i)=\frac{\sum_{l=1}^m(\widehat{\sigma}_{k_l}^2+\widehat{\sigma}_i^2)^{-1}}{\widehat{\sigma}_i^{-2}+\sum_{l=1}^m(\widehat{\sigma}_{k_l}^2+\widehat{\sigma}_i^2)^{-1}}, \tag{3.1}$$

where $\widehat{\sigma}_i=\sqrt{2}d_i$ and is an asymptotically unbiased estimate of σ_i, by (2.4). In contrast to the aforementioned methods, $\mathrm{PI}(j,i)$ is a direct measure of the conditional link probability. To implement the proposed PI method, three remarks are required.

Remark 1. To make a prediction, for each user j, other users i ($i\neq j$) are reordered based on decreasing PI values. Thus, a higher PI value indicates a larger probability of a link from j to i. In practice, a network platform only recommends the top L (for example, 3–5) users with the highest probability for user j.

Remark 2. If j and i have no connected nodes in common, then $\Gamma_i^{in}\bigcap\Gamma_j^{out}=\emptyset$, which means $\mathrm{PI}(j,i)=\widehat{\sigma}_i^2$. Thus, those i who have no connected nodes in common with j have the same PI values. If only $\Gamma_k^{in}\bigcap\Gamma_j^{out}\neq\emptyset$, we have $\mathrm{PI}(j,k)>\mathrm{PI}(j,i)$.

Remark 3. With regard to the computational complexity of PI, we need to estimate σ_i. Because $\widehat{\sigma}_i=\sqrt{2}d_i$, the cost of calculating PI is linear in the sample size n, that is, $O(n)$. Therefore, PI has computational complexity of the same order as the local similarity indices defined in Table 1.

3.2. Simulation studies

To demonstrate the finite-sample performance of the proposed PI method, we present three examples.

Example 1. (PSLSM with Power-law Popularity) In this example, the adjacency matrix is generated according to the PSLSM defined in (2.3), as follows. First, the popularity parameters $\sigma_i (1 \leq i \leq n)$ are independently and identically drawn from a power-law-type distribution $P(\sigma_i = k) = ck^{-\alpha}$, where c is a normalizing constant. The exponent parameter $\alpha \in \{1.7, 1.5, 1.3\}$. Second, for each node i, we generate the latent space variable Z_i from the standard normal distribution. Third, we generate the adjacency matrix A from (2.3).

Example 2. (PSLSM with Hub Nodes) We repeat the steps in Example 1 to obtain an initial adjacency matrix A, except that σ_i $(1 \leq i \leq n)$ is independently and identically drawn from a standard normal distribution. Then, we randomly select H nodes ($H \in \{0, 100, 200\}$) to be super popular nodes, or hub nodes (Newman, Barabási and Watts (2006)). A total of 600 nodes are randomly selected to follow each hub node.

Example 3. (Power-law Degree) We follow Clauset, Shalizi and Newman (2009) to simulate A using power-law distributed degrees. First, in-degree d_i is generated according to the discrete power-law distribution, that is, $P(d_i = k) = ck^{-\alpha}$, as in Example 1. The exponent parameter $\alpha \in \{1.7, 1.5, 1.3\}$. Here, a smaller α indicates a heavier tail of the distribution. Next, for the ith node, we randomly select d_i nodes as its followers, with $a_{ji} = 1$.

We compare the performance of PI with the local similarity indices listed in Table 1. The prediction accuracy is measured by the area under the receiver operating characteristic curve, referred to as AUC (Lü and Zhou (2011)). The network size n is fixed at 1,000. For a reliable evaluation, the experiment is randomly replicated $M = 1{,}000$ times. The value of $\widehat{\mathrm{AUC}}^{(m)}$ is calculated based on 10-fold cross-validation in the mth $(1 \leq m \leq M)$ replication. Here, 10-fold cross-validation means we randomly split the data into 10 parts. One part is used as a testing set, and the remainder are used as training sets. Then, $\widehat{\mathrm{AUC}}_k^{(m)}$ can be computed using all PI values and a_{ij} in the k-th fold. Thus, $\widehat{\mathrm{AUC}}^{(m)} = 10^{-1} \sum_{k=1}^{10} \widehat{\mathrm{AUC}}_k^{(m)}$ is the average of the AUC results from the testing set for all 10 parts. We report both $\widehat{\mathrm{AUC}} = M^{-1} \sum_m \widehat{\mathrm{AUC}}^{(m)}$ and the Monte Carlo standard deviation of $\widehat{\mathrm{AUC}}^{(m)}$, that is, $\mathrm{SE}^* = \{M^{-1} \sum_m (\widehat{\mathrm{AUC}}^{(m)} - \widehat{\mathrm{AUC}})^2\}^{1/2}$. Detailed results for these three examples are shown in Tables 2–4.

Based on Tables 2, 3, and 4, we make the following observations. First, in terms of $\widehat{\mathrm{AUC}}$, PI has a similar link prediction capability to the other indices when $\alpha = 1.7$ in Example 1 and $H = 0$ in Example 2. Second, in all other cases, PI outperforms the other methods, because the latter do not provide direct estimates

Table 2. Detailed results for the indices in Example 1.

Method	PI	RAI	CNI	AAI	HDI	HPI	LHNI	SI	SOI
$\alpha = 1.7$	0.750	0.747	0.749	0.749	0.749	0.749	0.746	0.749	0.749
	(0.015)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
$\alpha = 1.5$	0.795	0.772	0.789	0.788	0.788	0.789	0.784	0.789	0.788
	(0.010)	(0.011)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
$\alpha = 1.3$	0.845	0.799	0.817	0.817	0.814	0.817	0.808	0.818	0.815
	(0.007)	(0.009)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)

Table 3. Detailed results for the indices in Example 2.

Method	PI	RAI	CNI	AAI	HDI	HPI	LHNI	SI	SOI
$H = 0$	0.862	0.862	0.862	0.862	0.862	0.862	0.862	0.862	0.862
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
$H = 100$	0.832	0.811	0.811	0.811	0.812	0.808	0.785	0.812	0.812
	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.014)	(0.005)	(0.005)
$H = 200$	0.822	0.787	0.787	0.787	0.787	0.773	0.761	0.788	0.787
	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)

Table 4. Detailed results for the indices in Example 3.

Method	PI	RAI	CNI	AAI	HDI	HPI	LHNI	SI	SOI
$\alpha = 1.7$	0.714	0.699	0.707	0.706	0.702	0.709	0.689	0.703	0.702
	(0.034)	(0.031)	(0.032)	(0.033)	(0.031)	(0.031)	(0.033)	(0.031)	(0.032)
$\alpha = 1.5$	0.838	0.799	0.815	0.815	0.797	0.822	0.786	0.477	0.477
	(0.017)	(0.015)	(0.015)	(0.015)	(0.014)	(0.015)	(0.015)	(0.018)	(0.018)
$\alpha = 1.3$	0.9107	0.843	0.843	0.843	0.810	0.853	0.787	0.829	0.811
	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)	(0.006)	(0.005)	(0.005)

of the link probability. Last, the $\widehat{\mathrm{AUC}}$ of PI increases as α decreases, or as the number of popular nodes H increases. Therefore, PI is quite suitable for a degree distribution with a heavier tail.

3.3. A real-data example

We next illustrate the performance of the proposed method using a real-data analysis. The data are collected from Sina Weibo (`www.weibo.com`), which has more than 500 million registered users and is the largest social media platform, similar to Twitter in China. For this example, we start with the four official Weibo accounts of four well-respected MBA programs in China. However, owing to a constraint imposed by the Sina Weibo API, only $n = 8,591$ active followers are obtained. Their follower–followee relationships are recorded by the adjacency matrix $A = (a_{ij}) \in \mathbb{R}^{n\times n}$.

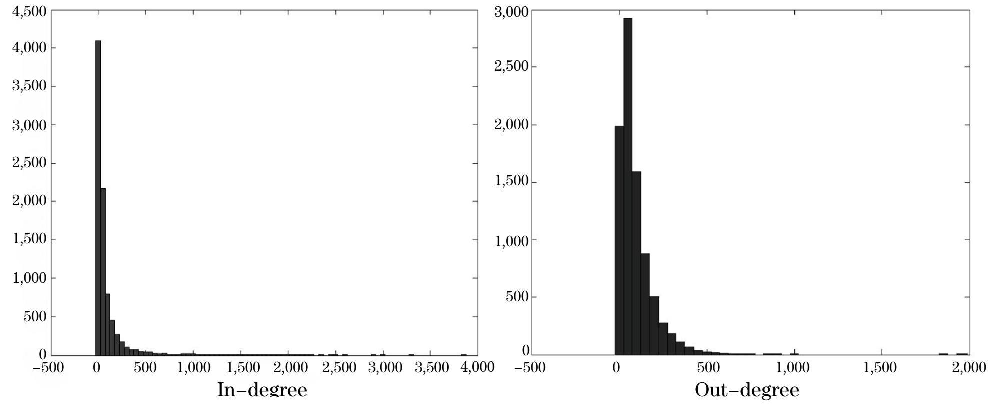

Figure 3. Histograms of the degrees in the real-data example. The left panel is the histogram of in-degrees and the right panel is that of out-degrees.

To better understand the network structure, we provide histograms for the in- and out-degrees in Figure 3. We find that both distributions are highly skewed. Comparatively speaking, the distribution of the in-degree has a far heavier tail. Next, we evaluate the reciprocity property. To this end, we calculate the empirical link probability (or network density) as $\widehat{P}(a_{ij} = 1) = (\sum a_{ij})/\{n(n-1)\} = 0.0107$, which is fairly close to zero. However, the conditional probability is $\widehat{P}(a_{ji} = 1|a_{ij} = 1) = \sum a_{ij}a_{ji}/\sum a_{ij} = 0.1892$, which is much larger and well bounded away from zero. This confirms that the reciprocity property does hold in this real data set.

Next, we examine the transitivity properties. We consider the Type 1 transitivity in Figure 1 first. This probability is estimated by $\widehat{P}_1 = \sum a_{ik}a_{jk}a_{ji}(1 - a_{ki})(1-a_{kj})/\sum a_{ik}a_{jk}(1-a_{ki})(1-a_{kj}) = 0.0289$. Similar calculations can be performed for the other transitivity types. This leads to $\widehat{P}_2 = 0.0219$, $\widehat{P}_3 = 0.2385$, $\widehat{P}_4 = 0.1929$, $\widehat{P}_5 = 0.1979$, $\widehat{P}_6 = 0.0954$, $\widehat{P}_7 = 0.0652$, $\widehat{P}_8 = 0.2238$, and $\widehat{P}_9 = 0.2495$. Thus, Type 3 has an excellent transitivity probability, $\widehat{P}_3 = 0.2385$, and ranks as the second highest of these probabilities. This implies that link predictions based on Type 3 transitivity should be useful. Even though the Type 9 transitivity probability is slightly larger than that of Type 3 (i.e., $\widehat{P}_3 = 0.2385$ vs. $\widehat{P}_9 = 0.2495$), the total number of triplets that qualify is over ten times smaller for Type 9: $\sum a_{ik}a_{ki}a_{jk}a_{kj} = 37,644,336$ for Type 9, and $\sum a_{ik}a_{jk} = 469,566,693$ for Type 3. As a result, focusing on Type 3 transitivity in (3.1) is empirically justified. Lastly, we compare the performance of PI with that of the competing methods listed in Table 1. The average AUCs based on 10-fold cross-validation

Table 5. AUC values of the various methods for the sina weibo data set.

PI	RAI	CNI	AAI	HDI	HPI	LHNI	SI	SOI
0.899	0.871	0.888	0.886	0.836	0.868	0.613	0.861	0.847

are summarized in Table 5. Once again, PI performs best, with an associated AUC value of 0.899. The second-best performance was achieved by CNI with an AUC of 0.888.

4. Conclusion

To conclude the paper, we discuss a number of interesting topics for future study. First, individual characteristics may also influence the network structure, and so should be taken into account. Second, the PSLSM provides a flexible framework model for large-scale network structures. However, extending the model to include other network structures, such as community structures, would be worthwhile.

Acknowledgment

We would like to thank the editors and the two anonymous reviewers for their consideration and constructive comments, which greatly improved the quality of the paper. Xiangyu Chang's research was supported by the National Natural Science Foundation of China (Project No. 11401462, 11771012, 61502342, 61603162) and the China Postdoctoral Science Foundation (Project No. 2015M582630). Huang's research was supported in part by National Natural Science Foundation of China (Grant No.11701560), the Beijing Municipal Social Science Foundation (Grant No.17GLC051), and the Center for Applied Statistics, School of Statistics, Renmin University of China. Hansheng Wang's research is partially supported by National Natural Science Foundation of China (No. 71532001, No. 11525101, No. 71332006). It is also supported in part by China's National Key Research Special Program (No. 2016YFC0207700). The corresponding author is Dr. Danyang Huang.

Appendix

Type 1 transitivity probability

Before the calculation, we define the following matrices,

$$B_{ij} = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, B_{jk} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{pmatrix} \text{ and } B_{ik} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix}.$$

To calculate the probability of Type 1 transitivity amounts to compute,

$$P_1 := P(a_{ji} = 1 | a_{ik}a_{jk} = 1, \gamma_k, \gamma_i) = \frac{P(a_{ji}a_{ik}a_{jk} = 1|\gamma_k, \gamma_i)}{P(a_{ik}a_{jk} = 1|\gamma_k)}. \tag{A.1}$$

First, by Eq. (2.1), we compute the denominator of the right hand side in (A.1),

$$\begin{aligned} P(a_{ik}a_{jk} = 1|\gamma_k, Z_i, Z_j, Z_k) &= \exp\left[-\left\{\frac{(Z_i - Z_k)^2}{2\gamma_k^2} + \frac{(Z_j - Z_k)^2}{2\gamma_k^2}\right\}\right] \\ &= \exp\left\{-\frac{1}{2}X^\top(\gamma_k^{-2}B_{ik} + \gamma_k^{-2}B_{jk})X\right\}, \end{aligned} \tag{A.2}$$

where $X = (Z_i, Z_j, Z_k)^\top$ satisfies a standard normal distribution $\mathcal{N}(0, I)$, I is an identical matrix. Define $\Sigma_1 = \gamma_k^{-2}(B_{ik} + B_{jk}) + I$, by integration over X, $P(a_{ik}a_{jk} = 1|\gamma_k) = (\det \Sigma_1)^{-1/2} = \gamma_k^2(\gamma_k^4 + 4\gamma_k^2 + 3)^{-1/2}$,

Second, we consider the numerator of the right hand side in (A.1). Similarly with (A.2), $P(a_{ji}a_{ik}a_{jk} = 1|\gamma_i, \gamma_k, Z_i, Z_j, Z_k) = \exp\{-0.5X^\top(\gamma_k^{-2}B_{ik} + \gamma_k^{-2}B_{jk} + \gamma_i^{-2}B_{ij})X\}$. Thus, define $\Sigma_1' = \Sigma_1 + \gamma_i^{-2}B_{ij}$, we have $P(a_{ji}a_{ik}a_{jk} = 1|\gamma_k, \gamma_i) = (\det \Sigma_1')^{-1/2} = \gamma_k^2\gamma_i(\gamma_k^4\gamma_i^2 + 2\gamma_k^4 + 4\gamma_k^2\gamma_i^2 + 6\gamma_k^2 + 3\gamma_i^2)^{-1/2}$.

Last, Type 1 transitivity is $P_1 = (\gamma_k^2\gamma_i^2 + \gamma_i^2)^{1/2}(\gamma_k^2\gamma_i^2 + \gamma_i^2 + 2\gamma_k^2)^{-1/2}$.

Type 2 transitivity probability

To calculate the probability of Type 2 transitivity amounts to compute,

$$P_2 := P(a_{ji} = 1|a_{ik}a_{kj} = 1, \gamma_k, \gamma_j, \gamma_i) = \frac{P(a_{ji}a_{ik}a_{kj} = 1|\gamma_k, \gamma_j, \gamma_i)}{P(a_{ik}a_{kj} = 1|\gamma_k, \gamma_j)}. \tag{A.3}$$

First, based on Eq. (2.1), the denominator of the right hand side in (A.3) is, $P(a_{ik}a_{kj} = 1|\gamma_k, \gamma_j, Z_i, Z_j, Z_k) = \exp\{-0.5X^\top(\gamma_k^{-2}B_{ik} + \gamma_j^{-2}B_{jk})X\}$. Thus, by integration over X, $P(a_{ik}a_{kj} = 1|\gamma_k, \gamma_j) = (\det \Sigma_2)^{-1/2} = (\gamma_k^2\gamma_j^2)^{1/2}$ $(3 + 2\gamma_k^2 + 2\gamma_j^2 + \ \gamma_k^2\gamma_j^2)^{-1/2}$. where $\Sigma_2 = \gamma_k^{-2}B_{ik} + \gamma_j^{-2}B_{jk} + I$. Second, by (2.1), $P(a_{ji}a_{ik}a_{kj} = 1|\gamma_j, \gamma_k, Z_i, Z_j, Z_k) = \exp\{-0.5X^\top(\gamma_i^{-2}B_{ij} + \gamma_k^{-2}B_{ik} + \gamma_j^{-2}B_{jk})X\}$.

Thus we have,

$$P(a_{ji}a_{ik}a_{kj}=1|\gamma_k,\gamma_j,\gamma_i)=(\det\Sigma_2')^{-1/2}$$
$$=\gamma_k\gamma_j\gamma_i(\gamma_k^2\gamma_j^2\gamma_i^2+2\gamma_i^2\gamma_j^2+2\gamma_j^2\gamma_k^2+2\gamma_k^2\gamma_i^2+3\gamma_i^2+3\gamma_j^2+3\gamma_k^2)^{-1/2},$$

where $\Sigma_2'=\Sigma_2+\gamma_i^{-2}B_{ij}$. Last, we have the Type 2 transitivity probability is,

$$P_2=\left(\frac{\gamma_k^2\gamma_j^2\gamma_i^2+2\gamma_j^2\gamma_i^2+2\gamma_k^2\gamma_i^2+3\gamma_i^2}{\gamma_k^2\gamma_j^2\gamma_i^2+2\gamma_i^2\gamma_j^2+2\gamma_j^2\gamma_k^2+2\gamma_k^2\gamma_i^2+3\gamma_i^2+3\gamma_j^2+3\gamma_k^2}\right)^{1/2}.$$

The other types of transitivity probability could be derived in the similar steps, we only list the key steps as follows.

Type 3 transitivity probability

Consider the probability $P_3:=P(a_{ji}=1|a_{jk}a_{ki}=1,\gamma_k,\gamma_i)$. First, by Eq. (2.1), $P(a_{jk}a_{ki}=1|\gamma_k,\gamma_i)=(\det\Sigma_3)^{-1/2}$, where, $\Sigma_3=\gamma_k^{-2}B_{jk}+\gamma_i^{-2}B_{ik}+I$. Second, $P(a_{ji}a_{jk}a_{ki}=1|\gamma_k,\gamma_i)=(\det\Sigma_3')^{-1/2}$, where $\Sigma_3'=\Sigma_3+\gamma_i^{-2}B_{ij}$. As a result, we can get the Type 3 transitivity probability as $P_3=(\gamma_i^4\gamma_k^2+2\gamma_i^4+2\gamma_i^2\gamma_k^2+3\gamma_i^2)^{1/2}\ (\gamma_i^4\gamma_k^2+2\gamma_i^4+4\gamma_i^2\gamma_k^2+6\gamma_i^2+3\gamma_k^2)^{-1/2}$.

Type 4 transitivity probability

Define $P_4:=P_4(a_{ji}=1|a_{kj}a_{ki}=1,\gamma_j,\gamma_i)$. Based on Eq. (2.1), $P(a_{kj}a_{ki}=1|\gamma_j,\gamma_i)=(\det\Sigma_4)^{-1/2}$, where $\Sigma_4=\gamma_j^{-2}B_{jk}+\gamma_i^{-2}B_{ik}+I$. Similarly, we know $P(a_{ji}a_{kj}a_{ki}=1|\gamma_j,\gamma_i)=(\det\Sigma_4')^{-1/2}$, where $\Sigma_4'=\Sigma_4+\gamma_i^{-2}B_{ij}$. Then, the Type 4 transitivity probability is $P_4=(\gamma_i^4\gamma_j^2+2\gamma_i^4+2\gamma_i^2\gamma_j^2+3\gamma_i^2)^{1/2}\ (\gamma_i^4\gamma_j^2+2\gamma_i^4+4\gamma_i^2\gamma_j^2+6\gamma_i^2+3\gamma_j^2)^{-1/2}$.

Type 5 transitivity probability

Define $P_5:=P(a_{ji}=1|a_{jk}a_{ik}a_{ki}=1,\gamma_k,\gamma_i)$. Thus we have, $P(a_{jk}a_{ik}a_{ki}=1|\gamma_k,\gamma_i)=(\det\Sigma_5)^{-1/2}$, where $\Sigma_5=\gamma_k^{-2}B_{jk}+\gamma_k^{-2}B_{ik}+\gamma_i^{-2}B_{ik}+I$. In the meanwhile, $P(a_{ji}a_{jk}a_{ki}a_{ki}=1|\gamma_j,\gamma_i)=(\det\Sigma_5')^{-1/2}$, where $\Sigma_5'=\Sigma_5+\gamma_i^{-2}B_{ij}$. Finally, we can get

$$P_5=\left(\frac{\gamma_i^4\gamma_k^4+4\gamma_i^4\gamma_k^2+3\gamma_i^4+2\gamma_i^2\gamma_k^4+3\gamma_i^2\gamma_k^2}{\gamma_i^4\gamma_k^4+4\gamma_i^4\gamma_k^2+4\gamma_i^2\gamma_k^4+3\gamma_i^4+3\gamma_k^4+9\gamma_i^2\gamma_k^2}\right)^{1/2}.$$

Type 6 transitivity probability

Define $P_6:=P(a_{ji}{=}1|a_{kj}a_{ik}a_{ki}{=}1,\gamma_k,\gamma_i,\gamma_j)$ Based on Eq. (2.1), $P(a_{kj}a_{ik}a_{ki}=1|\gamma_k,\gamma_i,\gamma_j)=(\det\Sigma_6)^{-1/2}$, where $\Sigma_6=\gamma_j^{-2}B_{jk}+\gamma_k^{-2}B_{ik}+\gamma_i^{-2}B_{ik}+I$. Furthermore, $P(a_{ji}a_{kj}a_{ik}a_{ki}=1|\gamma_k,\gamma_j,\gamma_i)=(\det\Sigma_6')^{-1/2}$, where $\Sigma_6'=\Sigma_6+\gamma_i^{-2}B_{ij}$.

Thus,

$$P_6 = \left(\frac{\gamma_i^4\gamma_k^2\gamma_j^2 + 2\gamma_i^4\gamma_k^2 + 2\gamma_i^4\gamma_j^2 + 2\gamma_i^2\gamma_k^2\gamma_j^2 + 3\gamma_i^4 + 3\gamma_i^2\gamma_k^2}{\gamma_i^4\gamma_k^2\gamma_j^2 + 2\gamma_i^4\gamma_k^2 + 2\gamma_i^4\gamma_j^2 + 3\gamma_i^4 + 4\gamma_i^2\gamma_k^2\gamma_j^2 + 6\gamma_i^2\gamma_k^2 + 3\gamma_i^2\gamma_j^2 + 3\gamma_j^2\gamma_k^2} \right)^{1/2}.$$

Type 7 transitivity probability

Define $P_7 := P(a_{ji} = 1 | a_{jk}a_{kj}a_{ik} = 1, \gamma_k, \gamma_j, \gamma_i)$. Based on Eq. (2.1), $P(a_{jk}a_{kj} a_{ik} = 1 | \gamma_k, \gamma_j) = (\det \Sigma_7)^{-1/2}$, where $\Sigma_7 = \gamma_k^{-2} B_{jk} + \gamma_j^{-2} B_{jk} + \gamma_k^{-2} B_{ik} + I$; and $P(a_{ji}a_{jk}a_{kj}a_{ik} = 1 | \gamma_j, \gamma_k, \gamma_i) = (\det \Sigma_7')^{-1/2}$, where $\Sigma_7' = \Sigma_7 + \gamma_i^{-2} B_{ij}$. Thus,

$$P_7 = \left(\frac{\gamma_i^2\gamma_k^4\gamma_j^2 + 2\gamma_i^2\gamma_k^4 + 4\gamma_i^2\gamma_j^2\gamma_k^2 + 3\gamma_i^2\gamma_k^2 + 3\gamma_i^2\gamma_j^2}{\gamma_i^2\gamma_k^4\gamma_j^2 + 2\gamma_i^2\gamma_k^4 + 4\gamma_i^2\gamma_j^2\gamma_k^2 + 3\gamma_i^2\gamma_k^2 + 3\gamma_i^2\gamma_j^2 + 3\gamma_k^4\gamma_j^2 + 3\gamma_k^4 + 3\gamma_k^2\gamma_j^2} \right)^{1/2}.$$

Type 8 transitivity probability

Define $P_8 := P(a_{ji} = 1 | a_{jk}a_{kj}a_{ki} = 1, \gamma_k, \gamma_j, \gamma_i)$. Based on Eq. (2.1), $P(a_{jk}a_{kj} a_{ki} = 1 | \gamma_k, \gamma_j, \gamma_i) = (\det \Sigma_8)^{-1/2}$, where $\Sigma_8 = \gamma_k^{-2} B_{jk} + \gamma_j^{-2} B_{jk} + \gamma_i^{-2} B_{ik} + I$; and we know that $P(a_{ji}a_{jk}a_{kj}a_{ki} = 1 | \gamma_j, \gamma_k, \gamma_i) = (\det \Sigma_8')^{-1/2}$, where $\Sigma_8' = \Sigma_8 + \gamma_i^{-2} B_{ij}$. Thus we can get

$$P_8 = \left(\frac{\gamma_i^4\gamma_k^2\gamma_j^2 + 2\gamma_i^4\gamma_k^2 + 2\gamma_i^4\gamma_j^2 + 2\gamma_i^2\gamma_k^2\gamma_j^2 + 3\gamma_i^2\gamma_j^2 + 3\gamma_i^2\gamma_k^2}{\gamma_i^4\gamma_k^2\gamma_j^2 + 2\gamma_i^4\gamma_k^2 + 2\gamma_i^4\gamma_j^2 + 4\gamma_i^2\gamma_k^2\gamma_j^2 + 6\gamma_i^2\gamma_k^2 + 6\gamma_i^2\gamma_j^2 + 3\gamma_j^2\gamma_k^2} \right)^{1/2}.$$

Type 9 transitivity probability

Define $P_9 := P(a_{ji} = 1 | a_{jk}a_{kj}a_{ik}a_{ki} = 1, \gamma_k, \gamma_j, \gamma_i)$. Based on Eq. (2.1), $P(a_{jk}a_{kj}a_{ik}a_{ki} = 1 | \gamma_k, \gamma_j, \gamma_i) = (\det \Sigma_9)^{-1/2}$, where $\Sigma_9 = \gamma_k^{-2} B_{jk} + \gamma_j^{-2} B_{jk} + \gamma_i^{-2} B_{ik} + \gamma_k^{-2} B_{ik} + I$. And $P(a_{ji}a_{jk}a_{kj}a_{ik}a_{ki} = 1 | \gamma_i, \gamma_j, \gamma_k) = (\det \Sigma_9')^{-1/2}$, where $\Sigma_9' = \Sigma_9 + \gamma_i^{-2} B_{ij}$. Define $\Gamma_{ijk} = \gamma_i^4\gamma_k^4\gamma_j^2 + 2\gamma_i^4\gamma_k^4 + 4\gamma_i^4\gamma_k^2\gamma_j^2 + 3\gamma_i^4\gamma_k^2 + 3\gamma_i^4\gamma_j^2$. Thus,

$$P_9 = \left(\frac{\Gamma_{ijk} + 2\gamma_i^2\gamma_k^4\gamma_j^2 + 3\gamma_i^2\gamma_k^4 + 3\gamma_i^2\gamma_j^2\gamma_k^2}{\Gamma_{ijk} + 4\gamma_i^2\gamma_k^4\gamma_j^2 + 6\gamma_i^2\gamma_k^4 + 9\gamma_i^2\gamma_j^2\gamma_k^2 + 3\gamma_k^4\gamma_j^2} \right)^{1/2}.$$

Technical details for (2.7)

The common neighbors probability could be written as,

$$P\left(a_{ji=1} \middle| \prod_{l=1}^{m} a_{jk_l} a_{k_l i} = 1, \mathcal{P} \right) = \frac{P(a_{ji} \prod_{l=1}^{m} a_{jk_l} a_{k_l i} = 1 | \mathcal{P})}{P(\prod_{l=1}^{m} a_{jk_l} a_{k_l i} = 1 | \mathcal{P})} \qquad \text{(A.4)}$$

First, we calculate the denominator of (A.4) as $P(\prod_{l=1}^m a_{jk_l} a_{k_l i} = 1|\mathcal{P}) =$

$$\frac{1}{(2\pi)^{m/2+1}} \int_{\mathbb{R}^{m+2}} \exp\left(-\frac{1}{2} X \Sigma_m X^\top\right) dX = (\det \Sigma_m)^{-1/2}$$

where $X = (Z_i, Z_j, Z_{k_1}, \ldots, Z_{k_m})^\top$, $\Sigma_m = B_m + I_m$. Further define $\Sigma_m = [M, N; N^\top, O]$, where,

$$M = \begin{pmatrix} m/\gamma_i^2 + 1 & 0 \\ 0 & \sum_{l=1}^m 1/\gamma_{k_l}^2 + 1 \end{pmatrix}_{2\times 2}, N = \begin{pmatrix} -1/\gamma_i^2, \ldots, -1/\gamma_i^2 \\ -1/\gamma_{k_1}^2, \ldots, -1/\gamma_{k_m}^2 \end{pmatrix}_{2\times m}.$$

$$O = \begin{pmatrix} 1/\gamma_{k_1}^2 + 1/\gamma_i^2 + 1 & 0 & \cdots \\ 0 & \ddots & 0 \\ 0 & 0 & 1/\gamma_{k_m}^2 + 1/\gamma_i^2 + 1 \end{pmatrix}_{m\times m},$$

So we have $\det(\Sigma_m) = \det(M - NO^{-1}N^\top)\det(O)$.

Second, we compute $P(a_{ji} \prod_{l=1}^m a_{jk_l} a_{k_l i} = 1|\mathcal{P}) = (\det \Sigma_m')^{-1/2}$, where $\Sigma_m' = B_m' + I_m$, $B_m' = B_m + L$, and

$$L = \begin{pmatrix} 1/\gamma_i^2 & -1/\gamma_i^2, 0, \ldots \\ -1/\gamma_i^2 & 1/\gamma_i^2, 0, \ldots \\ 0 & \ddots \end{pmatrix}_{(m+2)\times(m+2)}.$$

Thus, $\Sigma_m' = \begin{pmatrix} M' & N \\ N^\top & O \end{pmatrix}$ and $M' = \begin{pmatrix} (m+1)/\gamma_i^2 + 1 & -1/\gamma_i^2 \\ -1/\gamma_i^2 & \sum_{l=1}^m 1/\gamma_{k_l}^2 + 1 + 1/\gamma_i^2 \end{pmatrix}$.

Finally, the common neighbors probability is,

$$P\left(a_{ji}=1 \middle| \prod_{l=1}^m a_{ik_l} a_{jk_l}=1, \mathcal{P}\right) = \left\{\frac{(\det \Sigma_m)}{(\det \Sigma_m')}\right\}^{-1/2} = \left\{\frac{\det(M - NO^{-1}N^\top)}{\det(M' - NO^{-1}N^\top)}\right\}^{-1/2}.$$

Define $b = \sum_{l=1}^m \gamma_{k_l}^2 (\gamma_{k_l}^2 + \gamma_i^2 + \gamma_{k_l}^2 \gamma_i^2)^{-1/2}$, $c = \sum_{l=1}^m \gamma_i^2 (\gamma_{k_l}^2 + \gamma_i^2 + \gamma_{k_l}^2 \gamma_i^2)^{-1/2}$ and $d = \sum_{l=1}^m (\gamma_{k_l}^2 + \gamma_i^2 + \gamma_{k_l}^2 \gamma_i^2)^{-1/2}$ for simplicity, then $\det(M - NO^{-1}N^\top) = (d+c+1)(d+b+1) - d^2$ and $\det(M' - NO^{-1}N^\top) = (d+c+1+1/\gamma_i^2)(d+b+1+1/\gamma_i^2) - (d+1/\gamma_i^2)^2$. Thus, by $\sigma_i = n\gamma_i$, we have,

$$P\left(a_{ji} = 1 \middle| \prod_{l=1}^m a_{jk_l} a_{k_l i} = 1, \mathcal{P}\right) = \left\{\frac{\sum_{l=1}^m (\sigma_{k_l}^2 + \sigma_i^2)^{-1}}{1/\sigma_i^2 + \sum_{l=1}^m (\sigma_{k_l}^2 + \sigma_i^2)^{-1}}\right\}^{1/2} + O_p\left(\frac{1}{n^2}\right).$$

References

Adamic, L. and Huberman, B. (2000). Power-law distribution of the world wide web. *Science* **287**, 2115–2115.

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

Austin, A., Linkletter, C. and Wu, Z. (2013). Covariate-defined latent space random effects model. *Social Networks* **35**, 338–346.

Barabási, A. and Albert, R. (1999). Emergence of scaling in random networks. *science* **286**, 509–512.

Bickel, P., Choi, D., Chang, X. and Zhang, H. (2013). Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *The Annals of Statistics* **41**, 1922–1943.

Bickel, P. J. and Chen, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proceedings of the National Academy of Sciences* **106**, 21068–21073.

Choi, D., Wolfe, P. and Airoldi, E. (2012). Stochastic blockmodels with a growing number of classes. *Biometrika* **99**, 273–284.

Chowdhury, G. (2010). Introduction to modern information retrieval. *Facet publishing*.

Clauset, A., Shalizi, C. R. and Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review* **51**, 661–703.

Daniel, S. and Chen, Y. (2015). Latent space models for dynamic networks. *Journal of the American Statistical Association* **110**, 1646–1657.

Erdős, P. and Rényi, A. (1960). On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* **5**, 17–61.

Faust, K. (1988). Comparison of methods for positional analysis: Structural and general equivalences. *Social Networks* **10**, 313–341.

Frank, O. and Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association* **81**, 832–842.

Gao, C., Lu, Y. and Zhou, H. H. (2015). Rate-optimal graphon estimation. *The Annals of Statistics* **43**, 2624–2652.

Hoff, P. (2003). Random effects models for network data in dynamic social network modeling and analysis: Workshop summary and papers *The National Academies Press*, 303–312.

Hoff, P., Raftery, A. and Handcock, M. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association* **97**, 1090–1098.

Holland, P. W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association* **76**, 33–5.

Huang, D., Yin, J., Shi, T. and Wang, H. (2015). A Statistical Model for Social Network Labeling. *Journal of Business & Economic Statistics* **34**, 368–374.

Hunter, D. R. and Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics* **15**, 565–583.

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M. and Morris, M. (2008). Ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software* **24**, nipha54860.

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

Kim, M. and Leskovec, J. (2012). Multiplicative attribute graph model of real-world networks. *Internet Mathematics* **8**, 113–160.

Knoke, D. and Yang, S. (2008). *Social Network Analysis. Sage: London.*

Kossinets, G. (2006). Effects of missing data in social networks. *Social networks* **28**, 247–268.

Krivitsky, P. N., Handcock, M. S., Raftery, A. E. and Hoff, P. D. (2009). Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social Networks* **31**, 204–213.

Liben-Nowell, D., Novak, J., Kumar, R., Raghavan, P. and Tomkins, A. (2005). Geographic routing in social networks. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 11623–11628.

Lü, L. and Zhou, T. (2011). Link prediction in complex networks: A survey. *Physica A: Statistical Mechanics and its Applications* **390**, 1150–1170.

Newman, M. (2010). Networks: An introduction. *Oxford University Press, Inc.*

Newman, M., Barabási, A. L. and Watts, D. J. (2006). The structure and dynamics of networks. *Princeton University Press.*

Nowicki, K. and Snijders, T. A. B. (2001). Estimation and prediction for stochastic block structures. *Journal of the American Statistical Association* **96**, 1077–1087.

Olhede, S. C. and Wolfe, P. J. (2014). Network histograms and universality of blockmodel approximation. *Proceedings of the National Academy of Sciences* **111**, 14722–14727.

Sarkar, P., Chakrabarti, D. and Bickel, P. (2015). The consistency of common neighbors for link prediction in stochastic blockmodels. In *Advances in Neural Information Processing Systems*, 3016–3024.

Sarkar, P., Chakrabarti, D. and Moore, A. W. (2011). Theoretical justification of popular link prediction heuristics. *IJCAI Proceedings-International Joint Conference on Artificial Intelligence* **2**, 2722.

Sarkar, P. and Moore, A. W. (2005). Dynamic social network analysis using latent space Models. *SIGKDD Explorations Newsletter* **7**, 31–40.

Schweinberger, M. and Handcock, M. S. (2015). Local dependence in random graph models: Characterization, properties and statistical inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **77**, 647–676.

Van Duijn, M. A., Gile, K. J. and Handcock, M. S. (2009). A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models. *Social Networks* **31**, 52–62.

Wang, Y. J. and Wong, G. Y. (1987). Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association* **82**, 8–19.

Wasserman, S. and Faust, K. (1994). Social network analysis: Methods and applications. *Cambridge University Press.*

Watts, D. and Strogatz, S. (1998). Collective dynamics of small world networks. *Nature* **393**, 440–442.

Wolfe, P. J. and Olhede, S. C. (2013). Nonparametric graphon estimation. *arXiv preprint arXiv:1309.5936.*

Yu, Y. and Wang, X. (2014). Link prediction in directed network and its application in microblog. *Mathematical Problems in Engineering*, 2014.

Zhang, J. and Chen, Y. (2013). Sampling for conditional inference on network data. *Journal of the American Statistical Association* **108**, 1295–1307.

Center of Data Science and Information Quality, Department of Information Management and E-Business, School of Management, Xi'an Jiaotong University, Xi'an, Shaanxi, China.

E-mail: xiangyuchang@gmail.com

Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing, China.

E-mail: dyhuang89@126.com

Department of Business Statistics and Econometrics, Guanghua School of Management, Peking University, Beijing, China.

E-mail: hansheng@gsm.pku.edu.cn

(Received February 2017 ; accepted November 2017)