Statistics and Its Interface Volume 15 (2022) 267–282

Link prediction via latent space logistic regression model*

Rui Pan, Xiangyu Chang†, Xuening Zhu, and Hansheng Wang

Nowadays, link prediction is of vital importance in the operation of social network platforms. One typical application is to make accurate recommendation to enhance users' activeness. In this article, we propose a latent space logistic regression model for link prediction. The model takes both the users' attributes and the latent social space into consideration. Two pseudo maximum likelihood estimators are proposed for parameter estimation. They correspond to the concepts of reciprocity and transitivity, respectively, and are computationally efficient for large-scale social networks. Extensive simulation studies are provided to evaluate the finite sample performance of the newly proposed methodology. At last, a real data set of Sina Weibo is presented for illustration purposes.

Keywords and phrases: Latent Social Space, Link Prediction, Logistic Regression, Network Topology, Social Networks.

1. INTRODUCTION

A social network refers to a set of nodes (i.e., users) connected by various social relationships (i.e., links). Typical examples include but are not limited to Facebook (www.facebook.com), Twitter (www.twitter.com), WeChat (www.wechat.com), QQ (www.qq.com), Weibo (www.weibo.com), and many others. On one side, it is vital for a social network to be well connected so that information can diffuse efficiently. On the other side, most large-scale social networks are extremely sparse in structure. This means two arbitrarily selected nodes are hardly to be connected directly (Zhou et al., 2017). However, the small world property (Watts and Strogatz, 1998) ensures that disconnected nodes could be linked through a finite number of intermediate nodes. This can be achieved by recommending each node promising social relationships. Thus, it is of practical importance to establish a good recommendation system because many disconnected nodes could be linked. Furthermore, nodes' loyalties to the network platform can be potentially enhanced. The problem of making recommendation is also commonly known as making accurate *link prediction* (Aiello et al., 2012).

Link prediction is a classical research topic in network data analysis. Many researchers have made substantive progresses in the past literatures; see Lu and Zhou (2011) and Wang et al. (2015) for detailed reviews. Two kinds of information are extensively used for link prediction. They are *node-based information* and *topology-based information*. More specifically, node-based information, seen as exogenous predictive variables, can be further classified into two categories. The first is *nodal information*. Consider for example Twitter, whether a particular node is a popular movie star can be seen as useful nodal information. The second is specific to one particular pair of nodes, and we refer to it as *pairwise information*. As a concrete example, whether two nodes are graduated from the same high school can be seen as pairwise information. More examples of node-based information can be found in Hunter et al. (2008) and de Miguel Luken and Tranmer (2010).

The second type of information extensively is topology-based information, which is closely related to network structures. It has been widely proved that information derived from network structures is extremely useful (Libennowell and Kleinberg, 2007). To fix the idea, consider a particular topology-based information, the *common neighbors* (Newman, 2001). As discussed before, in a large-scale network, the probability that two arbitrary nodes (i.e., i and j) are directly connected is extremely small. However, if there exist many other nodes (i.e., common neighbors) connected with both i and j, the likelihood for i and j to be connected can be significantly improved. Other topology-based information, such as paths between two nodes, can also be utilized for link prediction (Lu et al., 2009).

*The research of Rui Pan is supported in part by National Natural Science Foundation of China (Nos. 11601539, 11631003), the disciplinary funding and the Emerging Interdisciplinary Project of Central University of Finance and Economics. The research of Xiangyu Chang is supported in part by the National Natural Science Foundation of China (No. 11771012) and Natural Science Foundation of Shannxi Province of China (No. 2021JC-01). The reserch of Xuening Zhu is supported by the National Natural Science Foundation of China (Nos. 11901105, 71991472, U1811461), the Shanghai Sailing Program for Youth Science and Technology Excellence (19YF1402700). The research of Hansheng Wang is partially supported by National Natural Science Foundation of China (No. 11831008) and also partially supported by the Open Research Fund of Key Laboratory of Advanced Theory and Application in Statistics and Data Science (KLATASDS-MOE-ECNU-KLATASDS2101).

†Corresponding author.

From the above discussions, it can be concluded that both node-based and topology-based information play a vital role in link prediction. Among the many link prediction techniques, logistic regression (LR) model is frequently used in practice, taking both information into account; see Herz (2015), Kruse et al. (2016), and many others. Note that one type of LR model, the latent space model proposed by Hoff et al. (2002), is of great interest in both theory and applications. In this model, nodes are assumed to be embedded in a latent social space, where nodes staying close to each other are more likely to be connected. Node-based information can be easily accommodated and many stylized network properties (e.g., reciprocity, transitivity, and so forth) can be well modeled. As a consequence, the latent space model and its extensions are widely studied and popularly used (Schweinberger and Snijders, 2003; Handcock et al., 2007; Krivitsky et al., 2009; Oconnor et al., 2015; Sewell and Chen, 2015; Rastelli et al., 2016; Han et al., 2019).

The latent space approach is attractive in interpreting network structures and existing estimation algorithms are mainly performed using Markov chain Monte Carlo (MCMC). To make the estimation procedure computationally fast, many efforts have been devoted (Xing et al., 2010; Salter-Townshend and Murphy, 2013; Sewell and Chen, 2017). In this work, we develop a novel *latent space logistic regression* (LSLR) model from the perspective of frequentist. The newly proposed LSLR model takes both latent social space and node-based information into account. Furthermore, the LSLR model can well express important characteristics of large-scale social networks, e.g., reciprocity and transitivity. Different from the existing latent space models, the LSLR model can express sparsity, which is commonly encountered in large-scale social network data. To estimate the regression coefficient, we develop two pseudo maximum likelihood estimators (PMLEs), which correspond to the concepts of reciprocity and transitivity. The newly proposed estimators are computationally efficient and particularly useful for large-scale network analysis. The asymptotic properties of the PMLEs are derived and two inference procedures are developed. Other than that, we also conduct various types of link prediction. Extensive simulation studies are conducted to evaluate their finite sample performances. Lastly, a real dataset about Sina Weibo is provided for illustration purposes.

The rest of article is organized as follows. In Section 2, we describe the LSLR model and investigate the distribution of some important network statistics under the LSLR model. In Section 3, we develop two estimation methods with their asymptotic properties and discuss the problem of link prediction. A number of simulation studies and real data example are conducted in Section 4. Some concluding remarks are given in Section 5. All the technical proofs are presented in the Appendix.

2. LATENT SPACE LOGISTIC REGRESSION MODEL

2.1 Model and notation

In order to describe the network structure, we define the $n \times n$ adjacency matrix $A = (A_{ij}) \in \{0, 1\}$, where n is the total number of nodes. If node i claims a certain relationship to node j, then $A_{ij} = 1$, otherwise, $A_{ij} = 0$. We follow the convention and let $A_{ii} = 0$ for $1 \le i \le n$. In this work, we focus on directed network so that A_{ij} can be unequal to A_{ji}. Further assume that $X_{ij} = (X_{ij,1}, \ldots, X_{ij,p})^\top \in \mathbb{R}^p$ is a p-dimensional predictive variable derived from nodes i and j for $i \neq j$. Any useful information for predicting A_{ij} can be included in X_{ij}. For instance, X_{ij} could be (1) whether node i or j carries certain characteristics (e.g., movie star, business elite or political leader), (2) whether i and j have the same nodal factor effect (e.g., gender), (3) quantitative measurement for social connection between i and j (e.g., the amount of phone calls made between i and j), and (4) difference in social status between i and j (e.g., absolute difference of annual income between i and j). For notation simplicity, we denote $\mathbb{X} = \{X_{ij}\}$ as the predictor set.

To model the stochastic mechanism of link formation (i.e., $A_{ij} = 1$), we assume that there exists a latent social space and each node has its own position in the space. To this end, define the unobserved position Z_i for node i. As commonly accepted, we suppose that $Z_i \in \mathbb{R}^1$ is treated as a random variable and $\mathbb{Z} = (Z_1, \cdots, Z_n)^\top \in \mathbb{R}^n$ according to the existing literature (Hoff et al., 2002; Sewell and Chen, 2015; Chang et al., 2019). Further define d_{ij} as the distance between node i and j in the latent space, where $d_{ij} = |Z_i - Z_j|$. Intuitively, if two nodes are close to each other in the latent social space (i.e., d_{ij} is small), the probability they are connected should be relatively large. To mimic this phenomenon, we propose the following LSLR model,

$$P(A_{ij} = 1 | \mathbb{X}, \mathbb{Z}) = \exp\big(-\alpha_n d_{ij}^2\big) \frac{e^{X_{ij}^\top \beta}}{1 + e^{X_{ij}^\top \beta}}, \tag{2.1}$$

where $\beta = (\beta_1, \cdots, \beta_p)^\top \in \mathbb{R}^p$ is the associated p-dimensional regression coefficient, with its true value given by $\beta_0 = (\beta_{01}, \cdots, \beta_{0p})^\top \in \mathbb{R}^p$. Further assume $c_1 n^\delta \le \alpha_n \le c_2 n^\delta$, where δ, c_1 and c_2 are some positive constants. As a result, α_n is an unknown parameter controlling the distance effect.

Note that (2.1) can be seen as a two-step process in link formation. First, nodes i and j should be close enough to each other (i.e., d_{ij}^2 should be small enough). So that the "probability" (i.e., $\exp(\alpha_n d_{ij}^2)$) they "meet" each other is nontrivial. In the meanwhile, nodes i and j should be similar to each other in terms of predictive variables. As a result, the "probability" (i.e., $\exp(X_{ij}^\top \beta)/\{1 + \exp(X_{ij}^\top \beta)\}$) they "resemble" each other is nontrivial. This makes the link formation between i and j. The whole process is further shown in Figure 1.

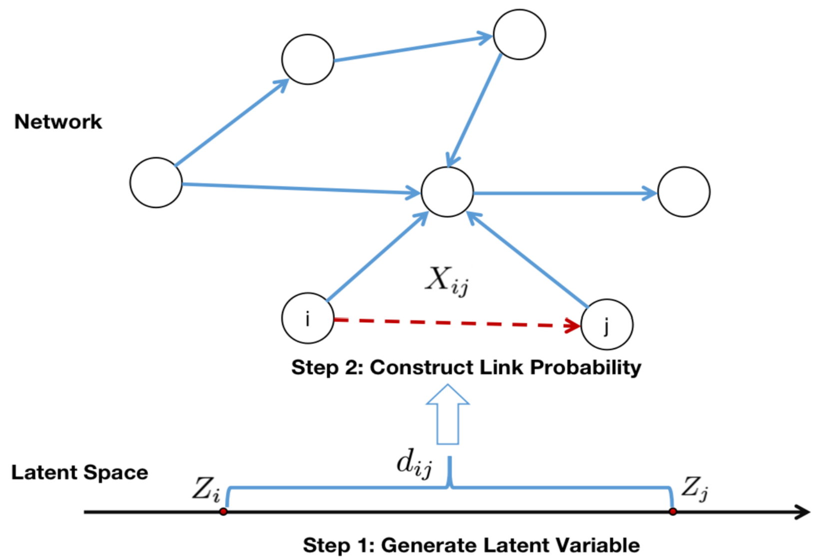

Figure 1. The formation of (2.1), which can be seen as a two-step process.

Regarding the LSLR model, we have the following three remarks.

Remark 1 (Network Sparsity)**.** Note that in the LSLR model (2.1), the unknown parameter α_n controls the strength of distance effect. We further require that α_n is positive, otherwise sparsity cannot be ensured. Even worse, the conditional probability of (2.1) might be greater than 1 if α_n is negative. It can be concluded that the probability nodes i and j are connected converges to 0 as $n \to \infty$, as long as $Z_i \neq Z_j$. As a result, the assumption on α_n guarantees network sparsity, especially for large-scale networks.

Remark 2 (Reciprocity and Transitivity)**.** The LSLR model possesses important network characteristics: reciprocity and transitivity. To see this, assume that node i follows node j, and node j follows node k (i.e., $A_{ij} = 1$ and $A_{jk} = 1$), then it is commonly observed with a high probability that $A_{ji} = 1$ and $A_{ik} = 1$. Theses two phenomena (i.e., $A_{ij}A_{ji} = 1$ and $A_{ij}A_{jk}A_{ik} = 1$) are referred to as reciprocity and transitivity. In the next subsection, we will show in detail that the LSLR model satisfies these two important characteristics. Further note that β is assumed to be size-invariant, which means that β does not change with the size of network n. Otherwise, transitivity may not be guaranteed.

Remark 3 (Relationship to LR model)**.** If two nodes are located at the same place in the latent social space (i.e., $Z_i = Z_j$), the probability that node i is connected to j becomes a constant, i.e., $\exp(X_{ij}^\top\beta)/\{1 + \exp(X_{ij}^\top\beta)\}$. This is a nontrivial probability which is bounded away from 0. In this case, the LSLR model degenerates to traditional LR model. In addition, we write the log-odds of our proposed LSLR model,

$$X_{ij}^\top\beta - \alpha_n d_{ij}^2 - \log\left\{1 + e^{X_{ij}^\top\beta}(1 - e^{-\alpha_n d_{ij}^2})\right\}.$$

It can be seen that the first two terms are the same as the latent space model. However, there is a third term $\log\{1+\exp(X_{ij}^\top\beta)(1-\exp(-\alpha_n d_{ij}^2))\}$, which makes the interpretation of β different from that of the latent space model. What we can conclude is that if the term $\alpha_n d_{ij}^2$ goes to 0, the third term goes to 0, and then the interpretation of β is the same as that of the latent space model.

Note that in model (2.1), the latent positions $\mathbb{Z}$ are unobserved. In order to make parameter estimation feasible, we will derive the marginal probability $P(A_{ij} = 1|\mathbb{X})$. To this end, define $U_{ij} = Z_i - Z_j$ to be a random variable. We further assume its probability density function as $f(u) \geq 0$, where $u \in \mathbb{R}^1$. As long as $f(0) > 0$, it is easy to have the following result as $n \to \infty$,

$$\begin{aligned} P(A_{ij} = 1|\mathbb{X}) &= \frac{e^{X_{ij}^\top\beta}}{1 + e^{X_{ij}^\top\beta}} \int_{\mathbb{R}} e^{-\alpha_n u^2} f(u)du \\ &= \frac{e^{X_{ij}^\top\beta}}{1 + e^{X_{ij}^\top\beta}} \int_{\mathbb{R}} e^{-t^2} f\Big(\frac{t}{\sqrt{\alpha_n}}\Big)\Big(\frac{1}{\sqrt{\alpha_n}}\Big)dt \\ &= \Big\{f(0) + O(n^{-\delta/2})\Big\}\Big(\frac{\pi}{\alpha_n}\Big)^{\frac{1}{2}} \frac{e^{X_{ij}^\top\beta}}{1 + e^{X_{ij}^\top\beta}}, \end{aligned} \tag{2.2}$$

where latent positions are integrated out. Note that the marginal probability of $A_{ij} = 1$ converges to 0 as the network size n increases. As a result, network sparsity can be guaranteed for large-scale networks. Based on the result in (2.2), we next study the properties of some important network statistics.

2.2 Network statistics

In order to explore social connections among nodes, various network statistics are extensively studied in the literature of social network analysis, e.g., degree, reciprocity, transitivity, and so forth (Wasserman and Faust, 1994). In this subsection, we intend to study the properties of these important network statistics under the LSLR model. We start with the simplest network statistic, that is the nodal degree.

Nodal degree

We first introduce notations A_{+i} and A_{i+}, denoting nodal in-degree and out-degree, respectively. Mathematically, they can be spelled as $A_{+i} = \sum_{j \neq i} A_{ji}$ and $A_{i+} = \sum_{j \neq i} A_{ij}$. In the study of sociology, a node with large nodal in-degree is believed to be attractive in the network, who draws many attentions from others. A node with large nodal out-degree is active, who initiatively follows many other nodes. As a result, nodal degree can reflect certain characteristics of a node and it is of great interest to investigate its properties.

From (2.2), the expectation of nodal in-degree can be easily derived as

$$E(A_{+i}|\mathbb{X}) = E\Big(\sum_{j\neq i} A_{ji}|\mathbb{X}\Big)$$
$$= \Big\{f(0) + O(n^{-\delta/2})\Big\}\Big(\frac{\pi}{\alpha_n}\Big)^{\frac{1}{2}} \sum_{j\neq i} \frac{e^{X_{ji}^\top \beta}}{1+e^{X_{ji}^\top \beta}}.$$

We then focus on the leading term of this expectation and consider two special scenarios regarding nodal in-degree for detailed discussions.

Scenario 1 ($X_{ij} = X_j$)**.** In the first scenario, we let $X_{ij} = X_j$, where $X_j = (X_{j,1}, \cdots, X_{j,p})^\top \in \mathbb{R}^p$ denotes nodal characteristics with respect to node j. As a result, the predictor between i and j is only determined by the nodal information from the receiver (i.e., node j). As a result, the leading term of $E(A_{+i}|\mathbb{X})$ is

$$(n-1)f(0)\Big(\frac{\pi}{\alpha_n}\Big)^{\frac{1}{2}} \frac{e^{X_i^\top \beta}}{1+e^{X_i^\top \beta}}.$$

The expected nodal in-degree for node i is only determined by its own nodal information X_i. If X_i is some information that exhibits large variability among nodes (i.e., personal social status), the nodal in-degree would vary tremendously for each node.

Scenario 2 ($X_{ij} = X_i$)**.** In the second scenario, we let $X_{ij} = X_i$. This means that the predictive variable is equivalent to the nodal information derived from the sender (i.e., node i). In this case, the leading term of $E(A_{+i}|\mathbb{X})$ can be expressed as

$$(n-1)f(0)\Big(\frac{\pi}{\alpha_n}\Big)^{\frac{1}{2}} \frac{1}{n-1}\sum_{j\neq i} \frac{e^{X_j^\top \beta}}{1+e^{X_j^\top \beta}}.$$

The expected nodal in-degree for a particular node i is determined by the average nodal effect generated from the other nodes (i.e., X_j with $j \neq i$). In this case, the variation of nodal in-degree will be relatively small under this scenario.

Similar conclusion can be obtained for nodal out-degree. Further note that $E(A_{+i}) = O(n^{1-\delta/2})$ and $E(A_{i+}) = O(n^{1-\delta/2})$. When $0 < \delta < 2$, the expected nodal in-degree and out-degree diverge with network size. When $\delta = 2$, the expected nodal in-degree and out-degree are well bounded, i.e., $E(A_{+i}) = O(1)$ and $E(A_{i+}) = O(1)$. Take nodal out-degree for instance, this means that when network size n gets larger, the number of nodes that one can follow is limited. When $\delta > 2$, the expected nodal in-degree and out-degree diminish along with the expansion of the network. Although this is counterintuitive in real practice, we still discuss it here for theoretical completeness.

Reciprocity

Next, we study another important network structure, that is reciprocity. To this end, define a *dyad* as $D_{ij} = (A_{ij}, A_{ji})$ with $i < j$ (Holland and Leinhardt, 1981). Then, a symmetric dyad with $D_{ij} = (1,1)$ is said to be reciprocated, where i and j are mutually connected. Using similar technique of deriving (2.2), we have that

$$P(A_{ij}A_{ji} = 1|\mathbb{X}) = \Big\{f(0) + O(n^{-\delta/2})\Big\}\Big(\frac{\pi}{2\alpha_n}\Big)^{\frac{1}{2}} p_{ij}p_{ji}, \tag{2.3}$$

where $p_{ij} = \exp(X_{ij}^\top\beta)/\{1+\exp(X_{ij}^\top\beta)\}$. It can be seen that $P(A_{ij}A_{ji} = 1|\mathbb{X})$ is $O(n^{-\delta/2})$. In a large-scale network, the probability that any two nodes are mutually connected is very small, and it converges to 0 as $n \to \infty$.

Although it is rare to observe a reciprocated dyad (i.e., $D_{ij} = (1,1)$), it is believed that given $A_{ji} = 1$, the probability of observing $A_{ij} = 1$ should be bounded away from 0. Under the framework of LSLR model, it can be derived that

$$P(A_{ij} = 1|A_{ji} = 1, \mathbb{X}) = P(A_{ij}A_{ji} = 1|\mathbb{X})/P(A_{ji} = 1|\mathbb{X})$$
$$= \sqrt{1/2}\frac{e^{X_{ij}^\top\beta}}{1+e^{X_{ij}^\top\beta}} + O(n^{-\delta/2}). \tag{2.4}$$

As a result, once node j follows node i (i.e., $A_{ji} = 1$), the probability that i follows back (i.e., $A_{ij} = 1$) is relatively large. In this sense, the phenomenon of reciprocity can be well explained by our newly proposed LSLR model.

Transitivity

We next study a more sophisticated network structure, which is transitivity. First of all, define a *triad* to be three different nodes (e.g., i, j, and k) and all possible edges involved among them. A triad is said to be transitive if i is related to j, j is related to k, and then i is related to k (i.e., $A_{ij}A_{jk}A_{ik} = 1$). A large number of transitive triads implies that the network structure is clustered (Wasserman and Faust, 1994).

We first derive the marginal probabilities of $P(A_{ij}A_{jk} = 1|\mathbb{X})$ and $P(A_{ij}A_{jk}A_{ik} = 1|\mathbb{X})$. Recall that $U_{ij} = Z_i - Z_j$. We assume $g(u, v)$ to be the joint probability density function of U_{ij} and U_{kl} for any $i \neq j$ and $k \neq l$, where $u \in \mathbb{R}^1$ and $v \in \mathbb{R}^1$. Given $g(0, 0) > 0$, we can use the same technique as in (2.2) and obtain

$$\begin{aligned} P(A_{ij}A_{jk} = 1|\mathbb{X}) &= p_{ij}p_{jk}\int_{\mathbb{R}}\int_{\mathbb{R}} e^{-\alpha_n(u^2+v^2)}g(u, v)dudv \\ &= \Big\{g(0, 0) + O(n^{-\delta/2})\Big\}\Big(\frac{\pi}{\alpha_n}\Big)p_{ij}p_{jk}. \end{aligned}$$

Further note that $U_{ik} = Z_i - Z_k = U_{ij} + U_{jk}$. As a result,

$$\begin{aligned} &P(A_{ij}A_{jk}A_{ik} = 1|\mathbb{X}) \\ &= p_{ij}p_{jk}p_{ik}\int_{\mathbb{R}}\int_{\mathbb{R}} e^{-\alpha_n\{u^2+v^2+(u+v)^2\}}g(u, v)dudv \\ &= \Big\{g(0, 0) + O(n^{-\delta/2})\Big\}\Big(\frac{\pi}{\sqrt{3}\alpha_n}\Big)p_{ij}p_{jk}p_{ik}. \end{aligned}$$

It can then be proved that

$$\begin{aligned} &P(A_{ik} = 1|A_{ij}A_{jk} = 1, \mathbb{X}) \\ &\quad = P(A_{ij}A_{jk}A_{ik} = 1|\mathbb{X})/P(A_{ij}A_{jk} = 1|\mathbb{X}) \\ &\quad = (\sqrt{1/3})\frac{e^{X_{ik}^\top\beta}}{1 + e^{X_{ik}^\top\beta}} + O(n^{-\delta/2}). \end{aligned} \tag{2.5}$$

From (2.5) we can see that once $A_{ij}A_{jk} = 1$, the probability of observing $A_{ik} = 1$ is bounded away from 0. Furthermore, this probability is only determined by the information obtained from nodes i or/and k (i.e., X_{ik}), and it has nothing to do with the intermediate node j. Hence, the LSLR model is able to express the transitivity property very well.

3. ESTIMATION METHODS AND LINK PREDICTION

In this section, we develop a novel estimation method for the unknown parameters in the LSLR model. To be more specific, we are particularly interested in the regression coefficient β. Since the social position Z_i is latent, the full likelihood function constructed on (2.1) involves integration, which is hard to calculate. As a straightforward alternative, one can start from the marginal probability (2.2). This leads to the following log-transformed pseudo likelihood function,

$$\begin{aligned} \ell(\alpha_n, \beta|\mathbb{X}) = \sum_{i\neq j}\Bigg\{&A_{ij}X_{ij}^\top\beta + A_{ij}\log\eta(\alpha_n) - \log\Big(1 + e^{X_{ij}^\top\beta}\Big) \\ &+ (1 - A_{ij})\log\Big(1 + \Big(1 - \eta(\alpha_n)\Big)e^{X_{ij}^\top\beta}\Big)\Bigg\}, \end{aligned} \tag{3.1}$$

where $\eta(\alpha_n) = f(0)(\pi/\alpha_n)^{1/2}$. Since α_n is not identifiable, we do not estimate α_n but focus on the estimation of β.

Remark 4. Note that the maximum likelihood estimator of β in (3.1) is not easy to obtain. Firstly, the summation in (3.1) involves all the possible pairs in the network, which is of $O(n^2)$. This leads to high computational cost for a large-scale network. Secondly, the probability density function of U_{ij}, i.e., $f(\cdot)$, needs to be evaluated. In order to avoid the computational burden as well as estimating the unknown probability density function, we develop two pseudo maximum likelihood estimators, which correspond to the concepts of reciprocity and transitivity.

3.1 Pseudo maximum likelihood estimator

As mentioned before, we develop two new estimators for β in this subsection. On one hand, the computational cost can be reduced. On the other hand, the computational procedure is irrelevant with α_n. We refer to the newly proposed estimator as a pseudo maximum likelihood estimator. More specifically, we propose two kinds of PMLE in this subsection, which correspond to the concepts of reciprocity and transitivity.

Recall that a dyad refers to a pair of nodes and the possible links formed between them. It is denoted by $D_{ij} = (A_{ij}, A_{ji})$ with $i < j$. In a large-scale network, most dyads are null with $D_{ij} = (0, 0)$ (Huang et al., 2016; Zhou et al., 2017), which carry less information. This motivates us to develop an estimator regarding $D_{ij} \neq (0, 0)$. Furthermore, we know the probability of $A_{ij} = 1$ given $A_{ji} = 1$ is $(\sqrt{1/2})\exp(X_{ij}^\top\beta)/\{1+\exp(X_{ij}^\top\beta)\}$, which is irrelevant with α_n. As a result, we construct the following log-transformed pseudo likelihood function

$$\begin{aligned} \ell_R(\beta|\mathbb{X}) &= \sum_{A_{ji}=1, i\neq j}\Bigg\{A_{ij}X_{ij}^\top\beta - \log(1 + e^{X_{ij}^\top\beta}) \\ &\qquad + (1 - A_{ij})\log\Big(1 + \gamma_0 e^{X_{ij}^\top\beta}\Big)\Bigg\} \\ &= \sum_{i\neq j}\Bigg\{A_{ij}X_{ij}^\top\beta - \log(1 + e^{X_{ij}^\top\beta}) \\ &\qquad + (1 - A_{ij})\log\Big(1 + \gamma_0 e^{X_{ij}^\top\beta}\Big)\Bigg\}A_{ji}, \end{aligned} \tag{3.2}$$

where some constants are ignored and $\gamma_0 = 1 - \sqrt{1/2}$. We denote the PMLE of β as $\hat{\beta}_R = \operatorname{argmax}_\beta \ell_R(\beta|\mathbb{X})$. Note that the conditional likelihood function (3.2) only takes summation of the reciprocated dyad (i.e., $D_{ij} = (1, 1)$) and asymmetric dyad (i.e., $D_{ij} = (1, 0)$ or $D_{ij} = (0, 1)$). Since the null dyad (i.e., $D_{ij} = (0, 0)$) is not taken into account, the computation of PMLE should be more efficient.

Similarly, transitivity refers to the network structure formed among three nodes, where $A_{ij}A_{jk}A_{ik} = 1$. We then develop a PMLE based on the concept of transitivity. More specifically, given $A_{ik}A_{kj} = 1$, we know that the conditional probability of $A_{ij} = 1$ is $(\sqrt{1/3})\exp(X_{ij}^\top\beta)/\{1 +$

$\exp(X_{ij}^\top\beta)\}$. As a result, we construct the following log-transformed likelihood function

$$\ell_T(\beta|\mathbb{X}) = \sum_{i\neq j}\sum_{k\neq i,j}^{A_{ik}A_{kj}=1}\Big\{A_{ij}X_{ij}^\top\beta - \log(1+e^{X_{ij}^\top\beta}) + (1-A_{ij})\log\Big(1+\nu_0 e^{X_{ij}^\top\beta}\Big)\Big\}$$

$$= \sum_{i\neq j}\sum_{k\neq i,j}\Big\{A_{ij}X_{ij}^\top\beta - \log(1+e^{X_{ij}^\top\beta}) + (1-A_{ij})\log\Big(1+\nu_0 e^{X_{ij}^\top\beta}\Big)\Big\}A_{ik}A_{kj}, \tag{3.3}$$

where some constants are ignored and $\nu_0 = 1-(\sqrt{1/3})$. Denote the PMLE as $\hat\beta_T = \operatorname{argmax}_\beta\ \ell_T(\beta|\mathbb{X})$. We next investigate the asymptotic properties of both $\hat\beta_R$ and $\hat\beta_T$.

3.2 Asymptotic properties

In order to establish asymptotic normality of the newly proposed PMLEs, we introduce some notations at the first hand. Define $l_{ij} = \{A_{ij}X_{ij}^\top\beta - \log(1+e^{X_{ij}^\top\beta}) + (1-A_{ij})\log(1+\gamma_0 e^{X_{ij}^\top\beta})\}A_{ji}$, then $\ell_R(\beta|\mathbb{X}) = \sum_{i\neq j}\ell_{ij}(\beta)$. Denote $s_{ij}(\beta) = \partial\ell_{ij}(\beta)/\partial\beta$ as the score function. Furthermore, let $\tilde s_{ij}(\beta) = s_{ij}(\beta) - E(s_{ij}(\beta)|\mathbb{X})$, and it can be proved that $E(\tilde s_{ij}(\beta)|\mathbb{X}) = 0$. Lastly, let $\ddot\ell_{ij}(\beta|\mathbb{X}) = \partial^2\ell_{ij}(\beta)/\partial\beta^\top\partial\beta$. The following technical conditions are needed for building asymptotic properties of PMLE $\hat\beta_R$.

- (C1) (Predictor Assumption) For any $i\neq j$, assume that X_{ij}s are random variables, with mean vector $\mathbf{0}$ and covariance matrix $\Sigma_X \in \mathbb{R}^{p\times p}$, where Σ_X is positive definite. Furthermore, for any X_{ij} and X_{kl} with $\{i,j\}\bigcap\{k,l\} = \emptyset$, assume that $\mathrm{cov}(X_{ij}, X_{kl}) = I_{p\times p}$, where $I_{p\times p}\in\mathbb{R}^{p\times p}$ is an identity matrix.
- (C2) (Coefficient Regularity) Assume $c_1 n^\delta \le \alpha_n \le c_2 n^\delta$, where c_1 and c_2 are some positive constants, and $1<\delta\le 2$.
- (C3) (Latent Position) Assume the latent positions $Z_1,\cdots,Z_n$ are independent and identically distributed. Define $U_{ij} = Z_i - Z_j$. Assume the probability density function of U_{ij} is $f(\cdot)\ge 0$ with $f(0)>0$. Furthermore, given $\mathbb{X}$ and $\mathbb{Z}$, all A_{ij}s are independent.

By (C1) we know that X_{ij}s are assumed to be random variables. Furthermore, X_{ij} and X_{kl} are mutually uncorrelated with $\{i,j\}\bigcap\{k,l\}=\emptyset$. This means that nontrivial dependence is allowed, for example, between X_{ij} and X_{ik} in a flexible manner. We next explain condition (C2) in detail. First of all, α_n is allowed to diverge with the sample size n. Denote the first- and second-order derivatives of the likelihood function $\ell_R(\beta)$ as $Q_1(\beta)\in\mathbb{R}^p$ and $Q_2(\beta)\in\mathbb{R}^{p\times p}$. It can be verified that the expectation of Q_1 is not exactly $\mathbf{0}$, so we use $Q_1^* = Q_1 - E(Q_1|\mathbb{X})$ for approximation. As shown in Appendix A, we can prove $\hat\beta_R - \beta_0 = \{Q_2\}^{-1}Q_1^* + o_p(1)$, and the normality of $\hat\beta_R$ comes from the asymptotic distribution of Q_1^*.

Recall the discussions of nodal degree in subsection 2.2, it is required that $\delta\le 2$. Otherwise, the expected nodal in- and out-degree will diminish along with the expansion of the network, which is counterintuitive in real practice. In the meanwhile, it is shown in Appendix A that $\{\mathrm{cov}^{-1/2}(Q_1^*)\}Q_1^*$ converges in distribution to a multivariate standard normal random vector under the condition $\delta>1$. The marginal probability of $A_{ij}=1$ given $\mathbb{X}$ is of $O(n^{-\delta/2})$. When $\delta=2$, this is an $O(n^{-1})$. When $\delta>2$, the probability that any two nodes are connected is smaller than $O(n^{-1})$, which will generate a really sparse network when n is large. As a result, we require $1<\delta\le 2$.

In condition (C3), it is assumed that the probability density function satisfying $f(0)>0$. Note that neither the support nor the function change with the number of nodes. This is commonly accepted in the past literatures where the distribution is assumed to be normal (Hoff et al., 2002; Krivitsky et al., 2009; Sewell and Chen, 2015; Chang et al., 2019). We then have the following theorem.

Theorem 1. *Assume the LSLR model (2.1) holds. Further assume conditions (C1)–(C3). As $n\to\infty$, we have that $\sqrt{n}(\hat\beta_R - \beta_0)\to_d N(0, H_1^{-1}C_1H_1^{-1})$, where C_1 is some positive definite matrix defined in Appendix A, and $H_1 = -E\{\ddot\ell_{ij}(\beta)\}$ is a positive definite matrix.*

The proof of Theorem 1 is given in Appendix A. As we can see in the conclusion of Theorem 1, the PMLE based on reciprocity is $\sqrt{n}$-consistent.

Similarly, we derive the asymptotic normality of $\hat\beta_T$, which is based on the concept of transitivity. To this end, let $\ell_T(\beta) = \sum_{i\neq j}\sum_{k\neq i,j}\ell_{ijk}(\beta)$. Denote $s_{ijk}(\beta) = \partial\ell_{ijk}(\beta)/\partial\beta$ as the score function. Furthermore, let $\tilde s_{ijk}(\beta) = s_{ijk}(\beta) - E(s_{ijk}(\beta)|\mathbb{X})$, where $E(\tilde s_{ijk}(\beta)|\mathbb{X}) = 0$. Lastly, denote the second order derivative of the likelihood function as $\ddot\ell_{ijk}(\beta) = \partial\ell_{ijk}(\beta)/\partial\beta^\top\partial\beta$.

Theorem 2. *Assume the LSLR model (2.1) holds. Further assume conditions (C1)–(C3). As $n\to\infty$, we have that $\sqrt{n}(\hat\beta_T - \beta_0)\to_d N(0, H_2^{-1}C_2H_2^{-1})$, where C_2 is some positive definite matrix defined in Appendix B, and $H_2 = -E\{\ddot\ell_{ijk}(\beta)\}$ $(i\neq j\neq k)$.*

The proof of Theorem 2 is given in Appendix B. It can be seen that $\hat\beta_T$ is also $\sqrt{n}$-consistent, which is the same as $\hat\beta_R$. However, the asymptotic covariances of $\hat\beta_R$ and $\hat\beta_T$ are totally different, which result in different inference procedures. The details are given in the next subsection.

3.3 Estimation of standard error

In order to make inference of the newly proposed PMLEs, the standard errors developed in Theorem 1 and 2 need to be estimated. First of all, by Theorem 1 we know that the asymptotic covariance of $\sqrt{n}(\hat{\beta}_R - \beta_0)$ is given by $H_1^{-1}C_1H_1^{-1}$. It can be estimated by $\hat{H}_1^{-1}\hat{C}_1\hat{H}_1^{-1}$, where

$$\hat{H}_1 = \frac{1}{n^{(2-\delta/2)}}\sum_{i\neq j}\frac{A_{ji}e^{X_{ij}^\top\hat{\beta}_R}}{\big(1+\gamma_0 e^{X_{ij}^\top\hat{\beta}_R}\big)^2} \times\left\{\gamma_0 A_{ij} + \frac{(1-\gamma_0)\big(1-\gamma_0 e^{2X_{ij}^\top\hat{\beta}_R}\big)}{\big(1+e^{X_{ij}^\top\hat{\beta}_R}\big)^2}\right\}X_{ij}X_{ij}^\top,$$

$$\hat{C}_1 = n^{-1+\delta}\sum_{i=1}^{n}(\hat{m}_i+\hat{h}_i)(\hat{m}_i+\hat{h}_i)^\top,$$

$$\hat{m}_i = (n-1)^{-1}\sum_{j\neq i}s_{ij}(\hat{\beta}_R),\quad \hat{h}_i = (n-1)^{-1}\sum_{j\neq i}s_{ji}(\hat{\beta}_R).$$

To test the statistical significance of one particular regression coefficient, i.e., $H_0: \beta_{0j} = 0$ versus $H_1: \beta_{0j}\neq 0$, a Z-type statistic can be established as $Z_R = \hat{\beta}_j^R/\widehat{\mathrm{SE}}_R(\hat{\beta}_j^R)$, where $\hat{\beta}_j^R$ is the jth element of $\hat{\beta}_R$ and $\widehat{\mathrm{SE}}_R^2(\hat{\beta}_j^R)$ is the jth diagonal element of $\hat{H}_1^{-1}\hat{C}_1\hat{H}_1^{-1}/n$. It can be seen that, there is no need to estimate the unknown coefficient δ. As a result, given a significant level α^*, the null hypothesis is rejected if $|Z_R| > z_{\alpha^*/2}$, where z_{α^*} is the α^*-th upper quantile of a standard normal distribution.

Similarly, the inference procedure can be developed for $\hat{\beta}_T$. To be more specific, the asymptotic covariance of $\sqrt{n}(\hat{\beta}_T - \beta_0)$ is given by $H_2^{-1}C_2H_2^{-1}$, which can be estimated by $\hat{H}_2^{-1}\hat{C}_2\hat{H}_2^{-1}$. The estimators are given by

$$\hat{H}_2 = \frac{1}{n^{(3-\delta)}}\sum_{i,j,k}\frac{A_{ik}A_{kj}e^{X_{ij}^\top\hat{\beta}_T}}{\big(1+\nu_0 e^{X_{ij}^\top\hat{\beta}_T}\big)^2} \times\left\{\nu_0 A_{ij} + \frac{(1-\nu_0)\big(1-\nu_0 e^{2X_{ij}^\top\hat{\beta}_R}\big)}{\big(1+e^{X_{ij}^\top\hat{\beta}_R}\big)^2}\right\}X_{ij}X_{ij}^\top,$$

$$\hat{C}_2 = n^{-1+2\delta}\sum_{i=1}^{n}(\hat{m}_i^*+\hat{h}_i^*+\hat{r}_i^*)(\hat{m}_i^*+\hat{h}_i^*+\hat{r}_i^*)^\top,$$

$$\hat{m}_i^* = (n-1)^{-1}(n-2)^{-1}\sum_{j\neq i}\sum_{k\neq i,j}s_{ijk}(\hat{\beta}_T),$$

$$\hat{h}_i^* = (n-1)^{-1}(n-2)^{-1}\sum_{j\neq i}\sum_{k\neq i,j}s_{jik}(\hat{\beta}_T),$$

$$\hat{r}_i^* = (n-1)^{-1}(n-2)^{-1}\sum_{j\neq i}\sum_{k\neq i,j}s_{jki}(\hat{\beta}_T).$$

In order to test the statistical significance of one particular regression coefficient β_{0j}, a Z-type statistics can be developed. More specifically, $Z_T = \hat{\beta}_j^T/\widehat{\mathrm{SE}}_T(\hat{\beta}_j^T)$, where $\hat{\beta}_j^T$ is the jth element of $\hat{\beta}_T$ and $\widehat{\mathrm{SE}}_T^2(\hat{\beta}_j^T)$ is the jth diagonal element of $\hat{H}_2^{-1}\hat{C}_2\hat{H}_2^{-1}/n$. Given a significant level α^*, the null hypothesis of $H_0: \beta_{0j} = 0$ can be rejected if $|Z_T| > z_{\alpha^*/2}$.

3.4 Link prediction via LSLR

In social networks analysis, link prediction attempts to estimate the likelihood of whether a link between two nodes exists, given both the observed edges and the attributes of nodes (Getoor and Diehl, 2005). In this subsection, we discuss how to estimate the link likelihood in social networks based on LSLR.

So far, we have obtained two PMLEs $\hat{\beta}_R$ and $\hat{\beta}_T$ of LSLR based on the concepts of reciprocity and transitivity. Therefore, link prediction can directly be performed, i.e., predicting $P(A_{ij} = 1|\mathbb{X})$ from (2.2). To be more specific, one only needs to calculate $\exp(X_{ij}^\top\hat{\beta})/\{1+\exp(X_{ij}^\top\hat{\beta})\}$ for $i\neq j$, where $\hat{\beta}$ can be either $\hat{\beta}_R$ or $\hat{\beta}_T$. As a result, denote the prediction of A_{ij} based on PMLEs as

$$\hat{A}_{ij}^P = I\Big(\exp(X_{ij}^\top\hat{\beta})/\{1+\exp(X_{ij}^\top\hat{\beta})\} > \xi_0\Big),$$

where the positive constant ξ_0 is some threshold. Note that there is no need to estimate $f(\mathbf{0})$ and α_n in (2.2), since they are irrelevant with i and j. In real practice, the choice of threshold ξ_0 can be determined according to specific research problem.

In a network without nodal information, topology-based metrics are extensively used for link prediction, among which the methods of "common neighbors" (CN) are proved to be useful (Yu and Wang, 2014; Wang et al., 2015; Chang et al., 2019). To be more specific, we adopt three different definitions of CN, i.e., transitivity, 2-in-star and 2-out-star. The corresponding network structures are shown in Figure 2. Mathematically, we define $CN_{ij}^1 = \sum_{k\neq i,j}A_{ik}A_{kj}$, $CN_{ij}^2 = \sum_{k\neq i,j}A_{ik}A_{jk}$ and $CN_{ij}^3 = \sum_{k\neq i,j}A_{ki}A_{kj}$ as the number of common neighbors between i and j. As a result, link prediction based on CN can be formulated as

$$\hat{A}_{ij}^{CN} = I\Big(CN_{ij} > \xi_1\Big),$$

where CN_{ij} can be CN_{ij}^1, CN_{ij}^2, CN_{ij}^3, and ξ_1 is some threshold.

In order to compare the method of the newly proposed PMLEs (i.e., $\hat{A}_{ij}^P$) and the results derived from CN (i.e., $\hat{A}_{ij}^{CN}$), we further draw the receiver operation curves (ROC) and calculate the value of the area under curve (AUC). As a result, AUC values are evaluated and more details can be found in numeric studies.

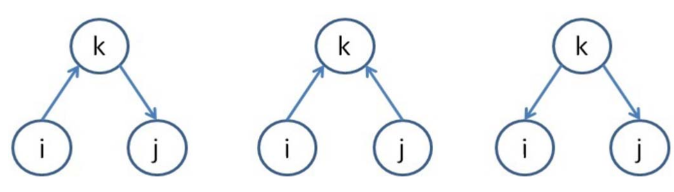

Figure 2. Three different network structures. From the left to the right: transitivity, 2-in-star, 2-out-star.

4. NUMERICAL STUDIES

4.1 Model setup

In order to demonstrate the finite sample performances of our proposed PMLEs, we conduct a number of simulation studies in this section. Different network sizes are considered with $n = 500$, 1000, and 2000. Accordingly, α_n is set to be cn^{δ}, where $\delta = 1.5$, 1.8, and 2 with appropriately chosen constant c. Nodal information $V_i = (V_{i1}, \cdots, V_{i5})^{\top} \in \mathbb{R}^5$ is generated from multivariate normal distribution with mean $\mathbf{0}$ and covariance $\Sigma_x = (\sigma_{k_1 k_2})$, where $\sigma_{k_1 k_2} = 0.5^{|k_1 - k_2|}$. These variables can be viewed as some standardized quantitative measures for each node (e.g., income). In order to obtain the adjacency matrix A, we introduce different simulation models for the latent position Z_i and the predictors X_{ij} for every $1 \le i \ne j \le n$.

Model I. In this model, we let $X_{ij} = (V_{j1}, \cdots, V_{j5})^{\top} \in \mathbb{R}^5$. As a result, the relationship between i and j is only determined by the the receiver j. Furthermore, generate the latent position Z_i independently from a standard normal distribution for $1 \le i \le n$. At last, let $\beta = (0.5, 1, 0, -0.5, 0.1)^{\top}$. The adjacency matrix can be derived according to (2.1).

Under this setup, whether $A_{ij} = 1$ is only determined by the characteristic of node j. If j is a node of great importance, he/she will attract many other nodes. As a result, the nodal in-degree of j is relatively large (i.e., large A_{+j}). This will result in highly skewed distribution of nodal in-degree, which is also known as the phenomenon of *power-law* distribution. That is, most of the nodes in a network have very small nodal in-degree, while very few nodes (e.g., movie star or political leader) have a large amount of followers.

Model II. In this model, let $X_{ij} = ((V_{i1} + V_{j1})/2, \cdots, (V_{i5} + V_{j5})/2)^{\top}$. As a result, the relationship between i and j depends on both i and j, which represents the average on a certain kind of measurement between i and j. It can be regarded as social tightness. Furthermore, we generate the latent position from a mixture normal distribution, i.e., $0.5N(-1, 0.5^2) + 0.5N(1, 0.5^2)$. Hypothetically, nodes are clustered in the latent space around two different centers and are rarely overlapped. At last, let $\beta = (0.1, 0.8, 0, 0.5, -0.5)^{\top}$. The adjacency matrix can then be simulated accordingly.

Model III. In the last model, let $X_{ij} = (|V_{i1} - V_{j1}|, \cdots, |V_{i5} - V_{j5}|)^{\top}$. As a result, the predictor represents the difference between some quantitative measures between nodes i and j. At last, let $\beta = (0.6, -0.6, 0.1, 0, 0.5)^{\top}$. The adjacency matrix is then simulated accordingly.

4.2 Assessment criteria and simulation results

For each simulation model, the experiment is randomly replicated $M = 1000$ times. Denote $\hat{\beta}_R^{(m)}$ and $\hat{\beta}_T^{(m)}$ to be the estimators corresponding to reciprocity and transitivity in the mth replication, where $1 \le m \le M$. We consider the following measures to gauge the performances. First of all, for a given parameter β_j with $1 \le j \le p$, the root mean square error is calculated as $\mathrm{RMSE}_j = \{M^{-1} \sum_{m=1}^{M} (\hat{\beta}_j^{(m)} - \beta_j)^2\}^{1/2}$, where $\hat{\beta}_j^{(m)}$ is the jth element of $\hat{\beta}_R^{(m)}$ or $\hat{\beta}_T^{(m)}$. Next, for each $1 \le j \le p$, a 95% confidence interval is constructed for β_j as $\mathrm{CI}_j = (\hat{\beta}_j^{(m)} - z_{0.025} \widehat{\mathrm{SE}}_j^{(m)}, \hat{\beta}_j^{(m)} + z_{0.025} \widehat{\mathrm{SE}}_j^{(m)})$, where $\widehat{\mathrm{SE}}_j^{(m)}$ is defined in the subsection of 3.4, i.e., the estimated standard error. Then the coverage probability (CP) for the jth parameter can be computed as $\mathrm{CP}_j = M^{-1} \sum_{m=1}^{M} I(\beta_j \in \mathrm{CI}_j^{(m)})$. Lastly, network statistics are also reported such as network density (ND, $\{n(n-1)\}^{-1} \sum_{i,j} A_{ij}$), the number of reciprocated pairs (NR) and the number of transitivity (NT).

The detailed simulation results are reported in Tables 1–3. Because the simulation results are similar for all the models, we focus on Model I (i.e., Table 1) for interpretation. Note that the top panel reports the simulation results based on reciprocity, and the bottom panel displays the results derived on transitivity. First of all, for a given δ, as the network size n increases, the values of RMSE for all the parameters decrease towards 0. Furthermore, the reported coverage probability values for each parameter are all close to the nominal level 95%. This implies that the estimated standard error approximates the true one very well. Moreover, network density drops slowly towards 0 as the network size increases, which indicates that the network structure is increasingly sparse. However, the number of reciprocated pairs and the number of transitivities increase along with the network size, which means that the effective sample size is increasing for parameter estimation.

Remark 5. We conduct more simulation studies for the comparison between the LSLR and LR model. First, when network density is moderate, parameter estimations for the LSLR and LR model are very similar to each other. Second, when network density is very low, e.g., under 10%, it can be detected that the LR model does not work and a biased estimation result can be found. This is because that when network density is very low, the number of positive response (i.e., number of observed edges) is extremely small. In addition, under the simulation setup, the true model is assumed to be the LSLR and the LR model ignores the effect of the latent space. This makes the LR model fail for sparse networks. In contrast, the estimation procedure we propose for the LSLR model takes advantages of network typologies (e.g., transitivity), which makes our model useful for sparse networks.

4.3 Real data analysis

To demonstrate the usefulness of our proposed methodology, we study a real example in this subsection. The data are collected form Sina Weibo (www.weibo.cn), which is the largest social network platform in China. The dataset con-

Table 1. Simulation Results for Model I with 1000 replications. The root mean square error (RMSE) values are reported for every β estimates. The corresponding coverage probability (CP) in percentage of every estimate is given in the parentheses. Network density (ND), the number of reciprocated pairs (NR) and the number of transitivities (NT) are also reported

δ	Network Size	ND (%)	NR ($\times 10^3$)	NT ($\times 10^3$)	β_1	β_2	β_3	β_4	β_5
					Reciprocity				
1.5	500	2.4	1.0	28.6	0.051(95.0)	0.061(95.0)	0.054(95.0)	0.056(95.0)	0.050(93.1)
	1000	1.4	2.5	81.6	0.033(94.9)	0.040(94.9)	0.037(93.3)	0.037(94.4)	0.031(94.3)
	2000	0.8	5.9	231.8	0.021(94.3)	0.026(94.4)	0.022(95.3)	0.024(94.0)	0.021(94.4)
1.8	500	1.3	0.6	8.9	0.069(94.7)	0.079(95.3)	0.075(93.7)	0.075(95.2)	0.068(95.2)
	1000	0.7	1.2	20.6	0.047(94.2)	0.058(94.0)	0.048(96.3)	0.052(94.2)	0.043(96.0)
	2000	0.4	2.7	47.5	0.032(94.9)	0.038(95.6)	0.034(94.8)	0.035(95.1)	0.030(95.0)
2.0	500	2.2	1.0	25.8	0.051(95.6)	0.065(94.0)	0.056(94.2)	0.057(95.5)	0.050(94.6)
	1000	1.1	2.0	51.6	0.037(94.4)	0.045(95.1)	0.039(94.4)	0.042(94.4)	0.035(94.9)
	2000	0.6	4.0	103.8	0.025(95.9)	0.031(96.1)	0.027(95.0)	0.029(94.5)	0.025(95.1)
					Transitivity				
1.5	500	2.4	1.0	28.6	0.046(95.0)	0.061(94.9)	0.045(96.2)	0.049(95.4)	0.041(96.1)
	1000	1.4	2.5	81.6	0.029(95.6)	0.039(95.0)	0.030(96.2)	0.033(95.0)	0.026(95.7)
	2000	0.8	5.9	231.8	0.018(95.9)	0.025(94.1)	0.019(96.0)	0.021(94.6)	0.017(95.9)
1.8	500	1.3	0.6	8.9	0.065(96.3)	0.085(96.5)	0.065(96.6)	0.070(96.4)	0.062(95.7)
	1000	0.7	1.2	20.6	0.044(96.4)	0.055(95.4)	0.045(96.0)	0.050(96.1)	0.038(96.8)
	2000	0.4	2.7	47.5	0.030(95.8)	0.040(95.7)	0.030(96.0)	0.033(95.8)	0.028(94.9)
2.0	500	2.2	1.0	25.8	0.046(95.1)	0.062(96.4)	0.049(95.0)	0.051(95.2)	0.042(95.8)
	1000	1.1	2.0	51.6	0.032(96.8)	0.045(95.4)	0.034(95.9)	0.037(95.0)	0.030(95.7)
	2000	0.6	4.0	103.8	0.023(96.4)	0.031(95.6)	0.024(96.3)	0.025(96.5)	0.021(95.8)

sists of $n = 3{,}030$ users, who are all followers of an official account. The adjacency matrix A is defined as follows. If the ith node follows the jth node on the platform, then $A_{ij} = 1$, otherwise $A_{ij} = 0$. We always set $A_{ii} = 0$ for $1 \le i \le n$. The estimated density of this network is around 3.51%, which indicates a sparse network structure. Other than that, nodal in-degree and out-degree are calculated according to the definition of A_{+i} and A_{i+}. The histogram of nodal in- and out-degree is shown in the left panel of Figure 3, which exhibits the power-law phenomenon. At last, the network structure is depicted in the right panel of Figure 3.

To explain the network structure, we consider the following nodal information: V_1, the location of each node (Beijing = 1, others = 0). V_2, whether the node is authorized (authorized = 1, not authorized = 0). Authorization is the mechanism created by Weibo platform and authorized nodes are believed to be more popular with relatively much more fans. V_3, the log transformed number of posts made by each node, which reflect the activity of each node. As a result, the covariates are constructed as follows. First of all, $X_{ij,1} = I(V_{i1} = V_{j1})$, which indicates whether node i and j are both from Beijing. Possessing the same characteristics is known as *homophily* in social network analysis (Hunter et al., 2008). Next, we let $X_{ij,2} = V_{j2}$, which represent the popularity of the *receiver* (i.e., node j in A_{ij}). Furthermore, we let $X_{ij,3} = |V_{i3} - V_{j3}|$, which represents the difference between node i and j on their activity. At last, we define $X_{ij,4} = \sum_k A^0_{ik} A^0_{kj}$ as the number of common neighbors between i and j. Note that (A^0_{ij}) represents network structure a month ago. As a result, $X_{ij,4}$ can be seen as a measurement of social tightness formed between i and j.

The detailed estimation result is shown in Table 4. Note that the estimation procedure as well as link prediction do not rely on the choice of $f(\cdot)$ defined in Section 2.1. The results based on reciprocity and transitivity are slightly different from each other. For the estimation from reciprocity, it can be seen that under 5% level of significance, all the estimates are statistically significant. More specifically, we can make the following conclusions. First, nodes both from Beijing are more likely to be connected. Second, the more popular the receiver (i.e., j), the more likely to observe an edge between i and j (i.e., $A_{ij} = 1$). Thirdly, the larger the difference between two nodes on their activity, the less likely they are connected. Lastly, the closer nodes i and j are socially connected, the larger probability to form an edge between i and j. On the other hand, the results from transitivity show that receiver's popularity is not statistically significant. What's more, the effect of social tightness is weaker compared with the method of reciprocity. In order to decide which model to be used, one can conduct model selection from the perspective of prediction accuracy.

Table 2. Simulation Results for Model II with 1000 replications. The root mean square error (RMSE) values are reported for every β estimates. The corresponding coverage probability (CP) in percentage of every estimate is given in the parentheses. Network density (ND), the number of reciprocated pairs (NR) and the number of transitivities (NT) are also reported

δ	Network Size	ND (%)	NR ($\times 10^3$)	NT ($\times 10^3$)	β_1	β_2	β_3	β_4	β_5
Reciprocity									
1.5	500	2.4	1.2	27.2	0.079(95.3)	0.097(95.1)	0.090(94.6)	0.090(95.5)	0.082(95.5)
	1000	1.4	2.8	77.4	0.051(95.4)	0.063(95.1)	0.058(94.9)	0.058(96.0)	0.057(94.1)
	2000	0.8	6.7	218.9	0.033(95.2)	0.040(94.5)	0.037(95.4)	0.039(94.7)	0.034(95.8)
1.8	500	0.9	0.5	4.2	0.131(95.1)	0.155(95.5)	0.146(95.1)	0.156(93.5)	0.137(95.0)
	1000	0.5	1.0	9.7	0.086(96.0)	0.106(94.9)	0.097(94.8)	0.100(95.6)	0.089(95.1)
	2000	0.3	2.1	22.4	0.062(94.8)	0.072(94.6)	0.068(94.6)	0.071(93.3)	0.063(93.4)
2.0	500	2.3	1.1	24.2	0.083(95.1)	0.099(95.6)	0.096(94.6)	0.098(93.7)	0.087(94.4)
	1000	1.1	2.2	48.9	0.057(95.6)	0.069(95.5)	0.064(95.7)	0.067(95.1)	0.058(95.5)
	2000	0.6	4.5	98.2	0.043(93.5)	0.050(94.4)	0.047(94.5)	0.047(94.5)	0.043(94.3)
Transitivity									
1.5	500	2.4	1.2	27.2	0.052(96.8)	0.067(95.1)	0.057(96.3)	0.061(95.5)	0.054(95.6)
	1000	1.4	2.8	77.4	0.034(95.4)	0.043(94.5)	0.038(96.5)	0.040(96.3)	0.036(96.4)
	2000	0.8	6.7	218.9	0.023(95.6)	0.027(95.8)	0.025(95.1)	0.027(94.8)	0.024(94.2)
1.8	500	0.9	0.5	4.2	0.093(97.7)	0.121(95.6)	0.110(95.4)	0.116(95.4)	0.098(96.4)
	1000	0.5	1.0	9.7	0.064(96.4)	0.081(95.8)	0.072(96.7)	0.075(96.2)	0.068(96.6)
	2000	0.3	2.1	22.4	0.046(95.2)	0.053(97.2)	0.048(96.5)	0.050(97.0)	0.046(96.0)
2.0	500	2.3	1.1	24.2	0.055(95.5)	0.068(95.5)	0.061(95.9)	0.064(96.2)	0.059(94.5)
	1000	1.1	2.2	48.9	0.039(95.1)	0.047(95.5)	0.043(96.1)	0.046(95.6)	0.041(95.7)
	2000	0.6	4.5	98.2	0.027(96.3)	0.034(95.2)	0.030(95.9)	0.031(96.8)	0.030(94.8)

In order to evaluate the prediction accuracy, the nodes are randomly separated into two parts. 1,000 nodes and their corresponding relationships are chosen to form the testing set, the left 2,030 are treated as training. Out-sample AUC can then be calculated for the methods of LSLR, CN, and logistic regression model. The experiment is randomly repeated for 100 times and the averaged AUC values are reported in Table 5. It can be seen that, the LSLR model together with the transitivity method achieve best out-sample AUC performance (i.e., more than 90%) than the other two CN methods.

5. CONCLUDING REMARKS

Logistic regression, as one of the most important data analysis tools, is widely used in a number of scientific and industrial fields. Due to its simplicity and interpretability, logistic regression is commonly applied to conduct link prediction in social science (Hoff et al., 2002; Herz, 2015; Kruse et al., 2016). Our newly proposed LSLR model can be seen as an essential extension of the logistic regression model for two reasons. First, we incorporate latent social space information in the logistic regression model, where nodes close to each other in the latent social space are more likely to be connected with each other. Second, useful network characteristics (i.e., sparsity, reciprocity, and transitivity) can be guaranteed under the setup of the LSLR model. Two PMLEs are constructed for estimating the unknown parameters in the LSLR model efficiently. Extensive simulation studies and a real example analysis both demonstrate the usefulness of the LSLR model.

We discuss here a number of interesting topics for future study. First, the LSLR model introduces a flexible framework for modeling network structure. Extending the framework to include other network topologies (e.g., community structure, dynamic networks and so forth) is an interesting topic worthwhile pursuing. Second, along with the development of technology, we are able to collect more abundant user information. Therefore, the number of nodal covariates (i.e., V_i) can be of high dimension. For example, it can be user self-created labels (Huang et al., 2016). Then, how to incorporate with such type of non-structural and high-dimensional data can be challenging and interesting. Third, we estimate the regression coefficients in the LSLR model by a pseudo maximum likelihood method, ignoring the distance between nodes in the latent space. It is then of great interest to predict the distance so that visualization can be performed to get more intuitive explanations.

Table 3. Simulation Results for Model III with 1000 replications. The root mean square error (RMSE) values are reported for every β estimates. The corresponding coverage probability (CP) in percentage of every estimate is given in the parentheses. Network density (ND), the number of reciprocated pairs (NR) and the number of transitivities (NT) are also reported

δ	Network Size	ND (%)	NR ($\times 10^3$)	NT ($\times 10^3$)	Parameter β_1	β_2	β_3	β_4	β_5
					Reciprocity				
1.5	500	3.0	1.8	51.3	0.062(94.6)	0.065(94.9)	0.057(96.0)	0.059(94.6)	0.061(94.5)
	1000	1.8	4.4	146.0	0.040(95.1)	0.042(95.2)	0.036(95.7)	0.037(94.9)	0.040(93.4)
	2000	1.1	10.4	413.7	0.025(95.5)	0.027(95.1)	0.024(95.0)	0.024(94.7)	0.024(95.3)
1.8	500	1.7	1.0	15.9	0.083(95.4)	0.089(94.8)	0.079(94.9)	0.081(93.9)	0.082(94.7)
	1000	0.9	2.2	36.9	0.055(95.8)	0.058(96.1)	0.054(94.4)	0.053(94.7)	0.055(94.9)
	2000	0.5	4.7	84.5	0.039(95.4)	0.041(94.6)	0.035(95.9)	0.036(95.4)	0.038(94.9)
2.0	500	2.9	1.7	46.1	0.063(95.7)	0.068(94.8)	0.063(93.5)	0.060(94.6)	0.064(93.7)
	1000	1.4	3.5	92.1	0.045(94.5)	0.048(94.5)	0.042(95.4)	0.043(94.3)	0.044(94.0)
	2000	0.7	7.0	185.2	0.031(95.4)	0.034(94.3)	0.031(94.6)	0.030(94.4)	0.031(93.9)
					Transitivity				
1.5	500	3.0	1.8	51.3	0.047(95.5)	0.042(94.6)	0.039(96.2)	0.041(94.9)	0.044(95.3)
	1000	1.8	4.4	146.0	0.031(95.7)	0.026(95.8)	0.025(95.8)	0.025(96.1)	0.028(95.7)
	2000	1.1	10.4	413.7	0.019(96.0)	0.017(95.6)	0.016(96.0)	0.016(95.4)	0.018(95.5)
1.8	500	1.7	1.0	15.9	0.063(96.3)	0.057(95.9)	0.053(96.3)	0.055(95.8)	0.062(95.7)
	1000	0.9	2.2	36.9	0.043(97.0)	0.038(97.3)	0.037(95.9)	0.038(96.3)	0.041(96.9)
	2000	0.5	4.7	84.5	0.030(97.2)	0.027(95.1)	0.025(97.1)	0.027(95.2)	0.029(95.6)
2.0	500	2.9	1.7	46.1	0.047(96.4)	0.042(95.5)	0.041(94.7)	0.040(96.8)	0.046(96.3)
	1000	1.4	3.5	92.1	0.034(95.4)	0.031(95.6)	0.029(95.2)	0.028(95.6)	0.033(95.6)
	2000	0.7	7.0	185.2	0.024(95.3)	0.022(95.1)	0.021(95.8)	0.020(96.3)	0.023(95.7)

Table 4. Detailed estimation results for real dataset

Covariate	Reciprocity Estimate	p-value	Transitivity Estimate	p-value
Intercept	−2.53	<.001	−3.18	<.001
Homophily of Location	0.27	<.001	0.24	<.001
Receiver's Popularity	0.50	<.001	0.09	0.159
Difference on No. of Posts	−0.12	<.001	−0.05	<.001
Social Tightness	0.36	<.001	0.09	<.001

Table 5. Out sample AUC values for different methods on Sina Weibo dataset

Method	LSLR	CN			LR
		Transitivity	2-in-star	2-out-star	
AUC	93.74%	93.53%	89.29%	85.29%	92.61%

APPENDIX A. PROOF OF THEOREM 1

The theorem conclusion can be proved in three steps. In the first step, we prove the consistency of the PMLE $\hat{\beta}_R$. In the second step, we establish the asymptotic normality of $Q_1^{(1)}(\beta)$, where $Q_1^{(1)}(\beta)$ is the first order derivative of the pseudo likelihood function (3.2). In the last step, we show the consistency of $Q_2^{(1)}(\beta)$, which is the second order derivative of the pseudo likelihood function. For notation simplicity, we denote $Q_1^{(1)}(\beta)$ and $Q_2^{(1)}(\beta)$ as $Q_1(\beta)$ and $Q_2(\beta)$ throughout the rest of the proof.

Step 1. Recall the pseudo likelihood function in (3.2), which is a convex function in β. As long as we can prove that there exists a $\sqrt{n}$-consistent local optimizer, it must be the global optimizer. By Fan and Li (2001), we know that this is implied by the following fact. For any arbitrary

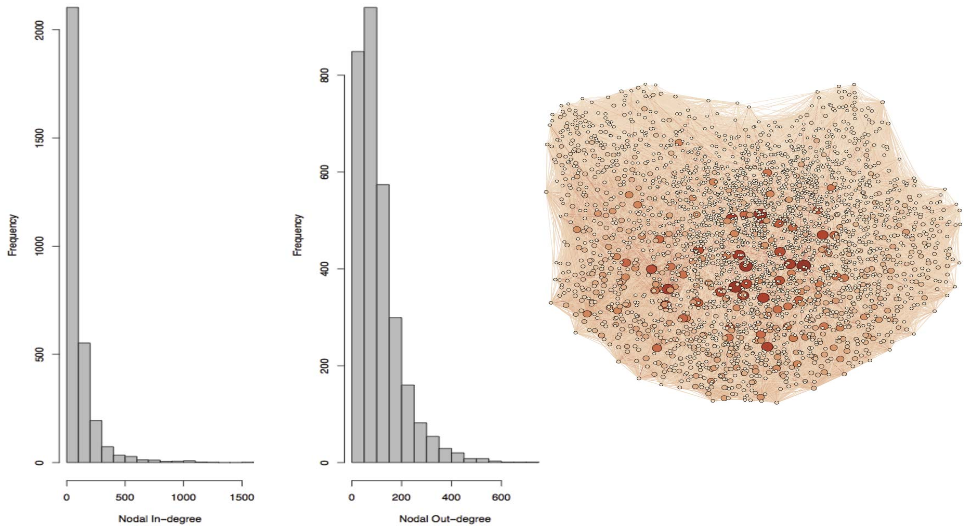

Figure 3. Left panel: histogram of nodal in-degree (left) and out-degree (right) of the Weibo data. The highly right skewed shape indicates that there may exist "super stars". Right panel: Network structure of the real dataset. Each dot represents a node, and each line represents an edge between two nodes. Larger size of the dot indicates higher number of nodal in-degrees.

small $\epsilon > 0$, there exists a sufficiently large constant C, such that

(A.1)

$$\lim_{n\to\infty} P\left[\sup_{\|\mathbf{u}\|=C}\left\{\ell_R(\beta+n^{-1/2}\mathbf{u})-\ell_R(\beta)\right\}<0\right]\geq 1-\epsilon,$$

where $\mathbf{u}=(u_1,\cdots,u_p)^\top\in\mathbb{R}^p$. Then by the standard argument on the Taylor expansion of the likelihood function, we have

(A.2)

$$\begin{aligned}
&n^{-(1-\delta/2)}\sup_{\|\mathbf{u}\|=C}\left\{\ell_R(\beta+n^{-1/2}\mathbf{u})-\ell_R(\beta)\right\}\\
&\quad= n^{-(3-\delta)/2}Q_1^\top(\beta)\mathbf{u}-\frac{1}{2}\mathbf{u}^\top\left\{-n^{-(2-\delta/2)}Q_2(\beta)\right\}\mathbf{u}+o_p(1)\\
&\quad\leq n^{-(3-\delta)/2}\|Q_1(\beta)\|C\\
&\qquad-\frac{1}{2}\lambda_{\min}\left\{-n^{-(2-\delta/2)}Q_2(\beta)\right\}C^2+o_p(1),
\end{aligned}$$

where $\lambda_{\min}(H)$ refers to the minimum eigenvalue of H. As we will prove in the next step, $n^{-(3-\delta)/2}Q_1(\beta)=O_p(1)$, which implies that $n^{-(3-\delta)/2}\|Q_1(\beta)\|=O_p(1)$. Similarly, we can prove that $\{-n^{-(2-\delta/2)}Q_2(\beta)\}\to_p H_1$, where H_1 is defined in Theorem 1 and we will prove in Step 3 that it is a positive definite matrix. Then the second term in (A.2) is quadratic in C. On the other hand, the first term in (A.2) is linear in C. Therefore, as long as the constant C is sufficient large, the second term will always dominate the other terms with arbitrary large probability. This implies the inequality (A.1), and thus completes the first step of the proof.

Step 2. In the second step, we show the asymptotic normality of $Q_1(\beta)$. To be more specific, we utilize the technique of U-statistics. Since the expectation of $Q_1(\beta)$ is not exactly 0, we first approximate it by $Q_1^*(\beta)$, a quantity with 0 expectation. Then we construct a U-statistic $\tilde{Q}_1^*(\beta)$ to approximate $Q_1^*(\beta)$. At last, the asymptotic normality of $Q_1(\beta)$ can be approximated by that of $\tilde{Q}_1^*(\beta)$.

Step 2.1 Recall that $\gamma_0=1-\sqrt{1/2}$ and we denote $p_{ij}=\exp(X_{ij}^\top\beta)/\{1+\exp(X_{ij}^\top\beta)\}$. As a result,

(A.3)

$$\begin{aligned}
Q_1(\beta)&=\partial\ell_R(\beta)/\partial\beta\\
&=\sum_{i\neq j}\left\{A_{ij}A_{ji}-A_{ji}\frac{e^{X_{ij}^\top\beta}}{1+e^{X_{ij}^\top\beta}}\right.\\
&\qquad\left.+(1-A_{ij})A_{ji}\frac{\gamma_0 e^{X_{ij}^\top\beta}}{1+\gamma_0 e^{X_{ij}^\top\beta}}\right\}X_{ij}\\
&=\sum_{i\neq j}\left\{A_{ij}A_{ji}\frac{1}{1+\gamma_0 e^{X_{ij}^\top\beta}}-A_{ji}p_{ij}\frac{1-\gamma_0}{1+\gamma_0 e^{X_{ij}^\top\beta}}\right\}X_{ij}\\
&=\sum_{i\neq j}s_{ij}(\beta).
\end{aligned}$$

Due to (2.2) and (2.3), we denote

$$\begin{aligned}
W_{ij}&=E(s_{ij}|\mathbb{X})\\
&=\frac{X_{ij}}{1+\gamma_0 e^{X_{ij}^\top\beta}}\left\{P(A_{ij}A_{ji}=1|\mathbb{X})\right.\\
&\qquad\left.-(1-\gamma_0)p_{ij}P(A_{ji}=1|\mathbb{X})\right\}\\
&=\eta_0\tilde{X}_{ij},
\end{aligned}$$

where $\tilde{X}_{ij} = p_{ij}p_{ji}X_{ij}/\{1+\gamma_0\exp(X_{ij}^\top\beta)\}$ and $\eta_0 = O(n^{-\delta})$. We then write

$$\text{(A.4)}\quad Q_1 = \sum_{i\neq j}(s_{ij} - W_{ij}) + \sum_{i\neq j} W_{ij} = \sum_{i\neq j}\tilde{s}_{ij} + \sum_{i\neq j} W_{ij} = Q_1^* + \sum_{i\neq j} W_{ij},$$

where $\tilde{s}_{ij} = s_{ij} - W_{ij}$. It is known that $E(\tilde{s}_{ij}) = 0$. We want to prove that Q_1 can be approximated by Q_1^*. To this end, we need to show

$$\text{(A.5)}\qquad \Big\{\text{cov}^{-1/2}\Big(Q_1^*\Big)\Big\}\Big\{E\Big(\sum_{i\neq j} W_{ij}\Big)\Big\} \;\to\; 0,$$

$$\text{(A.6)}\quad \Big\{\text{cov}^{-1/2}\Big(Q_1^*\Big)\Big\}\Big\{\text{cov}^{1/2}\Big(\sum_{i\neq j} W_{ij}\Big)\Big\} \;\to\; 0.$$

Step 2.1.1 In this step, we show the expectation and covariance of $\sum_{i\neq j} W_{ij}$.

First of all, note that $E(\sum_{i\neq j} W_{ij}) = \sum_{i\neq j} E(W_{ij}) = \eta_0 \sum_{i\neq j} E(\tilde{X}_{ij})$. Since $\eta_0 = O(n^{-\delta})$, along with condition (C1), it can be verified that $E(\sum_{i\neq j} W_{ij}) = O(n^{2-\delta})$.

We next study $\text{cov}(\sum_{i\neq j} W_{ij})$. By condition (C1), the covariance of $\sum_{i\neq j} W_{ij}$ is

$$\begin{aligned}\text{cov}\Big(\sum_{i\neq j}\eta_0\tilde{X}_{ij}\Big) = \eta_0^2\Big\{&\sum_{i\neq j}\text{cov}(\tilde{X}_{ij}) + \sum_{i\neq j}\text{cov}(\tilde{X}_{ij},\tilde{X}_{ji})\\ &+ \sum_{i\neq j}\sum_{k\neq i,j}\text{cov}(\tilde{X}_{ij},\tilde{X}_{ik})\\ &+ \sum_{i\neq j}\sum_{k\neq i,j}\text{cov}(\tilde{X}_{ji},\tilde{X}_{ki})\\ &+ 2\sum_{i\neq j}\sum_{k\neq i,j}\text{cov}(\tilde{X}_{ji},\tilde{X}_{ik})\Big\}.\end{aligned}$$

Recall that $\eta_0 = O(n^{-\delta})$. As a result, $\text{cov}(\sum_{i\neq j} W_{ij})$ is $O(n^{3-2\delta})$.

Step 2.1.2 In this step, we focus on the covariance of $Q_1^* = \sum_{i\neq j}\tilde{s}_{ij}$.

Note that $\text{cov}(Q_1^*) = E\{\text{cov}(Q_1^*|\mathbb{X})\} + \text{cov}\{E(Q_1^*|\mathbb{X})\}$, and $E(Q_1^*|\mathbb{X}) = 0$. So that $\text{cov}(Q_1^*) = E\{\text{cov}(Q_1^*|\mathbb{X})\}$. We then study $\text{cov}(Q_1^*|\mathbb{X})$, which is equal to

$$\text{(A.7)}\quad\begin{aligned}\text{cov}\Big(\sum_{i\neq j}\tilde{s}_{ij}|\mathbb{X}\Big) &= \sum_{i\neq j}\text{cov}\big(\tilde{s}_{ij}|\mathbb{X}\big) + \sum_{i\neq j}\text{cov}\big(\tilde{s}_{ij},\tilde{s}_{ji}|\mathbb{X}\big)\\ &+ \sum_{i\neq j}\sum_{k\neq i,j}\text{cov}\big(\tilde{s}_{ij},\tilde{s}_{ik}|\mathbb{X}\big)\\ &+ \sum_{i\neq j}\sum_{k\neq i,j}\text{cov}\big(\tilde{s}_{ji},\tilde{s}_{ki}|\mathbb{X}\big)\\ &+ 2\sum_{i\neq j}\sum_{k\neq i,j}\text{cov}\big(\tilde{s}_{ji},\tilde{s}_{ik}|\mathbb{X}\big)\\ &= M_1 + M_2 + M_{31} + M_{32} + 2M_{33}.\end{aligned}$$

We next study the terms in (A.7) one by one. Recall that $\tilde{s}_{ij} = s_{ij} - W_{ij} = \{A_{ij}A_{ji} - (1-\gamma_0)A_{ji}p_{ij} - \eta_0 p_{ij}p_{ji}\}X_{ij}/(1+\gamma_0 e^{X_{ij}^\top\beta})$ and $E(\tilde{s}_{ij}|\mathbb{X}) = 0$. As a result, $\text{cov}(\tilde{s}_{ij}|\mathbb{X}) = E(\tilde{s}_{ij}\tilde{s}_{ij}^\top|\mathbb{X})$. After simple calculation, it can be proved that $\text{cov}(\tilde{s}_{ij}|\mathbb{X}) = O_p(n^{-\delta/2})$. As a result, $M_1 = O_p(n^{2-\delta/2})$. Similarly, it can be shown that $M_2 = O_p(n^{2-\delta/2})$.

In order to derive the order of M_{31} to M_{33}, we need to utilize the following results,

$$\begin{aligned}&P(A_{ij}A_{ik} = 1|\mathbb{X})\\ &= p_{ij}p_{ik}\int\int e^{-\alpha_n(\|\mathbf{u}\|^2+\|\mathbf{v}\|^2)}g(\mathbf{u},\mathbf{v})d\mathbf{u}d\mathbf{v} = O(n^{-\delta}),\\ &P(A_{ij}A_{ji}A_{ki} = 1|\mathbb{X})\\ &= p_{ij}p_{ji}p_{ki}\int\int e^{-\alpha_n(2\|\mathbf{u}\|^2+\|\mathbf{v}\|^2)}g(\mathbf{u},\mathbf{v})d\mathbf{u}d\mathbf{v} = O(n^{-\delta}),\\ &P(A_{ij}A_{ji}A_{ik}A_{ki} = 1|\mathbb{X})\\ &= p_{ij}p_{ji}p_{ik}p_{ki}\int\int e^{-2\alpha_n(\|\mathbf{u}\|^2+\|\mathbf{v}\|^2)}g(\mathbf{u},\mathbf{v})d\mathbf{u}d\mathbf{v} = O(n^{-\delta}).\end{aligned}$$

It can then be derived that M_{31} to M_{33} are both of the order $O_p(n^{3-\delta})$. Under Condition (C2), the leading term is of $O(n^{3-\delta})$, and $\text{cov}(Q_1^*) = E\{\text{var}(Q_1^*|\mathbb{X})\} = O(n^{3-\delta})$.

Recall that we intend to prove (A.5) and (A.6). Under Condition (C2), we know that $2-\delta < (3-\delta)/2$ and $3-2\delta < 3-\delta$. As a result, Q_1 can be well approximated by Q_1^*.

Step 2.2 We have shown that Q_1 can be well approximated by $Q_1^* = \sum_{i\neq j}\tilde{s}_{ij}$. In this step, we construct a U-statistic to demonstrate the normality of Q_1. To this end, define $m_i = E(\tilde{s}_{ij}|Z_i,\mathbb{X})$ and $h_j = E(\tilde{s}_{ij}|Z_j,\mathbb{X})$. Then, let $\tilde{Q}_1^* = \sum_{i\neq j}(m_i + h_j)$. We next intend to show that Q_1^* can be well approximated by $\tilde{Q}_1^*$. To this end, we compare $\text{cov}(Q_1^*|\mathbb{X})$ and $\text{cov}(D|\mathbb{X})$, where $D = Q_1^* - \tilde{Q}_1^*$.

Remember that $\text{cov}(Q_1^*|\mathbb{X})$ has been studied in Step 2.1.2, which is $O_p(n^{3-\delta})$. We then only need to study $\text{cov}(D|\mathbb{X}) = \text{cov}\{\sum_{i\neq j}(\tilde{s}_{ij} - m_i - h_j|\mathbb{X})\}$, which is

$$\begin{aligned}&\sum_{i\neq j}\text{cov}\big(\tilde{s}_{ij} - m_i - h_j|\mathbb{X}\big)\\ &+ \sum_{i\neq j}\text{cov}\big(\tilde{s}_{ij} - m_i - h_j, \tilde{s}_{ji} - m_j - h_i|\mathbb{X}\big)\\ &+ \sum_i\sum_{j\neq i}\sum_{k\neq i,j}\Big\{\text{cov}\big(\tilde{s}_{ij} - m_i - h_j, \tilde{s}_{ik} - m_i - h_k|\mathbb{X}\big)\\ &\qquad + \text{cov}\big(\tilde{s}_{ji} - m_j - h_i, \tilde{s}_{ki} - m_k - h_i|\mathbb{X}\big)\\ &\qquad + 2\text{cov}\big(\tilde{s}_{ji} - m_j - h_i, \tilde{s}_{ik} - m_i - h_k|\mathbb{X}\big)\Big\}.\end{aligned}$$

Define the last three terms in the above equation by D_1, D_2 and D_3 respectively. We then have $\text{cov}\big(\tilde{s}_{ij} - m_i - h_j, \tilde{s}_{ik} - m_i - h_k|\mathbb{X}\big) = E(\tilde{s}_{ij}\tilde{s}_{ik}^\top|\mathbb{X}) - E(\tilde{s}_{ij}m_i^\top|\mathbb{X}) - E(\tilde{s}_{ij}h_k^\top|\mathbb{X}) - E(m_i\tilde{s}_{ik}^\top|\mathbb{X}) + E(m_i m_i^\top|\mathbb{X}) + E(m_i h_k^\top|\mathbb{X}) - E(h_j\tilde{s}_{ik}^\top|\mathbb{X}) + E(h_j m_i^\top|\mathbb{X}) + E(h_j h_k^\top|\mathbb{X})$. One can verify that $E(\tilde{s}_{ij}\tilde{s}_{ik}^\top|\mathbb{X}) = E\{E(\tilde{s}_{ij}\tilde{s}_{ik}^\top|\mathbb{X},Z_i)|\mathbb{X}\} = E\{E(\tilde{s}_{ij}|\mathbb{X},Z_i)E(\tilde{s}_{ik}^\top|\mathbb{X},Z_i)|\mathbb{X}\} =$

$E(m_i m_i^\top|\mathbb{X})$. Similarly, we can prove $E(s_{ij}m_i^\top|\mathbb{X}) = E(m_i m_i^\top|\mathbb{X})$ and $E(h_j h_k^\top|\mathbb{X}) = 0$. Consequently, we have $D_1 = 0$. The same technique can be used to derive that $D_2 = 0$ and $D_3 = 0$. Use similar techniques in Step 2.1.2, it can be verified that $\mathrm{cov}(D|\mathbb{X})$ is a smaller order as compared with $\mathrm{cov}(Q_1^*|\mathbb{X})$. This suggests that Q_1^* and $\tilde{Q}_1^*$ share the same asymptotic distribution.

Note that $\tilde{Q}_1^*$ can also be written as $\tilde{Q}_1^* = \sum_{i\neq j}(m_i + h_j) = (n-1)\sum_{i=1}^n (m_i + h_i)$. It can be seen that, given $\mathbb{X}$, $\tilde{Q}_1^*$ can be treated as the sum of independent random variables. In order to study $\mathrm{cov}(\tilde{Q}_1^*|\mathbb{X})$, we need to obtain the order of $P(A_{ij} = 1|\mathbb{X}, Z_i)$. Using similar technique in (2.2), it can be derived that $P(A_{ij} = 1|\mathbb{X}, Z_i) = O(n^{-\delta/2})$. As a result, by the definition of $\tilde{s}_{ij}$, it can be proved that $m_i = E(\tilde{s}_{ij}|Z_i, \mathbb{X}) = O(n^{-\delta/2})$ and $h_j = E(\tilde{s}_{ij}|Z_j, \mathbb{X}) = O(n^{-\delta/2})$. As a result, the conditional covariance of $n^{-(3-\delta)/2}\tilde{Q}_1^*$ can be calculated as the follows,

$$\begin{aligned}
&\mathrm{cov}\Big(n^{-(3-\delta)/2}\tilde{Q}_1^*|\mathbb{X}\Big)\\
&\quad = n^{-(3-\delta)}(n-1)^2\sum_{i=1}^n \mathrm{cov}(m_i + h_i|\mathbb{X})\\
&\quad = n^{-(3-\delta)}(n-1)^2\sum_{i=1}^n E\Big\{(m_i + h_i)(m_i + h_i)^\top|\mathbb{X}\Big\} \to_p C_1,
\end{aligned}$$

where C_1 is some positive definite matrix.

Step 3. In this step, we study the second order derivative of the conditional likelihood function, i.e., $Q_2(\beta)$. First of all, it can be derived that

$$\begin{aligned}
Q_2 = -\sum_{i\neq j}\Bigg\{&A_{ij}A_{ji}\frac{\gamma_0 e^{X_{ij}^\top\beta}}{\big(1+\gamma_0 e^{X_{ij}^\top\beta}\big)^2}\\
&+ A_{ji}p_{ij}\frac{(1-\gamma_0)(1-\gamma_0 e^{2X_{ij}^\top\beta})}{(1+e^{X_{ij}^\top\beta})\big(1+\gamma_0 e^{X_{ij}^\top\beta}\big)^2}\Bigg\}X_{ij}X_{ij}^\top.
\end{aligned}$$

According to (2.2) and (2.3), it can be proved that $E(Q_2|\mathbb{X})$ is $O_p(n^{2-\delta/2})$. As a result, $\{-n^{-(2-\delta/2)}Q_2\} \to_p H_1$, where H_1 is defined in Theorem 1. Along with the result derived in Step 2, it can be concluded that $\sqrt{n}(\hat{\beta}_R - \beta) \to_d N(0, H_1^{-1}C_1H_1^{-1})$. This completes the proof of Theorem 1.

APPENDIX B. PROOF OF THEOREM 2

Similar to Theorem 1, the conclusion in Theorem 2 can be proved in three steps. In the first step, we prove the consistency of the pseudo maximum likelihood estimator $\hat{\beta}_T$. In the second step, we establish the asymptotic normality of $Q_1^{(2)}(\beta)$, where $Q_1^{(2)}(\beta)$ is the first order derivative of the pseudo likelihood function (3.3). In the last step, we show the consistency of $Q_2^{(2)}(\beta)$, which is the second order derivative of the pseudo likelihood function. For notation simplicity, we denote $Q_1^{(2)}(\beta)$ and $Q_2^{(2)}(\beta)$ as $Q_1(\beta)$ and $Q_2(\beta)$ throughout the rest of the proof.

Step 1. Recall the pseudo likelihood function in (3.3), which is a convex function in β. As long as we can prove that there exists a $\sqrt{n}$-consistent local optimizer, it must be the global optimizer. By Fan and Li (2001), we know that this is implied by the following fact. For any arbitrary small $\epsilon > 0$, there exists a sufficiently large constant C, such that

(B.1)
$$\lim_{n\to\infty} P\left[\sup_{\|\mathbf{u}\|=C}\Big\{\ell_T(\beta + n^{-1/2}\mathbf{u}) - \ell_T(\beta)\Big\} < 0\right] \geq 1-\epsilon,$$

where $\mathbf{u} = (u_1, \cdots, u_p)^\top \in \mathbb{R}^p$. Then by the standard argument on the Taylor expansion of the likelihood function, we have

(B.2)
$$\begin{aligned}
&n^{-(2-\delta)}\sup_{\|\mathbf{u}\|=C}\Big\{\ell_T(\beta + n^{-1/2}\mathbf{u}) - \ell_T(\beta)\Big\}\\
&\quad = n^{-(5/2-\delta)}Q_1^\top(\beta)\mathbf{u} - \frac{1}{2}\mathbf{u}^\top\Big\{-n^{-(3-\delta)}Q_2(\beta)\Big\}\mathbf{u} + o_p(1)\\
&\quad \leq n^{-(5/2-\delta)}\|Q_1(\beta)\|C\\
&\qquad - \frac{1}{2}\lambda_{\min}\{-n^{-(3-\delta)}Q_2(\beta)\}C^2 + o_p(1).
\end{aligned}$$

As we will prove in the next step, $n^{-(5/2-\delta)}Q_1(\beta)$ is an $O_p(1)$, which implies that $n^{-(5/2-\delta)}\|Q_1(\beta)\| = O_p(1)$. Similarly, we can prove that $\{-n^{-(3-\delta)}Q_2(\beta)\} \to_p H_2$, where H_2 is defined in Theorem 2 and we will prove in Step 3 that it is a positive definite matrix. Then the second term in (B.2) is quadratic in C. On the other hand, the first term in (B.2) is linear in C. Therefore, as long as the constant C is sufficient large, the second term will always dominate the other terms with arbitrary large probability. This implies the inequality (B.1), and thus completes the first step of the proof.

Step 2. In the second step, we show the asymptotic normality of $Q_1(\beta)$. To be more specific, we utilize the technique of U-statistics. Since the expectation of $Q_1(\beta)$ is not exactly 0, we first approximate it by $Q_1^*(\beta)$, a quantity with 0 expectation. Then we construct a U-statistic $\tilde{Q}_1^*(\beta)$ to approximate $Q_1^*(\beta)$. At last, the asymptotic normality of $Q_1(\beta)$ can be approximated by that of $\tilde{Q}_1^*(\beta)$.

Step 2.1 Recall that $\nu_0 = 1 - \sqrt{1/3}$ and $p_{ij} = \exp(X_{ij}^\top\beta)/\{1+\exp(X_{ij}^\top\beta)\}$. As a result,

(B.3)
$$\begin{aligned}
Q_1(\beta) &= \partial\ell_T(\beta)/\partial\beta\\
&= \sum_{i,j,k} A_{ik}A_{kj}\Bigg\{A_{ij} - \frac{e^{X_{ij}^\top\beta}}{1+e^{X_{ij}^\top\beta}}\\
&\qquad + (1-A_{ij})\frac{\nu_0 e^{X_{ij}^\top\beta}}{1+\nu_0 e^{X_{ij}^\top\beta}}\Bigg\}X_{ij}\\
&= \sum_{i,j,k} A_{ik}A_{kj}\Bigg\{A_{ij}\frac{1}{1+\nu_0 e^{X_{ij}^\top\beta}} - p_{ij}\frac{1-\nu_0}{1+\nu_0 e^{X_{ij}^\top\beta}}\Bigg\}X_{ij}
\end{aligned}$$

$$= \sum_{i,j,k} s_{ijk}(\beta).$$

We next denote

$$\begin{aligned} W_{ijk} &= E(s_{ijk}|\mathbb{X}) \\ &= \frac{X_{ij}}{1+\nu_0 e^{X_{ij}^\top \beta}} \Big\{ P(A_{ik}A_{kj}A_{ij}=1|\mathbb{X}) \\ &\qquad - (1-\nu_0)p_{ij} P(A_{ik}A_{kj}=1|\mathbb{X}) \Big\} \\ &= \eta_1 \tilde{X}_{ijk}, \end{aligned}$$

where $\tilde{X}_{ijk} = p_{ik}p_{kj}p_{ij}X_{ij}/\{1+\nu_0 \exp(X_{ij}^\top\beta)\}$ and $\eta_1 = O(n^{-3\delta/2})$. We then write

$$\text{(B.4)} \qquad \begin{aligned} Q_1 &= \sum_{i,j,k}(s_{ijk} - W_{ijk}) + \sum_{i,j,k} W_{ijk} \\ &= \sum_{i,j,k} \tilde{s}_{ijk} + \sum_{i,j,k} W_{ijk} = Q_1^* + \sum_{i,j,k} W_{ijk}, \end{aligned}$$

where $\tilde{s}_{ijk} = s_{ijk} - W_{ijk}$. It is known that $E(\tilde{s}_{ijk}) = 0$. We want to prove that Q_1 can be approximated by Q_1^*. To this end, we need to show

$$\text{(B.5)} \qquad \Big\{\text{cov}^{-1/2}\Big(Q_1^*\Big)\Big\}\Big\{E\Big(\sum_{i,j,k} W_{ijk}\Big)\Big\} \quad \to \quad 0,$$

$$\text{(B.6)} \quad \Big\{\text{cov}^{-1/2}\Big(Q_1^*\Big)\Big\}\Big\{\text{cov}^{1/2}\Big(\sum_{i,j,k} W_{ijk}\Big)\Big\} \quad \to \quad 0.$$

Step 2.1.1 In this step, we show the expectation and covariance of $\sum_{i,j,k} W_{ijk}$.

First of all, note that $E(\sum_{i,j,k} W_{ijk}) = \sum_{i,j,k} E(W_{ijk}) = \eta_1 \sum_{i\neq j} E(\tilde{X}_{ijk})$. Since $\eta_1 = O(n^{-3\delta/2})$, along with condition (C1), it can be shown that $E(\sum_{i,j,k} W_{ijk}) = O(n^{3-3\delta/2})$.

We next study $\text{cov}(\sum_{i,j,k} W_{ijk})$, which is $\eta_1^2 \text{cov}(\sum_{i,j,k} \tilde{X}_{ijk})$. Recall that $\tilde{X}_{ijk}$ is a function of X_{ij}, X_{jk} and X_{ik}. Then by condition (C1), it can be verified that $\text{cov}(\sum_{i,j,k} \tilde{X}_{ijk})$ is $O(n^5)$. Recall that $\eta_1 = O(n^{-3\delta/2})$. As a result, $\text{cov}(\sum_{i,j,k} W_{ijk})$ is $O(n^{5-3\delta})$.

Step 2.1.2 In this step, we focus on the covariance of $Q_1^* = \sum_{i,j,k} \tilde{s}_{ijk}$.

Note that $\text{cov}(Q_1^*) = E\{\text{cov}(Q_1^*|\mathbb{X})\} + \text{cov}\{E(Q_1^*|\mathbb{X})\}$, and $E(Q_1^*|\mathbb{X}) = 0$. So that $\text{cov}(Q_1^*) = E\{\text{cov}(Q_1^*|\mathbb{X})\}$. We then study $\text{cov}(Q_1^*|\mathbb{X})$, which is equal to

(B.7)
$$\begin{aligned} \text{cov}\Big(\sum_{i,j,k}\tilde{s}_{ijk}|\mathbb{X}\Big) &= \sum_{i,j,k}\text{cov}\big(\tilde{s}_{ijk}|\mathbb{X}\big) + a_1 \sum_{i,j,k}\text{cov}\big(\tilde{s}_{ijk}, \tilde{s}_{jik}|\mathbb{X}\big) \\ &+ a_2 \sum_{i\neq j}\sum_{k_1\neq k_2}\text{cov}\big(\tilde{s}_{ijk_1}, \tilde{s}_{ijk_2}|\mathbb{X}\big) \\ &+ a_3 \sum_{i}\sum_{j_1\neq j_2}\sum_{k_1\neq k_2}\text{cov}\big(\tilde{s}_{ij_1k_1}, \tilde{s}_{ij_2k_2}|\mathbb{X}\big) \\ &= M_1 + M_2 + M_3 + M_4, \end{aligned}$$

where a_1, a_2 and a_3 are some constants. We next study the terms in (B.7) one by one. Recall that $\tilde{s}_{ijk} = s_{ijk} - W_{ijk} = \{A_{ik}A_{kj}A_{ij} - (1-\nu_0)A_{ik}A_{kj}p_{ij} - \eta_1 p_{ik}p_{kj}p_{ij}\}X_{ij}/(1+\nu_0 e^{X_{ij}^\top\beta})$ and $E(\tilde{s}_{ijk}|\mathbb{X}) = 0$. As a result, $\text{cov}(\tilde{s}_{ijk}|\mathbb{X}) = E(\tilde{s}_{ijk}\tilde{s}_{ji}^\top|\mathbb{X})$. After simple calculation, it can be proved that $\text{cov}(\tilde{s}_{ijk}|\mathbb{X}) = O_p(n^{-\delta})$. As a result, $M_1 = O_p(n^{3-\delta})$. Similarly, it can be shown that $M_2 = O_p(n^{3-\delta})$.

In order to derive the order of M_3 to M_4, we need to utilize the following results,

$$P(A_{ij}A_{ik_1}A_{ik_2}A_{k_1j}A_{k_2j} = 1|\mathbb{X}) = p_{ij}p_{ik_1}p_{ik_2}p_{k_1j}p_{k_2j}$$
$$\int\int\int e^{-\alpha_n(\|\mathbf{u}\|^2+\|\mathbf{v}\|^2+\|\mathbf{w}\|^2+\|\mathbf{u}-\mathbf{v}\|^2+\|\mathbf{u}-\mathbf{w}\|^2)}h(\mathbf{u},\mathbf{v},\mathbf{w})d\mathbf{u}d\mathbf{v}d\mathbf{w}$$
$$= O(n^{-3\delta/2}),$$
$$P(A_{ij_1}A_{ij_2}A_{ik_1}A_{ik_2}A_{k_1j_1}A_{k_2j_2} = 1|\mathbb{X}) = p_{ij_1}p_{ij_2}p_{ik_1}p_{ik_2}p_{k_1j_1}p_{k_2j_2}$$
$$\int\int\int\int\int e^{-\alpha_n(\|\mathbf{u_1}\|^2+\|\mathbf{u_2}\|^2+\|\mathbf{v_1}\|^2+\|\mathbf{v_2}\|^2+\|\mathbf{u_1}-\mathbf{u_2}\|^2+\|\mathbf{v_1}-\mathbf{v_2}\|^2)}$$
$$k(\mathbf{u_1},\mathbf{u_2},\mathbf{v_1},\mathbf{v_2})d\mathbf{u_1}d\mathbf{u_2}d\mathbf{v_1}d\mathbf{v_2} = O(n^{-2\delta}),$$

where $h(\cdot,\cdot,\cdot)$ and $k(\cdot,\cdot,\cdot,\cdot)$ are the joint probability density functions, satisfying $h(\mathbf{0},\mathbf{0},\mathbf{0}) > 0$ and $k(\mathbf{0},\mathbf{0},\mathbf{0},\mathbf{0}) > 0$. It can then be derived that M_3 to M_4 are of the order $O_p(n^{4-3\delta/2})$ and $O_p(n^{5-2\delta})$. By condition (C2), the leading term of $\text{cov}(Q_1^*|\mathbb{X})$ is M_4, which is of $O_p(n^{5-2\delta q})$. As a result $\text{cov}(Q_1^*) = E\{\text{var}(Q_1^*|\mathbb{X})\} = O(n^{5-2\delta q})$.

Recall that we intend to prove (B.5) and (B.6), By condition (C2), (B.5) and (B.6) are satisfied. As a result, Q_1 can be well approximated by Q_1^*, and $\text{cov}(Q_1^*) = O(n^{5-2\delta})$.

Step 2.2 We have shown that Q_1 can be well approximated by $Q_1^* = \sum_{i,j,k}\tilde{s}_{ijk}$. In this step, we construct a U-statistic to demonstrate the normality of Q_1. To this end, define $\tilde{m}_i = E(\tilde{s}_{ijk}|Z_i,\mathbb{X})$, $\tilde{h}_j = E(\tilde{s}_{ijk}|Z_j,\mathbb{X})$, and $\tilde{r}_k = E(\tilde{s}_{ijk}|Z_k,\mathbb{X})$. Then, let $\tilde{Q}_1^* = \sum_{i\neq j}\sum_{k\neq i,j}(\tilde{m}_i + \tilde{h}_j + \tilde{r}_k)$. We next intend to show that Q_1^* can be well approximated by $\tilde{Q}_1^*$. To this end, we compare $\text{cov}(Q_1^*|\mathbb{X})$ and $\text{cov}(D|\mathbb{X})$, where $D = Q_1^* - \tilde{Q}_1^*$.

Note that $\tilde{Q}_1^*$ can also be written as

$$\tilde{Q}_1^* = \sum_{i\neq j}\sum_{k\neq i,j}(\tilde{m}_i + \tilde{h}_j + \tilde{r}_k) = (n-1)(n-2)\sum_{i=1}^n(\tilde{m}_i + \tilde{h}_i + \tilde{r}_i).$$

It can be seen that, given $\mathbb{X}$, $\tilde{Q}_1^*$ can be treated as the sum of independent random variables. By similar techniques in Appendix A, the conditional covariance of $n^{-(3-\delta)/2}\tilde{Q}_1^*$ can be calculated as the follows,

$$\begin{aligned} &\text{cov}\Big(n^{-(5/2-\delta)}\tilde{Q}_1^*|\mathbb{X}\Big) \\ &= n^{-(5-2\delta)}(n-1)^2(n-2)^2\sum_{i=1}^n \text{cov}(\tilde{m}_i + \tilde{h}_i + \tilde{r}_i|\mathbb{X}) \\ &= n^{-(5-2\delta)}(n-1)^2(n-2)^2 \\ &\quad \sum_{i=1}^n E\Big\{(\tilde{m}_i + \tilde{h}_i + \tilde{r}_i)(\tilde{m}_i + \tilde{h}_i + \tilde{r}_i)^\top|\mathbb{X}\Big\} \to_p C_2, \end{aligned}$$

where C_2 is some positive definite matrix.

Step 3. In this step, we study the second order derivative of the conditional likelihood function, i.e., $Q_2(\beta)$. First of all, it can be derived that

$$Q_2 = -\sum_{i,j,k} A_{ik}A_{kj}\left\{A_{ij}\frac{\nu_0 e^{X_{ij}^\top\beta}}{(1+\nu_0 e^{X_{ij}^\top\beta})^2}\right.$$
$$\left.+ p_{ij}\frac{(1-\nu_0)(1-\nu_0 e^{2X_{ij}^\top\beta})}{(1+e^{X_{ij}^\top\beta})(1+\nu_0 e^{X_{ij}^\top\beta})^2}\right\}X_{ij}X_{ij}^\top.$$

According to the conclusions derived in *transitivity*, it can be proved that $E(Q_2|\mathbb{X})$ is $O_p(n^{3-\delta})$. As a result, $\{-n^{-(3-\delta)}Q_2\} \to_p H_2$, where H_2 is defined in Theorem 2. Along with the result derived in Step 2, it can be concluded that $\sqrt{n}(\hat{\beta}_T - \beta) \to_d N(0, H_2^{-1}C_2H_2^{-1})$. This completes the proof of Theorem 2.

Received 10 April 2020

REFERENCES

Aiello, L., Barrat, A., Schifanella R., Cattuto, C., Markines, B., and Menczer, F. (2012). Friendship prediction and homophily in social media. *ACM Transactions on the Web* **6**(2) 1–33.

Chang, X., Huang, D., and Wang, H. (2019). A popularity scaled latent space model for large-scale directed social network. *Statistica Sinica* **29**(3) 1277–1299. MR3932518

de Miguel Luken, V. and Tranmer, M. (2010). Personal support networks of immigrants to Spain: a multilevel analysis. *Social Networks* **32** 253–262.

Getoor, L. and Diehl, C. P. (2005). Link Mining: A Survey. *ACM SIGKDD Explorations Newsletter* **7**(2) 3–12.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96** 1348–1360. MR1946581

Han, X., Hsieh, C., and Ko, S. I. (2019). Spatial Modeling Approach for Dynamic Network Formation and Interactions. *Journal of Business & Economic Statistics* **2019** 1–16. MR4187179

Handcock, M. S., Raftery, A. E., and Tantrum, J. M. (2007). Model-based clustering for social networks. *Journal of the Royal Statistical Society, Series A* **170** 301–354. MR2364300

Herz, A. (2015). Relational constitution of social support in migrants' transnational personal communities. *Social Networks* **40** 64–74.

Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association* **97** 1090–1098. MR1951262

Holland, P. W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association* **76** 33–50. MR0608176

Huang, D., Yin, J., Shi, T., and Wang, H. (2016). A statistical model for social network labeling. *Journal of Business and Economics Statistics* **34**(3) 368–374. MR3523781

Hunter, D. R., Goodreau, S. M., and Handcock, M. S. (2008). Goodness of fit of social network models. *Journal of the American Statistical Association* **103** 248–258. MR2394635

Kruse, H., Smith, S., Tubergen, F., and Maas, I. (2016). From neighbors to school friends? How adolescents' place of residence relates to same-ethnic school friendships. *Social Networks* **44** 130–142.

Libennowell, D. and Kleinberg, J. M. (2007). The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology* **58** 1019–1031.

Lü, L., Jin, C. H., and Zhou, T. (2009). Similarity index based on local paths for link prediction of complex networks. *Physical Review E* **80** 046122.

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

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.

Newman, M. E. J. (2001). Clustering and preferential attachment in growing networks. *Physical Review E* **64**(2) 025102.

Oconnor, L., Medard, M., and Feizi, S. (2015). Maximum likelihood latent space embedding of logistic random dot product graphs. *Machine Learning*.

Rastelli, R., Friel, N., and Raftery, A. E. (2016). Properties of latent variable network models. *Network Science* **4**(4) 407–432.

Salter-Townshend, M. and Murphy, T. B. (2013). Variational Bayesian inference for the latent position cluster model for network data. *Computational Statistics and Data Analysis* **57**(1) 661–671. MR2981116

Schweinberger, M. and Snijders, T. A. B. (2003). Settings in social networks: a measurement model. *Sociological Methodology* **33** 307–341.

Sewell, D. K. and Chen, Y. (2015). Latent space models for dynamic networks. *Journal of the American Statistical Association* **110**(512) 1646–1658. MR3449061

Sewell, D. K. and Chen, Y. (2017). Latent space approaches to community detection in dynamic networks. *Bayesian Analysis* **12**(2) 351–377. MR3620737

Wang, P., Xu, B., Wu, Y., and Zhou, X. (2015). Link prediction in social networks: the state-of-the-art. *Science China: Information Sciences* **58** 1–38. MR3232207

Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*, Cambridge University Press.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature* **393** 440–442. MR1716136

Xing, E. P., Fu, W., and Song, L. (2010). A state-space mixed membership blockmodel for dynamic network tomography. *The Annals of Applied Statistics* **4**(2) 535–566. MR2758639

Yu, Y. and Wang, X. (2014). Link prediction in directed network and its application in microblog. *Mathematical Problems in Engineering* **1** 1–8.

Zhou, J., Tu, Y., Chen, Y., and Wang, H. (2017). Estimating spatial autocorrelation with sampled network data. *Journal of Business and Economic Statistics* **35**(1) 130–138. MR3591541

Rui Pan
Central University of Finance and Economics
China
E-mail address: panrui_cufe@126.com

Xiangyu Chang
Xi'an Jiaotong University
China
E-mail address: xiangyuchang@xjtu.edu.cn

Xuening Zhu
Fudan University
China
E-mail address: xueningzhu@fudan.edu.cn

Hansheng Wang
Peking University
China
E-mail address: hansheng@gsm.pku.edu.cn