

# A latent moving average model for network regression

RUI PAN\*, RONG GUAN<sup>†,§</sup>, XUENING ZHU,  
AND HANSHENG WANG<sup>‡</sup>

---

Different from traditional statistical analysis that concerns about individuals, network analysis focuses more on the dichotomous relationships between those individuals. It is then of interest to investigate the relationship against a set of predictive variables. The widely used generalized linear model is no longer applicable, since it implicitly assumes that different subjects are completely independent. To solve this problem, we propose a latent moving average model (LMAM), which allows for nontrivial dependence for overlapped relationships. It is only assumed that the non-overlapped relationships are independent. Under such an assumption, the asymptotic theory, including the rate of convergence and asymptotic normality, can be established. A number of numerical studies are conducted to demonstrate the finite sample performance of our proposed method. At last, a real dataset is analyzed for illustration purpose.

KEYWORDS AND PHRASES: Generalized linear model, Latent moving average model, Network regression, Social networks.

---

## 1. INTRODUCTION

In the past decade, the whole world has witnessed the tremendously rapid growth of various social network platforms. As a good example, Facebook ([www.facebook.com](http://www.facebook.com)) has attracted more than one billion active users across all over the world. As its Chinese counter part, Sina Weibo ([www.weibo.com](http://www.weibo.com)), the most popular microblog service in China, also gained more than one hundred million active users in the domestic market. As a result, network data are becoming increasingly available, which possess huge commercial values.

---

\*The research of Rui Pan is supported in part by National Natural Science Foundation of China (NSFC, 11601539, 11631003).

†The research of Rong Guan is supported in part by National Natural Science Foundation of China (NSFC, 71401192, 71771204), and Program for Innovation Research in Central University of Finance and Economics (011650317002).

‡The research of Hansheng Wang is supported in part by National Natural Science Foundation of China (NSFC, 71532001, 11525101), China's National Key Research Special Program grant 2016YFC0207700, and the Business Intelligence Research Center at Peking University.

§Corresponding author.

Inspired by their tremendous achievements, it is of great interest to develop novel statistical methods, which are useful for network analysis. Consider a social network with  $n$  players. Mathematically, one can use an index  $i$  ( $1 \leq i \leq n$ ) to denote the  $i$ th network player and then define a binary variable  $Y_{ij}$  to represent the social relationship between player  $i$  and  $j$ . Typically, we define  $Y_{ij} = 1$  if  $i$  is related to  $j$ , otherwise  $Y_{ij} = 0$  [10, 20]. For instance, if the  $i$ th player adds the  $j$ th as its friend on Facebook, then  $Y_{ij} = 1$ . For convenience, we always set  $Y_{ii} = 0$ . Note that the relationship between  $i$  and  $j$  can be both symmetric (i.e.,  $Y_{ij} = Y_{ji}$ ) and asymmetric (i.e.,  $Y_{ij}$  may not equal to  $Y_{ji}$ ).

Network data analysis is a classical and important research problem. Many researchers have made profound contributions to this field. For example, an exponential random graph model (typically referred to as  $p_1$  model) was developed by [8], which assumes that different dyads, i.e.,  $(Y_{ij}, Y_{ji})$ s, are independent. In order to allow for dyad dependence, a Markov graph model was considered by [4]. Later, stochastic block models were studied by [13] and [19]. As a computationally effective way, the method of pseudo-likelihood estimation was investigated by [18]. Subsequently, in a seminal work of [6], the concept of latent space was proposed, which incorporates predictive variables into a Bayesian modelling framework. It is extensively studied and popularly used recently [5, 11, 15]; however, to our best knowledge, none of the existing theory has been rigorously established for network regression under an asymptotic framework. That motivates us to accomplish this work.

In order to run a social network site successfully, one needs to have a good understanding about the statistical behavior of  $Y_{ij}$ . This amounts to building a regression relationship between  $Y_{ij}$  and  $X_{ij}$ , where  $X_{ij} = (X_{ij,1}, \dots, X_{ij,p})^\top \in \mathbb{R}^p$  is a  $p$ -dimensional predictive variable. Useful information for predicting  $Y_{ij}$  could be included in  $X_{ij}$ . For example, (1) whether player  $i$  has some distinguished features (e.g., a political leader), (2) whether  $i$  and  $j$  share some common characteristics (e.g., graduate from the same university), and (3) certain quantitative measures for the social tightness between  $i$  and  $j$  (e.g., the amount of phone calls made between  $i$  and  $j$ ); see [1] and [9] for more examples about the predictive variables. Consequently, developing regression methods for network data is becoming a problem of interest.

To model the relationship between  $Y_{ij}$  and  $X_{ij}$ , the classical method of generalized linear model [12] could be used in practice. The method implicitly assumes that given  $\mathbb{X}$ , different  $Y_{ij}$ s are mutually independent, where  $\mathbb{X} = (X_{ij})$  is the information set. For simplicity, we refer to this as a complete independence (CI) assumption. Obviously, CI assumption is very restrictive, since players in the network are usually related to each other through various relationships. In a recent work of [14], the authors proposed various test statistics for the CI assumption. When CI assumption is invalid, it is then of great interest to re-investigate the theoretical properties of the resulting maximum likelihood estimator (MLE). More specifically, we concern about the consistency of the MLE, the rate of convergence, and the asymptotic normality. All those important theoretical problems have not been thoroughly investigated in the past literature, according to our best knowledge.

To fulfill this theoretical gap, we propose a latent moving average model (LMAM). In contrast to the CI assumption, LMAM only requires non-overlapped relationships to be conditionally independent. This means that given  $\mathbb{X}$ , we assume  $Y_{ij}$  and  $Y_{kl}$  are mutually independent if  $\{i, j\} \cap \{k, l\} = \emptyset$ . We refer to this assumption as non-overlapped independence (NI) assumption throughout this work. As a result, nontrivial dependence is allowed, for instance, between  $Y_{ij}$  and  $Y_{ik}$  in a flexible manner. Under such an assumption, we find that the MLE is still consistent and asymptotically normal; however, the convergence rate under these two assumptions are different, which is to be demonstrated later. This immediately arises an important research problem. That is under the NI assumption, how to make asymptotically statistical inference.

The rest of the article is organized as follows: Section 2 illustrates the methodology, including the MLE and its asymptotic normality under the NI assumption. Section 3 presents a set of simulation studies and a real social network example. Concluding remarks are made in Section 4. All the technical proofs are left in the Appendix.

## 2. METHODOLOGY

### 2.1 The maximum likelihood estimator

Recall that  $Y_{ij}$  is the relationship indicator from  $i$  to  $j$ , and  $X_{ij}$  is the associated  $p$ -dimensional predictor. For convenience, we use  $\mathbb{Y} = (Y_{ij}) \in \mathbb{R}^{n \times n}$  to denote the  $n \times n$  adjacency matrix. In practice, the following model has been popularly used, that is  $P(Y_{ij}|\mathbb{X}) = g(X_{ij}^\top \beta)$ , where  $g(\cdot)$  is a pre-specified link function and  $\beta = (\beta_1, \dots, \beta_p)^\top \in \mathbb{R}^p$  is the unknown regression coefficient vector. For simplicity, we define  $p_{ij} = E(Y_{ij}|\mathbb{X}) = g(X_{ij}^\top \beta)$ . It is well known that two popular link functions have been extensively employed practically. They are, respectively, the logit model with  $g(t) = e^t/(1+e^t)$  and the probit model with  $g(t) = \Phi(t)$ , where  $\Phi(\cdot)$  is the cumulative distribution function of the standard normal distribution.

As we mentioned before, a full specification about the likelihood of  $\mathbb{Y}$  given  $\mathbb{X}$  is extremely difficult because the dependence structure between different  $Y_{ij}$ s could be rather complicated. Nevertheless, to facilitate an easy estimation, one can naïvely assume that given  $\mathbb{X}$ , different  $Y_{ij}$ s are mutually independent (i.e., CI assumption). Then the corresponding likelihood function can be constructed as the follows,

$$(1) \quad L(\beta) = \prod_{i \neq j} \left\{ g(X_{ij}^\top \beta) \right\}^{Y_{ij}} \left\{ 1 - g(X_{ij}^\top \beta) \right\}^{1-Y_{ij}}.$$

By maximizing the above objective function, an MLE can be obtained, which is denoted as  $\hat{\beta} = \operatorname{argmax}_{\beta} L(\beta)$ . However, the corresponding asymptotic properties under the NI assumption are still unknown and need to be investigated.

### 2.2 A latent moving average model

In order to obtain the asymptotic properties (i.e., consistency and asymptotic normality) of  $\hat{\beta}$ , certain type of independence assumption has to be made. The simplest assumption, as implicitly used by practitioners, would be the CI assumption. Under such an assumption, the optimal solution of (1) becomes the exact MLE. However, according to our limited empirical experiences, such an assumption is not appropriate for network data analysis.

As another interesting solution, we propose here a latent moving average model (i.e., LMAM). More specifically, LMAM assumes that there exists an unobserved and independent latent nodal information  $V_i$  for  $i = 1, \dots, n$ , such that the network relationship is determined as

$$(2) \quad Y_{ij} = G(X_{ij}, V_i, V_j, \varepsilon_{ij}),$$

where  $\varepsilon_{ij}$  is the random noise and is assumed to be independent of  $X_{ij}$ ,  $V_i$  and  $V_j$ . Furthermore,  $G$  is an unspecified function such that  $E(Y_{ij}|\mathbb{X}) = E\{G(X_{ij}, V_i, V_j, \varepsilon_{ij})|\mathbb{X}\} = g(X_{ij}^\top \beta)$ . Similar model was also used by [7], which is called the bilinear mixed-effects model. The author proposed an MCMC algorithm for model estimation. However, no theoretical results are further investigated.

From (2), on one hand, LMAM allows non-overlapped relationships to be independent. That is  $Y_{ij}$  and  $Y_{kl}$  with  $\{i, j\} \cap \{k, l\} = \emptyset$  are mutually independent given  $\mathbb{X}$ . On the other hand, overlapped relationships, that is  $Y_{ij}$  and  $Y_{kl}$  with  $\{i, j\} \cap \{k, l\} \neq \emptyset$ , are assumed to be mutually dependent given  $\mathbb{X}$ , because they share the same latent nodal information. This contains four important special cases, which are respectively, the reciprocated dependence (i.e.,  $Y_{ij}$  and  $Y_{ji}$ ), the 2-out-star dependence (i.e.,  $Y_{ij}$  and  $Y_{ik}$ ), the 2-in-star dependence (i.e.,  $Y_{ji}$  and  $Y_{ki}$ ), and the transitive dependence (i.e.,  $Y_{ij}$  and  $Y_{jk}$ ).

### 2.3 Asymptotic distribution

We consider first the asymptotic distribution of  $\hat{\beta}$  under the CI assumption. Under this situation, different  $Y_{ij}$ s

are assumed to be conditionally independent so that the classical theory of MLE can be applied directly [16]. As a result, the sample size is  $N = n(n - 1)$ . This gives  $\sqrt{N}(\hat{\beta} - \beta) \rightarrow_d N(0, H^{-1})$  as  $n \rightarrow \infty$ , where

$$(3) \quad H = E \left[ \frac{\dot{g}^2(X_{ij}^\top \beta)}{g(X_{ij}^\top \beta) \{1 - g(X_{ij}^\top \beta)\}} X_{ij} X_{ij}^\top \right],$$

and  $\dot{g}(\cdot)$  is the first order derivative of  $g(\cdot)$ . As one can see, under the CI assumption, the convergence rate of  $\hat{\beta}$  is very fast, which is  $\sqrt{N} \approx n$ . However, the convergence rate and asymptotic distribution of  $\hat{\beta}$  under the NI assumption are not clear. The answer is given by the following theorem.

**Theorem 1.** *Assume  $g(t)$  is a continuous function in  $t$ , where the first- and second-order derivative exist. Further assume  $\log(g(t))$  is concave in  $t$ . Under the LMAM (2), define  $m_i = E\{s_{ij}(\beta)|V_i, \mathbb{X}\}$ ,  $h_i = E\{s_{ji}(\beta)|V_i, \mathbb{X}\}$  and*

$$(4) \quad s_{ij}(\beta) = \dot{g}(X_{ij}^\top \beta) X_{ij} \left[ \frac{Y_{ij} - g(X_{ij}^\top \beta)}{g(X_{ij}^\top \beta) \{1 - g(X_{ij}^\top \beta)\}} \right].$$

Assume that  $n^{-1} \sum_{i=1}^n E\{(m_i + h_i)(m_i + h_i)^\top | \mathbb{X}\} \rightarrow_p C_0$  for some positive definite matrix  $C_0$ . Then,  $\sqrt{n}(\hat{\beta} - \beta) \rightarrow_d N(0, H^{-1} C_0 H^{-1})$ , as  $n \rightarrow \infty$ .

The proof of Theorem 1 is given in Appendix A. By Theorem 1, we know that the asymptotic behavior of  $\hat{\beta}$  might be totally different, according to whether the assumption of CI or NI holds. If CI assumption is correct, the convergence rate of  $\hat{\beta}$  is  $\sqrt{N} \approx n$ . In contrast, the convergence rate of  $\hat{\beta}$  reduces to the level  $\sqrt{n}$  under the NI assumption. In addition to that, the analytical formulas for the asymptotic variances are totally different. Accordingly, the resulting inference procedures under different independence assumptions should be distinguished. The details are given in the next subsection.

**Remark 1.** We would like to make a remark discussing the likelihood function under different dependence assumptions. First of all, under the CI assumption, the likelihood function can be easily spelled out, which is exactly (1). Second, under the NI assumption, the likelihood function is complicated, because the dependence between overlapped relationships is not specified. As a result, we still refer to the likelihood function in (1) as the one under the NI assumption. However, it is actually a pseudo likelihood function under the NI assumption.

**Remark 2.** It is not intuitive that the consistency result of  $\hat{\beta}$  only depends on the margin  $g(\cdot)$  but not  $G(\cdot)$ . This is mainly because the first-order derivative  $S = \sum_{i \neq j} s_{ij}$  can be well approximated by a  $U$ -statistic. As shown in the proof of Theorem 1, the  $U$ -statistic has the form  $\tilde{S} = \sum_{i \neq j} (m_i + h_j)$ , where  $m_i$  and  $h_j$  are conditional expectations only determined by  $g(\cdot)$ . As a result, only  $g(\cdot)$  matters for the consistency of the MLE.

**Remark 3.** In practice, we may know  $V_i$  and  $V_j$ . As a result, it is of great interest to model function  $G(\cdot)$ . Here, we give one particular scenario, where  $G(\cdot)$  is the indicator function such that  $Y_{ij} = I(X_{ij}^\top \beta + \delta_1 V_i + \delta_2 V_j + \varepsilon_{ij} > 0)$ . If we further assume that  $V_i$ s and  $\varepsilon_{ij}$ s are independently distributed as normal, then the margin  $g(\cdot)$  function is just the probit link. In this case, the asymptotic properties of  $\hat{\beta}$  can also be derived under the NI assumption. Since this is out of the scope of this article, we leave this as a future topic here.

## 2.4 Estimation for standard error

First of all, under the situation of CI assumption, the asymptotic variance of  $\sqrt{N}(\hat{\beta} - \beta)$  is given by  $H^{-1}$ , which can then be estimated by  $\hat{H}^{-1}$ , where

$$(5) \quad \hat{H} = \frac{1}{N} \sum_{i \neq j} \left[ \frac{\dot{g}^2(X_{ij}^\top \hat{\beta})}{g(X_{ij}^\top \hat{\beta}) \{1 - g(X_{ij}^\top \hat{\beta})\}} X_{ij} X_{ij}^\top \right].$$

To test the statistical significance of one particular regression coefficient, i.e.,

$$(6) \quad H_0 : \beta_j = 0 \text{ vs. } H_1 : \beta_j \neq 0,$$

a  $Z$ -type statistic can be constructed as  $Z_C = \hat{\beta}_j / \widehat{SE}_C(\hat{\beta}_j)$ , where  $\widehat{SE}_C^2(\hat{\beta}_j)$  is the  $j$ th diagonal element of  $\hat{H}^{-1}/N$ . Given a significant level  $\alpha$ , we reject the null hypothesis if  $|Z_C| > z_{1-\alpha/2}$ , where  $z_\alpha$  is the  $\alpha$ -th lower quantile of a standard normal distribution.

On the other hand, under the NI assumption, the asymptotic variance of  $\sqrt{n}(\hat{\beta} - \beta)$  is given by a sandwich form formula  $H^{-1} C_0 H^{-1}$  as shown in Theorem 1. It can then be estimated by  $\hat{H}^{-1} \hat{C}_0 \hat{H}^{-1}$ , where  $\hat{H}$  is given in (5), and

$$\begin{aligned} \hat{C}_0 &= \frac{1}{n} \sum_{i=1}^n (\hat{m}_i + \hat{h}_i)(\hat{m}_i + \hat{h}_i)^\top, \\ \hat{m}_i &= \frac{1}{n-1} \sum_{j \neq i} s_{ij}(\hat{\beta}), \\ \hat{h}_i &= \frac{1}{n-1} \sum_{j \neq i} s_{ji}(\hat{\beta}). \end{aligned}$$

Similarly, to test the hypothesis (6), a  $Z$ -type statistic can be constructed as  $Z_N = \hat{\beta}_j / \widehat{SE}_N(\hat{\beta}_j)$ , where  $\widehat{SE}_N^2(\hat{\beta}_j)$  is the  $j$ th diagonal element of  $\hat{H}^{-1} \hat{C}_0 \hat{H}^{-1}/n$ . As a result, for a given significance level  $\alpha$ , we reject the null hypothesis of (6) if  $|Z_N| > z_{1-\alpha/2}$ .

**Remark 4.** Note that for different assumptions (both CI and NI), only one estimation  $\hat{\beta}$  is utilized. It is defined as the MLE of (1), and numerically obtained by Newton-Raphson iterative algorithm. For different assumptions, the main difference lies in the inference procedure, i.e., different standard errors are used for the hypothesis test.

### 3. NUMERICAL STUDIES

#### 3.1 Simulation studies

To demonstrate the finite sample performance of the proposed methods, we present here a number of simulation studies. The network size is set to  $n=100, 200,$  and  $500,$  and a total of 1,000 replications are conducted.

We first generate a set of nodal information for each node,  $Z_i = (Z_{i1}, \dots, Z_{i6})^\top \in \mathbb{R}^6$  with  $1 \leq i \leq n$ . In particular,  $Z_{i1} \in \{0, 1\}$  is a binary random variable with  $P(Z_{i1} = 1) = 0.15$ . This variable can be treated as a rare feature (i.e., movie star or political leader). We then let  $X_{ij,1} = Z_{j1}$  so that the first predictor between nodes  $i$  and  $j$  only depends on the characteristic of node  $j$ . Next, we simulate  $Z_{i2}$  as a discrete random variable, taking the values of 0, 1, 2 with probabilities 0.1, 0.3 and 0.6 respectively.  $Z_{i2}$  can be seen as education level, which takes a number of discrete values. Accordingly,  $X_{ij,2}$  is designed to represent the effect of homophily, i.e.,  $X_{ij,2} = I(Z_{i2} = Z_{j2})$ , where  $I(\cdot)$  is the indicator function. Lastly, we simulate  $(Z_{i3}, \dots, Z_{i6})^\top \in \mathbb{R}^4$  from a multivariate normal distribution with mean 0 and covariance  $\Sigma = (\sigma_{j_1 j_2}) \in \mathbb{R}^{4 \times 4}$ , where  $\sigma_{j_1 j_2} = 0.5^{|j_1 - j_2|}$ . In practice,  $(Z_{i3}, \dots, Z_{i6})^\top$  can be viewed as the centralized quantitative measures for node  $i$  in social network analysis, e.g., age, income, and many others. We further let  $(X_{ij,3}, X_{ij,4})^\top = (Z_{i3} + Z_{j3}, Z_{i4} + Z_{j4})^\top / 2$ , and  $(X_{ij,5}, X_{ij,6})^\top = (|Z_{i5} - Z_{j5}|, |Z_{i6} - Z_{j6}|)^\top$ . They represent the average value and the difference of quantitative measures between nodes  $i$  and  $j$  respectively. As a result, the predictor dimension  $p$  is fixed to be 6.

After deriving the predictive variables  $X_{ij}$ , we generate the latent nodal information  $V_i$  independently from a standard normal distribution for  $1 \leq i \leq n$ . As a result, network dependence can be considered in our model by letting  $V_{ij} = V_i + V_j$ . Accordingly, different  $Y_{ij}$ s are independently generated by

$$(7) \quad Y_{ij} = I(X_{ij}^\top \beta + \delta V_{ij} / \sqrt{2} + \sqrt{1 - \delta^2} \epsilon_{ij} > 0),$$

where  $\epsilon_{ij}$  is simulated independently from a standard normal distribution with  $1 \leq i, j \leq n$ . The marginal model is then given by  $P(Y_{ij} = 1 | \mathbb{X}) = g(X_{ij}^\top \beta)$ , where  $g(\cdot) = \Phi(\cdot)$  is chosen to be the probit link function. Note that  $\delta$  is a scalar between 0 and 1, which controls the effect of network dependence. Furthermore,  $\delta = 0$  corresponds to the null hypothesis of the CI assumption, while  $\delta > 0$  indicates the NI assumption. We try three  $\delta$  values in our simulation study, i.e.,  $\delta = 0, 0.5,$  and  $0.9$ . At last, the regression coefficient is set to be  $\beta = (1.5, 1, 0, 0.5, 0, -1)^\top \in \mathbb{R}^6$ .

As we mentioned in Section 2.4, different independence assumptions (i.e., CI and NI) result in different asymptotic variances, which in turn produce different convergence rates and estimates for standard error. This leads to two types of Z-test for the estimated regression coefficients. Specifically, we report the resulting empirical rejection rates for  $\beta_3$  and

$\beta_5$ , whose true values are 0, under the nominal level 5% in Table 1. As one can see, if the CI assumption holds (i.e.,  $\delta = 0$ ), the empirical rejection rates are very close to 5%; see the 3rd and 5th columns in the top panel. In contrast, the empirical rejection rates under the NI assumption are substantially smaller than the nominal level 5%; see the 4th and 6th columns in the top panel. This is because the formula under the NI assumption seriously overestimates the standard error. On the other hand, when the NI assumption is valid (i.e.,  $\delta = 0.5$  or  $0.9$ ), the associated empirical rejection rates are close to 5%; see columns 4 and 6 in the bottom two panels. While the Z-test under CI assumption has larger rejection rates than 5%; see the 3rd and 5th columns in the bottom two panels. The 7th column of Table 1 reports the mean absolute error (MAE) of the regression coefficient estimates, which is averaged across 1,000 simulation replications. It decreases as the network size gets larger under all the values of  $\delta$  (i.e.,  $\delta = 0, 0.5,$  and  $0.9$ ), which demonstrates the consistency of the estimator. At last, the average value of network density (i.e.,  $ND = \sum_{i \neq j} Y_{ij} / \{n(n-1)\}$ ) is also reported in percentage; see the last column of Table 1.

#### 3.2 Renren dataset

To further demonstrate our method, we present here a real example about Renren website ([www.renren.com](http://www.renren.com)), which can be treated as a Facebook-type social network media in Chinese, whose majority users are university students. Similar to Facebook, Renren allows users to create their profiles, including basic information (e.g., gender, birthday, hometown), university information (e.g., university, admission year, department), and personal information (e.g., interests, favorites). In addition to that, users can write blogs, upload photos and videos. A user is also allowed to add another as a friend as long as his/her invitation is accepted. Consequently, a friend-type relationship can be established.

Our dataset is collected from a small sub-network of Renren, containing a total of 104 nodes. The users are particularly selected so that they are from one common university but two different departments. Those 104 users constitute a small network with its adjacency matrix given by  $\mathbb{Y} \in \mathbb{R}^{104 \times 104}$ . The density of this network is given by  $\sum_{i \neq j} Y_{ij} / \{n(n-1)\} = 2,909 / 10,712 = 27.16\%$ . In addition, the network structure is visualized in the left of Figure 1, where each circle represents a user and a line denotes the existence of friendship between two users. One can observe two clusters in the network clearly, which correspond to two departments. Besides, the histogram of nodal degrees (i.e.,  $d_i = \sum_j Y_{ij}$ ) is given in the right of Figure 1, which exhibits a bimodal shape.

In order to explain such a friendship network  $\mathbb{Y}$ , the following nodal information are collected. Specifically, the categorical information include gender, hometown, admission year and department of each user, which can be seen as their demographic information. The quantitative information include the number of friends of each user, the number

Table 1. Simulation Results with 1,000 replications for different  $\delta$  values. Empirical rejection rates for  $\beta_3 = 0$  and  $\beta_5 = 0$  are reported under two different assumptions (i.e., CI and NI). Mean absolute error for all the regression coefficient estimates is calculated. At last, network density is reported in percentage

$\delta$ value	Network Size	$\beta_3$		$\beta_5$		MAE	Network Density(%)
		CI	NI	CI	NI		
$\delta = 0$	100	0.046	0.002	0.050	0.010	0.022	39.019
	200	0.034	0.005	0.057	0.007	0.011	38.925
	500	0.049	0.003	0.036	0.006	0.004	39.067
$\delta = 0.5$	100	0.554	0.057	0.443	0.054	0.059	39.029
	200	0.655	0.065	0.579	0.046	0.040	38.959
	500	0.816	0.059	0.712	0.050	0.025	39.041
$\delta = 0.9$	100	0.731	0.065	0.666	0.067	0.110	39.048
	200	0.787	0.073	0.765	0.056	0.076	38.993
	500	0.904	0.057	0.850	0.050	0.048	39.026

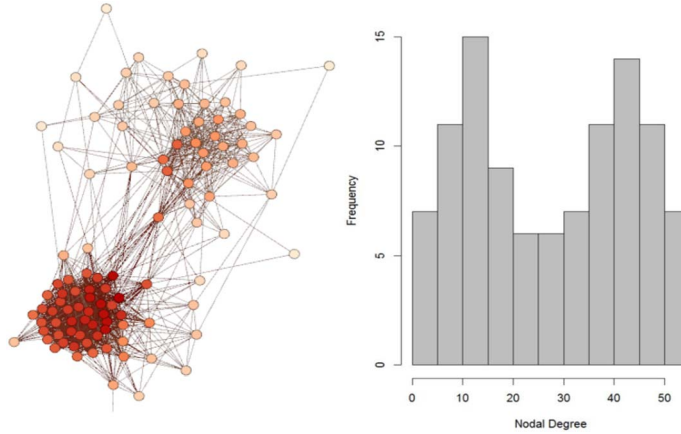


Figure 1. The Renren dataset with 104 users. The left: network structure, circle represents user and line denotes the relationship. The right: histogram of nodal degrees.

of blogs they upload, and the number of visitors to their home pages, which represent their social activeness levels.

To conduct the regression analysis, the following predictive variables are constructed. Following [9], we construct a set of predictors related to the concept of homophily. That is, if two nodes share the same level of a factor (e.g., gender), then the corresponding homophily statistic takes the value of 1, otherwise it equals to 0. The homophily predictors we consider include, whether two nodes share the same level of gender, hometown, admission year and department. Secondly, following [6], we consider the concept of social distance in our regression. To this end, we standardize all the quantitative variables. Then their inter-node absolute difference is included as the predictor.

We next compute the MLE, where the link function is chosen to be probit. The standard errors according to both CI and NI assumptions are also reported. As shown in Table 2, the results are totally different. The CI based results suggest that except for the distance of the number of visitors, all the other estimated regression coefficients are significant un-

der the nominal level 5%. While under NI assumption, only the intercept, the homophily of gender, admission year and department are significant under the nominal level 5%. To test which assumption is more appropriate, we refer to the test procedures of [14]. The resulting transitivity test statistic is 3.11, which provides strong evidence to reject the null hypothesis (i.e., CI assumption) at 5% significance level. As a result, we decide to employ the regression results under the NI assumption for interpretation. More specifically, it suggests that students with the same gender, university admission year, and from the same department are more likely to be friends on Renren.

#### 4. CONCLUDING REMARKS

In this article, we propose a novel assumption describing the network dependence structure, that is the NI assumption. It allows for nontrivial dependence for overlapped dyads (e.g.,  $\{i, j\}$  and  $\{i, k\}$ ). Under the NI assumption, we are able to investigate the asymptotic behavior of the MLE

Table 2. Regression results of the Renren dataset based on CI and NI assumptions

Variable	Coeff.	Z-Test Based On			
		CI		NI	
		S.E.	p-value	S.E.	p-value
Intercept	-2.393	0.056	<0.001	0.357	<0.001
Homophily					
Gender	0.141	0.031	<0.001	0.067	0.036
Hometown	0.191	0.041	<0.001	0.110	0.083
Admission Year	0.834	0.040	<0.001	0.245	<0.001
Department	1.848	0.033	<0.001	0.210	<0.001
Distance					
Friends	0.089	0.021	<0.001	0.129	0.490
Logs	0.051	0.014	<0.001	0.041	0.208
Visitors	-0.026	0.016	0.111	0.066	0.700

$\hat{\beta}$ . To conclude the article, we discuss two topics for further research. First of all, it is noteworthy that whether LMAM is the best dependence assumption for network data is unknown. Further studies on the network dependence structure are of great interest. Secondly, in this article, we mainly discuss dichotomous relationship between nodes (i.e.,  $Y_{ij} = 0$  or 1). However, in real practice, the relationships may exist in terms of varying strengths, which leads to valued relationships. Then how to model such type of response should be further investigated.

## APPENDIX

### Appendix A. Proof of Theorem 1

In order to prove Theorem 1, we introduce a Lemma beforehand.

**Lemma 1.** *Let  $\ell(\beta)$  be the log-likelihood function of (1). Define  $u = (u_1, \dots, u_p)^\top \in \mathbb{R}^p$ . For an arbitrary small  $\epsilon > 0$ , there exists a sufficiently large constant  $C$ , such that*

$$\begin{aligned}
 & n^{-1} \sup_{\|u\|=C} \left\{ \ell(\beta + n^{-1/2}u) - \ell(\beta) \right\} \\
 \text{(A.1)} \quad & = n^{-3/2} \dot{\ell}^\top(\beta)u - \frac{1}{2}u^\top \left\{ -n^{-2} \ddot{\ell}(\beta) \right\} u + o_p(1) \\
 & \leq n^{-3/2} \|\dot{\ell}(\beta)\|C - \frac{1}{2} \lambda_{\min} \left\{ -n^{-2} \ddot{\ell}(\beta) \right\} C^2 + o_p(1),
 \end{aligned}$$

where  $\dot{\ell}(\cdot) \in \mathbb{R}^p$  and  $\ddot{\ell}(\cdot) \in \mathbb{R}^{p \times p}$  are respectively the first- and second-order derivatives of  $\ell(\cdot)$ . Furthermore,  $\lambda_{\min}(H)$  refers to the minimum eigenvalue of  $H$ .

**PROOF OF LEMMA 1.** First of all, the quality holds in (A.1) because of Taylor expansion. Second, as we will prove in Step 2,  $n^{-3/2} \dot{\ell}(\beta) = O_p(1)$ , which implies that  $n^{-3/2} \|\dot{\ell}(\beta)\| = O_p(1)$ . Similarly, we can prove that  $\{-n^{-2} \ddot{\ell}(\beta)\} \rightarrow_p H$ , where  $H$  is defined in (3) and it is a positive definite matrix. Then the second term in (A.1) is quadric in  $C$ . On the other hand, the first term in (A.1) 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 completes the proof of Lemma 1.

The theorem conclusion can be proved in two steps. In the first step, we show that  $\hat{\beta}$  is  $\sqrt{n}$ -consistent. In the second step, we establish its asymptotic normality.

STEP 1. By definition, the log-likelihood function is given by

$$\ell(\beta) = \sum_{i \neq j} \left[ Y_{ij} \log \{g(X_{ij}^\top \beta)\} + (1 - Y_{ij}) \log \{1 - g(X_{ij}^\top \beta)\} \right].$$

It can be easily shown that  $\ell(\beta)$  is a convex function in  $\beta$ . Thus as long as we can prove there exists a  $\sqrt{n}$ -consistent local optimizer, it must be the global optimizer  $\hat{\beta}$ . By [3], we know that this is implied by the following fact, that is for an arbitrary small  $\epsilon > 0$ , there exists a sufficiently large constant  $C$ , such that

$$\text{(A.2)} \quad \lim_{n \rightarrow \infty} P \left[ \sup_{\|u\|=C} \left\{ \ell(\beta + n^{-1/2}u) - \ell(\beta) \right\} < 0 \right] \geq 1 - \epsilon,$$

where  $u = (u_1, \dots, u_p)^\top \in \mathbb{R}^p$ . By Lemma 1, (A.2) can be easily shown, and thus completes the first step of the proof. STEP 2. Because  $\hat{\beta}$  is  $\sqrt{n}$ -consistent, we can apply the standard Taylor expansion augment to establish its asymptotic normality. As a result, we obtain  $\sqrt{n}(\hat{\beta} - \beta) = -\{n^{-2} \ddot{\ell}(\beta)\}^{-1} \{n^{-3/2} \dot{\ell}(\beta)\} + o_p(1)$ . Then the conclusion follows if we can prove  $\{-n^{-2} \ddot{\ell}(\beta)\} \rightarrow_p H$  and  $n^{-3/2} \dot{\ell}(\beta) \rightarrow_d N(0, C_0)$ .

To this end, define  $S = \dot{\ell}(\beta) = \sum_{i \neq j} s_{ij}(\beta)$ . Further define  $m_i = E(s_{ij}|V_i, \mathbb{X})$  and  $h_j = E(s_{ij}|V_j, \mathbb{X})$ . Then, let  $\tilde{S} = \sum_{i \neq j} (m_i + h_j)$ . We next want to show that  $S$  can be well approximated by  $\tilde{S}$ . To this end, we compare  $\text{var}(S|\mathbb{X})$  and  $\text{var}(D|\mathbb{X})$ , where  $D = S - \tilde{S}$ . We first consider  $\text{var}(S) = E\{\text{var}(S|\mathbb{X})\}$ , since  $E(S|\mathbb{X}) = 0$ . Next, we

calculate  $\text{var}(S|\mathbb{X})$  as

$$\begin{aligned} \text{var}(S|\mathbb{X}) &= \text{var}\left(\sum_i \sum_{j \neq i} s_{ij}|\mathbb{X}\right) \\ &= \sum_i \sum_{j \neq i} \text{var}(s_{ij}|\mathbb{X}) + \sum_i \sum_{j \neq i} \text{cov}(s_{ij}, s_{ji}|\mathbb{X}) \\ &\quad + \sum_i \sum_{j \neq i} \sum_{k \neq i, j} \left\{ \text{cov}(s_{ij}, s_{ik}|\mathbb{X}) + \text{cov}(s_{ji}, s_{ki}|\mathbb{X}) \right. \\ &\quad \left. + 2\text{cov}(s_{ji}, s_{ik}|\mathbb{X}) \right\} = O_p(n^3). \end{aligned}$$

We then compare this result against  $\text{var}(D|\mathbb{X})$ , which is further given by the following result.

$$\begin{aligned} \text{(A.3)} \quad &\text{var}\left(\sum_{i \neq j} (s_{ij} - m_i - h_j)|\mathbb{X}\right) \\ &= \sum_{i \neq j} \text{var}(s_{ij} - m_i - h_j|\mathbb{X}) \\ &\quad + \sum_{i \neq j} \text{cov}(s_{ij} - m_i - h_j, s_{ji} - m_j - h_i|\mathbb{X}) \\ &\quad + \sum_i \sum_{j \neq i} \sum_{k \neq i, j} \left\{ \text{cov}(s_{ij} - m_i - h_j, s_{ik} - m_i - h_k|\mathbb{X}) \right. \\ &\quad \left. + \text{cov}(s_{ji} - m_j - h_i, s_{ki} - m_k - h_i|\mathbb{X}) \right. \\ &\quad \left. + 2\text{cov}(s_{ji} - m_j - h_i, s_{ik} - m_i - h_k|\mathbb{X}) \right\}. \end{aligned}$$

Define the last three terms in (A.3) by  $I_1$ ,  $I_2$  and  $I_3$  respectively. We then have  $I_1 = \text{cov}(s_{ij} - m_i - h_j, s_{ik} - m_i - h_k|\mathbb{X}) = E(s_{ij}s_{ik}^\top|\mathbb{X}) - E(s_{ij}m_i^\top|\mathbb{X}) - E(s_{ij}h_k^\top|\mathbb{X}) - E(m_i s_{ik}^\top|\mathbb{X}) + E(m_i m_i^\top|\mathbb{X}) + E(m_i h_k^\top|\mathbb{X}) - E(h_j s_{ik}^\top|\mathbb{X}) + E(h_j m_i^\top|\mathbb{X}) + E(h_j h_k^\top|\mathbb{X})$ . One can verify that  $E(s_{ij}s_{ik}^\top|\mathbb{X}) = E\{E(s_{ij}s_{ik}^\top|\mathbb{X}, V_i)|\mathbb{X}\} = E\{E(s_{ij}|\mathbb{X}, V_i)E(s_{ik}^\top|\mathbb{X}, V_i)|\mathbb{X}\} = E(m_i m_i^\top|\mathbb{X})$ . 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  $I_1 = 0$ . The same technique can be used to derive that  $I_2 = 0$  and  $I_3 = 0$ . So that  $\text{var}(D|\mathbb{X}) = O_p(n^2)$ , which is a smaller order as compared with  $\text{var}(S|\mathbb{X})$ . This suggests that  $S$  and  $\tilde{S}$  share the same asymptotic distribution.

On the other hand,  $\tilde{S}$  can be also written as,  $\tilde{S} = \sum_{i \neq j} (m_i + h_j) = (n-1) \sum_i (m_i + h_i)$ . It can be seen that, given  $\mathbb{X}$ ,  $\tilde{S}$  can be treated as the sum of independent variables. The conditional variance of  $n^{-3/2}\tilde{S}$  can be calculated as the follows,

$$\begin{aligned} \text{var}(n^{-3/2}\tilde{S}|\mathbb{X}) &= n^{-3}(n-1)^2 \sum_i \text{var}(m_i + h_i|\mathbb{X}) \\ &= n^{-3}(n-1)^2 \sum_i E\left\{(m_i + h_i)(m_i + h_i)^\top|\mathbb{X}\right\} \rightarrow_p C_0, \end{aligned}$$

where the last convergence result is due to the theorem assumption.

As a result, by the central limit theorem, we obtain that  $n^{-3/2}S|\mathbb{X}$  is distributed as  $N(0, C_0)$ , where  $C_0$  is a positive definite matrix. This completes the whole proof.

Received 4 November 2016

## REFERENCES

- [1] DE MIGUEL LUKEN, V. and TRANMER, M. (2010), "Personal support networks of immigrants to Spain: a multilevel analysis," *Social Networks*, 32, 253–262.
- [2] ENNETT, S. T. and BAUMAN, K. E. (1993), "Peer group structure and adolescent cigarette smoking: a social network analysis," *Journal of Health and Social Behavior*, 34, 226–236.
- [3] 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.
- [4] FRANK, O. and STRAUSS, D. (1986), "Markov graphs," *Journal of the American Statistical Association*, 81, 832–842.
- [5] 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.
- [6] 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](#)
- [7] HOFF, P. D. (2005), "Bilinear mixed-effects models for dyadic data," *Journal of the American Statistical Association*, 100(469), 286–295. [MR2156838](#)
- [8] 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](#)
- [9] 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.
- [10] KNOKE, D. and YANG, S. (2008), *Social Network Analysis*, London: Sage.
- [11] 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.
- [12] MCCULLAGH, P. and NELDER, J. A. (1989), *Generalized Linear Models*, Chapman and Hall, New York.
- [13] NOWICKI, K. and SNIJDERS, T. A. B. (2001), "Estimation and prediction for stochastic block structures," *Journal of the American Statistical Association*, 96, 1077–1087.
- [14] PAN, R. and WANG, H. (2015), "A note on testing conditional independence for social network analysis," *Science China: Mathematics*, 1179–1190.
- [15] SEWELL, D. K. and CHEN, Y. (2015), "Latent space models for dynamic networks," *Journal of the American Statistical Association*, 110(512), 1646–1658.
- [16] SHAO, J. (1997), *Mathematical Statistics*, New York: Springer.
- [17] SNIJDERS, T. A. B. (2002), "Markov chain monte carlo estimation of exponential random graph models," *Journal of Social Structure*, 3(2).
- [18] STRAUSS, D. and IKEDA, M. (1990), "Pseudolikelihood estimation for social networks," *Journal of the American Statistical Association*, 85, 204–212.
- [19] WANG, Y. J. and WONG, G. Y. (1987), "Stochastic blockmodels for directed graphs," *Journal of the American Statistical Association*, 82, 8–19. [MR0883333](#)
- [20] WASSERMAN, S. and FAUST, K. (1994), *Social Network Analysis: Methods and Applications*, Cambridge University Press.

Rui Pan  
School of Statistics and Mathematics  
Central University of Finance and Economics  
Beijing, 100081  
P. R. China  
E-mail address: [panrui\\_cufe@126.com](mailto:panrui_cufe@126.com)

Rong Guan  
School of Statistics and Mathematics  
Central University of Finance and Economics  
Beijing, 100081  
P. R. China  
E-mail address: [rongguan77@gmail.com](mailto:rongguan77@gmail.com)

Xuening Zhu  
School of Data Science  
Fudan University  
Shanghai, 200433  
P. R. China  
E-mail address: [puddingnnn529@gmail.com](mailto:puddingnnn529@gmail.com)

Hansheng Wang  
Guanghua School of Management  
Peking University  
Beijing, 100871  
P. R. China  
E-mail address: [hansheng@pku.edu.cn](mailto:hansheng@pku.edu.cn)