

## LOGISTIC REGRESSION WITH NETWORK STRUCTURE

Xu Zhang<sup>1</sup>, Rui Pan<sup>2</sup>, Guoyu Guan<sup>1</sup>, Xuening Zhu<sup>3</sup> and Hansheng Wang<sup>4</sup>

<sup>1</sup>*Northeast Normal University*, <sup>2</sup>*Central University of Finance and Economics*,  
<sup>3</sup>*Fudan University* and <sup>4</sup>*Peking University*

*Abstract:* As one of the most popular classification methods, the logistic regression model has been studied extensively. Essentially, the model assumes that an individual's class label is influenced by a set of predictors. However, with the rapid advance of social network services, social network data are becoming increasingly available. As a result, incorporating this additional network structure in order to improve classification accuracy has become an important research problem. To this end, we propose a network-based logistic regression (NLR) model that takes the network structure into consideration. Four interesting scenarios are used to investigate the link formation of the network structure under the NLR model. Furthermore, we determine the impact of the network structure on classification by deriving the asymptotic properties for the prediction rule under different sparsities of network. Lastly, simulation studies are conducted to demonstrate the finite-sample performance of the proposed method, and a real Sina Weibo data set is analyzed for illustrative purposes.

*Key words and phrases:* Classification, logistic regression, network structure.

### 1. Introduction

Classification methods have been used extensively in empirical studies to analyze data with categorical responses. These methods have a wide range of applications in fields such as finance, medicine, and sociology, among others (Michie, Spiegelhalter and Taylor (1994); Hastie, Tibshirani and Friedman (2001)). Of the many existing classification methods, the logistic regression (LR) model has been widely studied (McCullagh and Nelder (1989); Hosmer, Lemeshow and Sturdivant (2013); Lindquist and Mckeague (2012); Polson, Scott and Windle (2013)). The LR model essentially assumes that all individuals are independent of each other, and only the data (i.e., predictors or attributes) obtained from each individual are utilized for classification. For instance, when predicting a person's credit risk (i.e., good or bad), predictors such as status of residence, age, and annual income could be considered (Hand and Henley (1997)).

Today, with the rapid advance of information technology, social network-

related services and websites have emerged (e.g., Twitter, Facebook, Sina Weibo, and WeChat). As a result, social network data are becoming increasingly available. A social network refers to a group of individuals (i.e., nodes) connected to each other through various relationships (i.e., edges), such as friendship, kinship, and common interest (Wasserman and Faust (1994)). Social network analyses have attracted much attention in fields, such as economics, marketing, and finance (Carrington, Scott and Wasserman (2005)).

In the field of statistics, the exponential random graph model (ERGM) is one of the most popular methods used for network modeling (Lusher, Koskinen and Robins (2012)). The  $p_1$  model (Holland and Leinhardt (1981)), which can be regarded as the seminal work in this area, is a simple ERGM. The Markov graphs model of Frank and Strauss (1986) provided a milestone in the development of the ERGM. However, the model was not widely adopted by practitioners until the late 1990s, when a series of studies (Wasserman and Pattison (1996, 1999); Robins, Pattison and Wasserman (1999)) extended the log-linear type model to the  $p^*$  model, that is, an ERGM. Thereafter, the ERGM became a powerful method for network modeling and has been widely explored (Robins et al. (2007); Schweinberger (2011); Almquist and Butts (2014)). In addition to the ERGM, the network block model (Holland, Laskey and Leinhardt (1983); Wang and Wong (1987); Nowicki and Snijders (2001); Kolaczyk and Csardi (2014)) and the latent network model (Hoff, Raftery and Handcock (2002); Sarkar and Moore (2005); Handcock, Raftery and Tantrum (2007); Sewell and Chen (2015)) are two popular network modeling methods that have been widely investigated. Moreover, several recent studies focus on modeling the network evolution (Almquist and Butts (2014); Arabshahi et al. (2015)), where the data are the network structure and each individual's predictors at different time points. The aim is to predict the entry and exit of a node on given days.

Our data contain both the class label and the predictors for each individual, as well as the network structure between these individuals. Unlike the network modeling methods mentioned above, which regard the edges of a network as responses, our main focus is on incorporating the network structure into the traditional classification problem (i.e., the class label is regarded as the response).

In the field of statistics, the autologistic actor-attribute model (Robins, Pattison and Elliott (2001) ALAAM) utilizes network information for classification. Essentially, it is an LR model that assumes that the class label of each individual depends on both the predictors and its neighbors' information on the network (e.g., as actor 2-star, event 2-star, etc.). An approximate algorithm, the pseudo-

likelihood method (Strauss and Ikeda (1990)), is adopted for inference purposes. The ALAAM treats some network statistics as additional predictors. However, there is no scientific guidance on how to decide which network statistics should be employed. Furthermore, this method does not model the network structure; that is, the network is regarded as deterministic.

To this end, we propose a network-based logistic regression (NLR) model that takes the network structure into consideration. The NLR model assumes that whether two nodes are connected is influenced by their class labels and by the similarity in their predictors. Furthermore, the attributes of each node are employed to predict the labels using the classical LR model. The proposed NLR model offers greater flexibility, and provides intuitive explanations under various scenarios about the link formation of the network structure. Moreover, we examine the impact of the network structure on classification by developing approximation algorithms under different sparsities of network and establish their corresponding theoretical properties.

The rest of the paper is organized as follows. Section 2 introduces the NLR model, including the approximate prediction procedures and their corresponding theoretical properties. In Section 3, a number of simulation studies are conducted to demonstrate the finite-sample performance of our proposed method, and a real Sina Weibo data set is analyzed to provide an empirical explanation. Section 4 concludes the paper. All technical proofs are left to the Appendix.

## 2. Network-based Logistic Regression Model

### 2.1. Model and notation

Let  $\{Y_i, X_i\}$  be the observation collected from the  $i$ th ( $1 \leq i \leq n$ ) node, where  $n$  is the total sample size,  $Y_i \in \{0, 1\}$  is the binary response, and  $X_i = (X_{i1}, \dots, X_{ip})^\top \in \mathbb{R}^p$  is the associated  $p$ -dimensional predictor, where  $p$  is assumed to be fixed. Given  $X_i$ s, assume that different  $Y_i$  are independent of each other, and that the regression relationship is given by

$$P(Y_i = 1|X_i) = \frac{\exp(X_i^\top \beta)}{1 + \exp(X_i^\top \beta)}, \quad (2.1)$$

where  $\beta = (\beta_1, \dots, \beta_p)^\top \in \mathbb{R}^p$  is a  $p$ -dimensional regression coefficient vector.

Write  $\mathbb{Y} = (Y_1, \dots, Y_n)^\top \in \mathbb{R}^n$  and  $\mathbb{X} = (X_1, \dots, X_n)^\top \in \mathbb{R}^{n \times p}$ . To describe the network structure, define  $A = (a_{ij}) \in \mathbb{R}^{n \times n}$  as the adjacency matrix, where  $a_{ij} = 1$  if node  $i$  follows node  $j$ , otherwise  $a_{ij} = 0$ . Note that  $A$  can be asymmetric because node  $i$  following node  $j$  does not necessarily mean that the reverse is true.

We follow tradition, and let  $a_{ii} = 0$ , for any  $1 \leq i \leq n$  (Lee (2004)). Given  $\mathbb{Y}$  and  $\mathbb{X}$ , assume that the edges (i.e.,  $a_{ij}$ ) are independent of each other, such that

$$P(a_{ij} = 1 | Y_i = k_1, Y_j = k_2, X_i, X_j) = \frac{\exp(\phi s_{ij} + \omega_{k_1 k_2})}{1 + \exp(\phi s_{ij} + \omega_{k_1 k_2})} \doteq \pi_{ij}^{k_1 k_2}, \quad (2.2)$$

where  $k_1, k_2 \in \{0, 1\}$ ,  $s_{ij}$  represents the similarity of node  $i$  and node  $j$ , according to their predictors,  $\phi$  indicates the strength of the influence of the similarity on the link probability, and  $\omega_{k_1 k_2}$  is a coefficient that reflects the effect of the class labels on the link probability. This is also an LR model. Together, (2.1) and (2.2) constitute the NLR model. Note that the absolute values of the similarities are assumed to be no more than one. In our numerical studies, we use the cosine similarity ( $s_{ij} = \|X_i\|^{-1} \|X_j\|^{-1} X_i^\top X_j$ ) for continuous variables and a simple matching coefficient for binary variables ( $s_{ij} = p^{-1} \sum_{k=1}^p I(X_{ik} = X_{jk})$ ) (Kaufman and Rousseeuw (2009)). Write  $\theta = (\beta^\top, \phi, \omega^\top)^\top \in \mathbb{R}^{p+5}$ , where  $\omega = (\omega_{11}, \omega_{10}, \omega_{01}, \omega_{00})^\top \in \mathbb{R}^4$ . From (2.1) and (2.2), given  $\mathbb{X}$ , the likelihood function can be expressed as

$$\begin{aligned} \mathcal{L}(\theta) &= P(\mathbb{Y}, A | \mathbb{X}) = P(\mathbb{Y} | \mathbb{X}) P(A | \mathbb{Y}, \mathbb{X}) \\ &= \prod_{i=1}^n P(Y_i | X_i) \prod_{i \neq j} P(a_{ij} | Y_i, Y_j, X_i, X_j) \\ &= \prod_{i=1}^n \left\{ \frac{\exp(X_i^\top \beta)}{1 + \exp(X_i^\top \beta)} \right\}^{Y_i} \left\{ \frac{1}{1 + \exp(X_i^\top \beta)} \right\}^{1-Y_i} \\ &\quad \prod_{i \neq j} \prod_{k_1, k_2} \left\{ (\pi_{ij}^{k_1 k_2})^{a_{ij}} (1 - \pi_{ij}^{k_1 k_2})^{1-a_{ij}} \right\}^{I(Y_i=k_1, Y_j=k_2)}, \end{aligned} \quad (2.3)$$

where  $I(\cdot)$  is the indicator function. Denote the maximum likelihood estimator (MLE) of (2.3) as  $\hat{\theta} = (\hat{\beta}^\top, \hat{\phi}, \hat{\omega}^\top)^\top = \arg \max_{\theta} \mathcal{L}(\theta)$ . Furthermore, we can verify that  $\hat{\theta}$  can be derived from two separate LR models. Note that the collinearity of the statistics in (2.2) means we can rewrite the model as  $\text{logit}(\pi_{ij}^{Y_i Y_j}) = \omega_{00} + \phi s_{ij} + (\omega_{11} - \omega_{00}) Y_i Y_j + (\omega_{10} - \omega_{00}) Y_i (1 - Y_j) + (\omega_{01} - \omega_{00}) (1 - Y_i) Y_j$  and derive the corresponding estimator, where  $\text{logit}(x) = \log\{x/(1-x)\}$ .

## 2.2. Prediction using the NLR model

As a standard classification method, we should consider how to predict the unknown class labels after the model is fitted. To this end, the leave-one-out method is adopted to construct a prediction rule. Statistically, this amounts to evaluating the conditional probability  $P(Y_i = 1 | \mathbb{Y}_{(-i)}, \mathbb{X}, A)$ , where  $i$  is an arbitrarily selected node, and  $\mathbb{Y}_{(-i)} = (Y_j : j \neq i)^\top \in \mathbb{R}^{n-1}$ . After algebraic

calculation, we have  $P(Y_i = 1 | \mathbb{Y}_{(-i)}, \mathbb{X}, A) \propto$

$$\frac{\exp(X_i^\top \beta)}{1 + \exp(X_i^\top \beta)} \prod_{j \neq i} \prod_{k=0,1} \left\{ (\pi_{ij}^{1k})^{a_{ij}} (1 - \pi_{ij}^{1k})^{1-a_{ij}} (\pi_{ji}^{k1})^{a_{ji}} (1 - \pi_{ji}^{k1})^{1-a_{ji}} \right\}^{I(Y_j=k)},$$

with some constants ignored. Then, the logit transformation is given by

$$\begin{aligned} \text{logit}\{P(Y_i = 1 | \mathbb{Y}_{(-i)}, \mathbb{X}, A)\} &= X_i^\top \beta + Q_1 + Q_2 + Q_3 + Q_4 \\ &\quad + Q_5 + Q_6 + Q_7 + Q_8, \end{aligned} \tag{2.4}$$

where  $Q_j$  is defined as follows:

$$\begin{aligned} Q_1 &= \sum_{j \neq i}^n a_{ij} Y_j \log \frac{\pi_{ij}^{11}}{\pi_{ij}^{01}}, Q_2 = \sum_{j \neq i}^n a_{ji} Y_j \log \frac{\pi_{ji}^{11}}{\pi_{ji}^{10}}, \\ Q_3 &= \sum_{j \neq i}^n a_{ij} (1 - Y_j) \log \frac{\pi_{ij}^{10}}{\pi_{ij}^{00}}, Q_4 = \sum_{j \neq i}^n a_{ji} (1 - Y_j) \log \frac{\pi_{ji}^{01}}{\pi_{ji}^{00}}, \\ Q_5 &= \sum_{j \neq i}^n (1 - a_{ij}) Y_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}}, Q_6 = \sum_{j \neq i}^n (1 - a_{ji}) Y_j \log \frac{1 - \pi_{ji}^{11}}{1 - \pi_{ji}^{10}}, \\ Q_7 &= \sum_{j \neq i}^n (1 - a_{ij}) (1 - Y_j) \log \frac{1 - \pi_{ij}^{10}}{1 - \pi_{ij}^{00}}, Q_8 = \sum_{j \neq i}^n (1 - a_{ji}) (1 - Y_j) \log \frac{1 - \pi_{ji}^{01}}{1 - \pi_{ji}^{00}}. \end{aligned}$$

In order to better understand (2.4), we consider the following four scenarios.

**SCENARIO 1. (SIMILARITY MODEL)** In the first scenario, we consider a simple network structure that assumes that  $\omega_{11} = \omega_{10} = \omega_{01} = \omega_{00}$ . As a result, the terms  $Q_1$  to  $Q_8$  in (2.4) are all zero. The prediction rule is equivalent to the classical LR model. In particular, if  $\phi = 0$ , model (2.2) degenerates to the ER model of Erdős and Rényi (1959).

**SCENARIO 2. (HOMOPHILY)** In the second scenario, we study the phenomenon of “homophily,” also known as “love of the same.” More specifically, it is assumed that  $\omega_{00} > \omega_{01}$ ,  $\omega_{00} > \omega_{10}$ , and  $\omega_{11} > \omega_{01}$ ,  $\omega_{11} > \omega_{10}$ . Thus, given  $\mathbb{X}$ , nodes with the same label are more likely to be connected. As a result, the log-terms in  $Q_1$ ,  $Q_2$ ,  $Q_7$ , and  $Q_8$  are positive, and in  $Q_3$ ,  $Q_4$ ,  $Q_5$ , and  $Q_6$ , they are negative. For example, for  $Q_1$ , a greater number of  $i$ ’s receivers (i.e., node  $j$  with  $a_{ij} = 1$ ) being labeled as 1 (i.e.,  $Y_j = 1$ ) makes it more likely that node  $i$  will be labeled as 1 (i.e.,  $Y_i = 1$ ). To summarize, under this scenario, the probability that node  $i$  is labeled as 1 (i.e.,  $Y_i = 1$ ) is positively influenced by (a) the number of  $i$ ’s friends (i.e.,  $j$  with  $a_{ij} = 1$  or  $a_{ji} = 1$ ) with a label of 1, and (b) the number of  $i$ ’s nonfriends (i.e.,  $j$  with  $a_{ij} = 0$  or  $a_{ji} = 0$ ) with a label of 0.

**SCENARIO 3. (HETEROPHILY)** In contrast to scenario 2, we focus on the phenomenon of “heterophily,” also known as “love of the different.” Here, nodes with distinct labels are more likely to be connected; that is,  $\omega_{00} < \omega_{01}$ ,  $\omega_{00} < \omega_{10}$ , and  $\omega_{11} < \omega_{01}$ ,  $\omega_{11} < \omega_{10}$ . In such a case, the log-terms in  $Q_1$ ,  $Q_2$ ,  $Q_7$ , and  $Q_8$  are negative, and in  $Q_3$ ,  $Q_4$ ,  $Q_5$ , and  $Q_6$ , they are positive. Consequently, given  $\mathbb{X}$ , the probability that node  $i$  is labeled as 1 is positively influenced by (a) the number of  $i$ 's friends with a label of 0, and (b) the number of  $i$ 's nonfriends with a label of 1.

**SCENARIO 4. (CORE-PERIPHERY)** In the last scenario, class 1 is regarded as the “core” and class 0 is regarded as the “periphery,” without loss of generality. Accordingly, assume  $\omega_{00} < \min\{\omega_{01}, \omega_{10}\} \leq \max\{\omega_{01}, \omega_{10}\} < \omega_{11}$ , which means that nodes are most likely to be connected between the core, and least likely to be connected between the periphery. Under this assumption, the log-terms in  $Q_1$ ,  $Q_2$ ,  $Q_3$ , and  $Q_4$  are positive, and in  $Q_5$ ,  $Q_6$ ,  $Q_7$ , and  $Q_8$ , they are negative. Therefore, given  $\mathbb{X}$ , the more friends node  $i$  has, regardless of their labels, the more likely it is that node  $i$  will be labeled as 1. In addition, the more nonfriends node  $i$  has, the more likely it is that node  $i$  will belong to class 0.

### 2.3. Asymptotic results

So far, we have discussed four scenarios with distinct relationships between the  $\omega$  values. Next, we consider a more general situation of the sparsity of real social networks. Note that a network is called sparse if the network density (i.e.,  $\{n(n-1)\}^{-1} \sum_{i \neq j}^n a_{ij}$ ) tends to zero as the network size  $n$  tends to infinity (Kolaczyk (2009)). To impose sparsity, assume there exist positive constants  $c_{k_1 k_2}$  and  $\gamma$ , where  $\omega_{k_1 k_2} = c_{k_1 k_2} - \gamma \log n$ , for  $k_1, k_2 \in \{0, 1\}$ , from which it can be deduced that there exist some positive constants  $\kappa_1$  and  $\kappa_2$ , such that for any  $i, j, k_1, k_2$ ,

$$P(\kappa_1 n^{-\gamma} \leq \pi_{ij}^{k_1 k_2} \leq \kappa_2 n^{-\gamma}) \rightarrow 1, \quad (2.5)$$

which indicates that as  $n$  gets larger, the link probabilities become smaller. Next, we need to determine a sensible choice for  $\gamma$ . From (2.5), we know immediately that the network is sparse, because  $E(a_{ij}) = O(n^{-\gamma})$ , which shows that the network density tends to zero as  $n$  tends to infinity. Then, the expected out-degree is  $E(d_i) = O(n^{1-\gamma})$ , where  $d_i = \sum_{j \neq i}^n a_{ij}$  is defined as the out-degree of node  $i$ . As a result, we must have  $\gamma \leq 1$ . Otherwise,  $E(d_i) \rightarrow 0$  as  $n \rightarrow \infty$ . This implies that a larger network means fewer nodes are followed. In practice, this is obviously incorrect. Thus, we need to examine the asymptotic properties of the

proposed NLR model with (1)  $\gamma = 1$ , and (2)  $\gamma < 1$ .

CASE I. ( $\gamma = 1$ ) In this case,  $E(d_i) = O(1)$ , which means there is an upper bound for the expected nodal out-degree. In practice, this is a common requirement by many social network platforms. Consider, for example, the Sina Weibo, which is the largest Twitter-type social media in China. Each user in Weibo is limited to following 2,000 followees. This implies that  $d_i \leq 2,000$ , which sets an upper bound for the nodal out-degree. The next theorem follows from (2.4).

**Theorem 1.** *Assume  $\gamma = 1$ , and that  $X_1, \dots, X_n$  are independent and identically distributed (i.i.d.). As  $n \rightarrow \infty$ , we have (1)  $\max_{5 \leq j \leq 8} \{var(Q_j)\} / \min_{1 \leq j \leq 4} \{var(Q_j)\} \rightarrow 0$ , and (2)  $\max_{5 \leq j \leq 8} \{var(Q_j)\} / var(X_i^\top \beta) \rightarrow 0$ .*

The proof of Theorem 1 is given in Appendix A. From Theorem 1, we know that the variability of  $Q_j$  with  $1 \leq j \leq 4$  and  $X_i^\top \beta$  is much larger than that of  $Q_{j'}$ , with  $5 \leq j' \leq 8$ . As a result, we propose the following approximate logit function for prediction purposes:

$$ANLR1 = X_i^\top \beta + Q_1 + Q_2 + Q_3 + Q_4. \tag{2.6}$$

CASE II. ( $\gamma < 1$ ) By assuming  $\gamma < 1$ , there is no upper bound on the nodal out-degree. This may be unrealistic in real life, but we still consider this case for theoretical completeness. Moreover, with the rapid advance of information technology, the upper bound imposed on the nodal out-degree could be unlimited. Theoretically, we can treat it as infinity; thus,  $\gamma < 1$  (i.e.,  $E(d_i) \rightarrow \infty$  as  $n \rightarrow \infty$ ) should be allowed. This leads to the following theorem.

**Theorem 2.** *Assume  $0 < \gamma < 1$ , and that  $X_1, \dots, X_n$  are i.i.d.. As  $n \rightarrow \infty$ , we have (1)  $\max_{5 \leq j \leq 8} \{var(Q_j)\} / \min_{1 \leq j \leq 4} \{var(Q_j)\} \rightarrow 0$ , and (2)  $var(X_i^\top \beta) / \min_{1 \leq j \leq 4} \{var(Q_j)\} \rightarrow 0$ .*

The proof of Theorem 2 is given in Appendix B. From Theorem 2, we can conclude that  $var(Q_j) / var(Q_{j'}) = o(1)$ , for  $5 \leq j \leq 8$  and  $1 \leq j' \leq 4$ . As a result, we propose the following different approximate logit function for predictions when  $0 < \gamma < 1$ :

$$ANLR2 = Q_1 + Q_2 + Q_3 + Q_4. \tag{2.7}$$

To conclude this subsection, recall that  $Q_5$  to  $Q_8$  represent the effects of disconnected nodes. As a result, the above theorems tell us that the contribution of disconnected nodes to a classification is limited under certain sparsity conditions. This implies that we can predict the class label of a particular node only from its neighbors. Moreover, as the network becomes more dense (i.e., Case II

with  $E(d_i) \rightarrow \infty$  as  $n \rightarrow \infty$ ), the effects of the nodal predictors (i.e.,  $X_i$ s) on the classification could be omitted. The simulation studies in the next section further verify that (2.6) and (2.7) perform almost as well as (2.4). As a result, the approximate prediction is both theoretically and practically sound. In practice, we substitute the MLE (i.e.,  $\hat{\theta}$ ) into (2.6) and (2.7) when performing predictions.

### 3. Numerical Studies

#### 3.1. Simulation models

To demonstrate the finite-sample performance of the proposed method, we present three simulation models. The models vary in terms of  $\omega$  and  $\phi$  (i.e., the generating mechanism of the adjacency matrix  $A$ ) and the value of the regression coefficient  $\beta$ . The predictor dimension is fixed to  $p = 5$ . Next,  $X_i = (X_{i1}, \dots, X_{i5})^\top \in \mathbb{R}^5$  is simulated from a multivariate normal distribution with mean  $\mathbf{0}$  and covariance  $\Sigma_X = (\sigma_{j_1 j_2}) \in \mathbb{R}^{5 \times 5}$ , where  $\sigma_{j_1 j_2} = 0.5^{|j_1 - j_2|}$ . The class label  $Y_i$  is then generated according to (2.1).

**MODEL 1. (HOMOPHILY)** In the “homophily” scenario, nodes are more likely to be connected if they have the same label. As a result, it is required that  $\omega_{00} > \omega_{01}$ ,  $\omega_{00} > \omega_{10}$ , and  $\omega_{11} > \omega_{01}$ ,  $\omega_{11} > \omega_{10}$ . We then let  $\omega = (\omega_{11}, \omega_{10}, \omega_{01}, \omega_{00})^\top = (\log(0.8\delta) - \gamma \log n, \log(0.5\delta) - \gamma \log n, \log(0.5\delta) - \gamma \log n, \log(0.8\delta) - \gamma \log n)^\top$  and  $\phi = 2$ , where  $\gamma > 0$  and  $\delta > 0$  control the density of the network, and will be specified subsequently. The network adjacency matrix  $A$  can then be derived according to (2.2). Lastly, we let  $\beta = (-1, 0.8, 1, 0.5, 1.5)^\top$ .

**MODEL 2. (HETEROPHILY)** In the second model, we consider the phenomenon of “heterophily.” That is, nodes with different labels are more likely to be connected. In order to mimic this kind of network structure, we set  $\omega = (\omega_{11}, \omega_{10}, \omega_{01}, \omega_{00})^\top = (\log(0.5\delta) - \gamma \log n, \log(0.8\delta) - \gamma \log n, \log(0.8\delta) - \gamma \log n, \log(0.5\delta) - \gamma \log n)^\top$  and  $\phi = -2$ . Similarly to Model 1, the network adjacency matrix  $A$  can be derived accordingly. The regression coefficient is set to  $\beta = (-0.3, 1, -1, 2, 0.5)^\top$ .

**MODEL 3. (CORE-PERIPHERY)** In the third simulation model, the phenomenon of “core-periphery” is studied. The core is defined as the node labeled as 1 in the network. In practice, the core may denote important people (e.g., movie star, political leader, or business elite). As a result, it is assumed that the probability of important people being connected is much larger. We then let  $\omega = (\omega_{11}, \omega_{10}, \omega_{01}, \omega_{00})^\top = (\log(0.8\delta) - \gamma \log n, \log(0.4\delta) - \gamma \log n, \log(0.4\delta) - \gamma \log n, \log(0.2\delta) - \gamma \log n)^\top$  and  $\phi = 2$ . Lastly, fix  $\beta = (-1, -0.2, 1, 0, 3)^\top$ .

### 3.2. Performance measurements and simulation results

For each simulation model, different network sizes are considered (i.e.,  $n = 200, 500, 1,000$ ). We also try different  $\gamma$  values, namely,  $\gamma = 0.1, 0.5$ , and  $1$ , with corresponding  $\delta$  values of  $0.1, 0.5$ , and  $1$ . The experiment is randomly replicated  $T = 1,000$  times. Let  $A^{(t)} = (a_{ij}^{(t)})$  be the generated network structure in the  $t$ th replication, and let  $\hat{\theta}^{(t)} = (\hat{\beta}^{(t)\top}, \hat{\phi}^{(t)}, \hat{\omega}^{(t)\top})^\top$  be the corresponding estimator. We then consider the following measurements to gauge the performance of our methodology.

First, the root mean squared errors (RMSEs) of the parameter estimators are evaluated using  $\text{RMSE}_\beta = (T^{-1} \sum_{t=1}^T \|\hat{\beta}^{(t)} - \beta\|^2)^{1/2}$ ,  $\text{RMSE}_\phi = (T^{-1} \sum_{t=1}^T \|\hat{\phi}^{(t)} - \phi\|^2)^{1/2}$ , and  $\text{RMSE}_\omega = (T^{-1} \sum_{t=1}^T \|\hat{\omega}^{(t)} - \omega\|^2)^{1/2}$ . Next, in order to evaluate the model’s prediction accuracy, we generate another  $n_0 = 500$  testing samples in each replication, as follows. Consider the first testing sample, indexed by  $i = n + 1$ . In the  $t$ th replication, generate its predictor  $X_{n+1}^{(t)}$  and class label  $Y_{n+1}^{(t)}$ , according to (2.1). Then, generate the network links between the testing sample and the existing  $n$  subjects using (2.2), which we denote as  $\mathbb{E}_{(n+1)}^{(t)} = \{a_{i(n+1)}^{(t)}, a_{(n+1)i}^{(t)}, 1 \leq i \leq n\}$ .

As long as we obtain  $X_{n+1}^{(t)}$  and  $\mathbb{E}_{(n+1)}^{(t)}$ , we can then predict the conditional probability of  $P(Y_{n+1}^{(t)} = 1 | \mathbb{Y}^{(t)}, \mathbb{X}^{(t)}, X_{n+1}^{(t)}, A^{(t)}, \mathbb{E}_{(n+1)}^{(t)})$  using the following four competing methods: (1) the NLR model, that is, prediction formula (2.4); (2) the approximated NLR (ANLR) methods, that is, prediction formula (2.6) when  $\gamma = 1$  or (2.7) when  $\gamma < 1$ ; (3) the classical LR model; and (4) the ALAAM. Note that, for the sake of fairness, the statistics used in the ALAAM are the same as those in the NLR. After removing the effect of collinearity, the final statistics are  $X_i, \sum_{j \neq i}^n a_{ij} Y_j, \sum_{j \neq i}^n a_{ji} Y_j, \sum_{j \neq i}^n a_{ij} (1 - Y_j),$  and  $\sum_{j \neq i}^n a_{ji} (1 - Y_j)$ . We then use the area under the receiver operating characteristic curve (AUC) value to evaluate the out-of-sample prediction accuracy. We also report the mean value of the network density (ND), as  $\text{ND} = (n^2 - n)^{-1} T^{-1} \sum_{t=1}^T \sum_{i \neq j}^n I(a_{ij}^{(t)} = 1)$ . The simulation results are summarized in Table 1.

Because the performance of the model is qualitatively similar, we focus on Model 1 for interpretation. First, we find that for different  $\gamma$ , the RMSE values steadily decrease to zero as  $n$  increases, for both  $\hat{\beta}, \hat{\phi}$ , and  $\hat{\omega}$ . For example, when  $\gamma = 1$ , the RMSE value for  $\hat{\beta}$  drops from 66.653% to 28.017% as  $n$  increases from 200 to 1,000. At the same time, the network density drops from 0.537% to 0.108%, which indicates that the network is becoming increasingly sparse. Second, for

Table 1. Simulation results for three simulation models, with 1,000 replications. The RMSE values (%) are reported for  $\beta$ ,  $\phi$ , and  $\omega$  estimates. The AUC values are given for the network-based logistic regression (NLR), approximate network logistic regression (ANLR), traditional logistic regression (LR), and autologistic actor-attribute (ALAAM) models. The network density (ND) is also reported.

Model	RMSE(%)			AUC				Density	
	$n$	$\beta$	$\phi$	$\omega$	NLR	ANLR	LR	ALAAM	ND(%)
Homophily	$\gamma = 0.1$								
	200	68.679	5.651	10.359	0.975	0.964	0.895	0.965	5.882
	500	39.337	2.367	4.392	0.996	0.977	0.900	0.987	5.404
	1,000	27.661	1.157	2.211	1.000	0.981	0.900	0.994	5.065
	$\gamma = 0.5$								
	200	67.276	6.998	13.164	0.958	0.949	0.896	0.950	3.643
	500	40.879	3.452	6.473	0.977	0.964	0.899	0.969	2.345
	1,000	27.884	2.069	3.862	0.987	0.971	0.900	0.979	1.675
	$\gamma = 1$								
	200	66.653	18.424	35.134	0.908	0.908	0.896	0.904	0.537
	500	39.719	11.580	21.647	0.912	0.912	0.899	0.911	0.216
	1,000	28.017	7.992	15.188	0.913	0.913	0.900	0.912	0.108
Heterophily	$\gamma = 0.1$								
	200	65.385	5.538	10.318	0.971	0.953	0.871	0.961	5.807
	500	39.050	2.323	4.361	0.995	0.973	0.876	0.987	5.334
	1,000	26.770	1.167	2.213	1.000	0.978	0.877	0.995	5.000
	$\gamma = 0.5$								
	200	64.737	6.983	13.076	0.949	0.933	0.871	0.941	3.598
	500	39.162	3.484	6.491	0.972	0.954	0.875	0.964	2.312
	1,000	26.616	2.001	3.803	0.985	0.963	0.876	0.977	1.651
	$\gamma = 1$								
	200	65.757	17.848	34.097	0.887	0.887	0.871	0.883	0.529
	500	38.909	11.181	21.351	0.892	0.891	0.875	0.891	0.213
	1,000	26.954	8.033	15.187	0.894	0.893	0.877	0.893	0.107
Core-Periphery	$\gamma = 0.1$								
	200	89.847	6.802	12.603	0.991	0.977	0.928	0.985	4.027
	500	49.125	2.851	5.360	0.999	0.997	0.931	0.997	3.689
	1,000	34.153	1.390	2.744	1.000	1.000	0.931	1.000	3.452
	$\gamma = 0.5$								
	200	84.265	8.406	16.471	0.981	0.949	0.928	0.975	2.478
	500	49.200	4.060	8.094	0.991	0.979	0.930	0.988	1.586
	1,000	33.184	2.434	4.715	0.996	0.991	0.931	0.993	1.133
	$\gamma = 1$								
	200	87.528	22.414	44.651	0.940	0.940	0.929	0.937	0.362
	500	49.527	13.986	27.013	0.944	0.943	0.931	0.942	0.146
	1,000	33.028	9.810	19.243	0.944	0.944	0.932	0.944	0.073

the out-of-sample prediction accuracy, the NLR model outperforms the classical LR model and the ALAAM in terms of the AUC values. Moreover, when  $\gamma = 1$ , the ANLR1 method is better than the ALAAM because the estimation using our method is more accurate. When  $\gamma < 1$ , the  $X_i^\top \beta$  term is ignored in ANLR2, making it perform slightly worse than the ALAAM.

### 3.3. A Sina Weibo data set

We next present a real-data example from Sina Weibo ([www.weibo.com](http://www.weibo.com)), which is the largest Twitter-type social media in China. The data set is collected from an official Weibo account of an MBA program.

The sample includes  $n = 6,416$  users, and each user is a follower of the official account. The binary response considered is whether the user belongs to the community called “Tsinghua,” which represents one of the most famous universities in China; a user with this label means that he/she has taken the MBA course in “Tsinghua.” Therefore, we have  $Y_i = 1$  for nodes carrying the “Tsinghua” label, otherwise  $Y_i = 0$ . To apply the NLR model, a set of nodal predictors are collected, including (1) the gender of each user (i.e., male = 1 and female = 0), (2) whether the user is authenticated by Sina Weibo (e.g., movie star, business elite, or political leader), (3) the number of Weibo posts, (4) the number of personal labels (created by users to describe their lifestyles and career status), and (5) the tenure, measured in days (i.e., how long the user has been registered). Furthermore, a  $\log(1 + x)$  transformation is applied to the continuous predictors (i.e., number of posts, number of labels, and tenure). The final network density is around 0.33%, indicating a high level of sparsity.

We then apply the NLR model to this data set. The estimation results are given in Table 2. From the first part of Table 2 (i.e., the results for model (2.1)), authentication and the number of personal labels are statistically significant. This shows that authenticated users with more personal labels are more likely to belong to the “Tsinghua” community. Intuitively, the users authenticated by the platform tend to be more active and professional. At the same time, having more personal labels indicates a broader range of interests. We may consider that a person with more professional abilities and broader interests is more likely to be a member of the business elite and to take the MBA course. In addition, from the results in the second part of Table 2 (i.e., the results for model (2.2)), we can see that all coefficients are significant, indicating that both the similarity of the predictors and the class labels affect the link probability. More specifically,  $\hat{\omega} = (-1.32, -3.63, -4.17, -5.13)^\top$  and  $\hat{\phi} = -0.99$ . First, the estimation of  $\omega$  shows

Table 2. The estimation of the NLR model for the Sina Weibo data set.

Regression Coefficient	Estimate	S.E.	<i>p</i> -value
Model-(2.1)			
INTERCEPT	-5.77	1.51	< 0.001
GENDER	0.32	0.19	0.094
AUTHENTICATION	0.65	0.18	< 0.001
NUMBER OF POSTS	-0.08	0.07	0.211
NUMBER OF PERSONAL LABELS	1.48	0.19	< 0.001
TENURE	-0.11	0.42	0.793
Model-(2.2)			
$\omega_{00}$	-5.13	0.02	< 0.001
$\phi$	-0.99	0.03	< 0.001
$\omega_{11} - \omega_{00}$	3.82	0.02	< 0.001
$\omega_{10} - \omega_{00}$	1.50	0.01	< 0.001
$\omega_{01} - \omega_{00}$	0.96	0.01	< 0.001

that the Sina Weibo network is a typical core-periphery network, because  $\hat{\omega}_{00} < \min\{\hat{\omega}_{01}, \hat{\omega}_{10}\} \leq \max\{\hat{\omega}_{01}, \hat{\omega}_{10}\} < \hat{\omega}_{11}$ . Second, the significant difference between  $\hat{\omega}_{01}$  and  $\hat{\omega}_{10}$  shows that the link pattern is asymmetric. Lastly, the coefficient of feature similarity is significant and negative (i.e.,  $\hat{\phi}$ ), indicating that users with different backgrounds are more likely to follow each other, whereas users with similar backgrounds are less likely to follow each other, given their class labels. Moreover, to evaluate the goodness fit of models (2.1) and (2.2), we calculate their deviance values, which are 1,289.1 and 1,806,786.0. The corresponding *p*-values for the goodness of fit are less than 0.001, which means that both models fit the data well.

In order to evaluate the prediction accuracy of our proposed method, we randomly split the samples into two sets. The first set contains 70% of the samples for training, and the other contains the remaining 30% for testing. We then apply the proposed NLR model to the training set and derive the model estimation. Accordingly, the AUC is adopted to evaluate the prediction accuracy of five competing methods (NLR, ANLR1, ANLR2, LR, ALAAM) on the testing set. We randomly repeat this procedure 200 times. The AUC values, averaged over 200 replications, are 0.923 (NLR), 0.923 (ANLR1), 0.891 (ANLR2), 0.749 (LR), and 0.913 (ALAAM). Thus, NLR and ANLR1 perform best, showing that the first approximation (2.6) is reasonable, and that the influence of nonfriends can be ignored. Moreover, the result for the ANLR2 method illustrates that the effects of the predictors should not be dropped. Lastly, we provide the ROC curves for the five competing methods in Figure 1. The figure shows that the

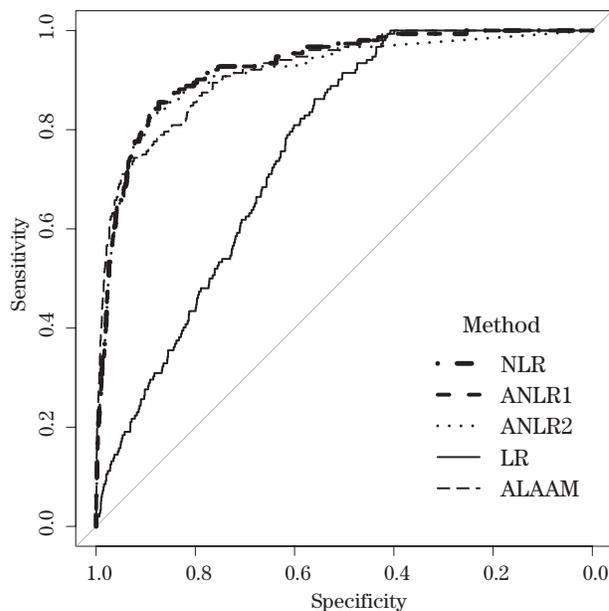


Figure 1. ROC curves of five methods (NLR, ANLR1, ANLR2, LR, and ALAAM) applied to the Sina Weibo data set, where NLR, ANLR1, and ANLR2 show similar results and outperform LR and ALAAM.

NLR model and its corresponding approximation methods have similar results, and that they outperform the LR model and the ALAAM. Additionally, the LR performs worst, because it ignores the network structure.

#### 4. Conclusion

In this paper, we propose a novel probabilistic model for classification, that incorporates a network's structure into the classical LR model. We refer to the proposed model as the network-based logistic regression (NLR) model. The model assumes that the network links are dependent on the class labels and on the nodes' similarity. We also discuss the link formation of the network structure. We further develop two approximate algorithms for model prediction, and derive the asymptotic properties under different sparsities of network.

To conclude, we identify possible topics for further research. First, in practice, the number of classes may be more than two, which leads to a multiclass classification problem. Extending the NLR model to solve this problem could be an important research problem. Second, the dimension of the predictors (i.e.,  $p$ ) is assumed to be fixed. Thus, it would be of great interest to investigate the

problem of variable selection under a high-dimensional framework for the NLR model. Lastly, in addition to the classical LR model, there exist many other methods for classification. Such methods include, but are not limited to support vector machines (Zhang (2006); Wu and Liu (2007) SVM), boosting (Buhlmann and Yu (2003); Zou, Zhu and Hastie (2008)), and so on. Whether the network structure can be incorporated into these methods is also left for further study.

### Acknowledgement

The research of Xu Zhang and Guoyu Guan is supported in part by National Natural Science Foundation of China (NSFC, 11501093, 11690012, 11571068) and the Fundamental Research Funds for the Central Universities (Grant No.2412017 FZ030). The research of Rui Pan is supported in part by National Natural Science Foundation of China (NSFC, 11601539, 71771224) and the Fundamental Research Funds for the Central Universities (QL18010). The research of Hansheng Wang is supported in part by National Natural Science Foundation of China (NSFC, 11831008, 11525101, 71532001) and China's National Key Research Special Program (No. 2016YFC0207704). The research of Xuening Zhu is supported by Shanghai Sailing Program 19YF1402700, National Natural Science Foundation of China (NSFC, U1811461), and Fudan-Xinzailing joint research centre for big data, School of Data Science, Fudan University. The research of all the authors is supported by the Fundamental Research Funds for the Central Universities (Grant No.130028613), and National Natural Science Foundation of China (NSFC, 11631003).

### Appendix

#### Appendix A: Proof of Theorem 1

The proof of Theorem 1 consists of three steps. The variances of  $Q_1$  to  $Q_4$  and  $Q_5$  to  $Q_8$  are calculated respectively in Steps 1 and 2. Then, the variance of  $X_i^\top \beta$  is derived in Step 3. For convenience, denote  $E(Y_i|X_i) = \{1 + \exp(-X_i^\top \beta)\}^{-1} = p_i$  throughout the rest of the proof. It can be easily shown that  $p_i$ s are independent and identically distributed with finite expectation.

STEP 1. In this step,  $\text{var}(Q_1)$  is mainly derived, and the result can be extended to  $Q_2$  to  $Q_4$  easily. By (2.5) and  $\gamma = 1$ , we have  $\pi_{ij}^{11} = O_p(n^{-1})$  and  $\pi_{ij}^{01} = O_p(n^{-1})$ , for any  $i$  and  $j$ . As a result, it can be shown that  $\log(\pi_{ij}^{11}/\pi_{ij}^{01}) = O_p(1)$ . Next, we evaluate the variance of  $Q_1$ , i.e.,  $\text{var}(Q_1) =$

$$\begin{aligned}
 & \sum_{j \neq i}^n \text{var} \left( a_{ij} Y_j \log \frac{\pi_{ij}^{11}}{\pi_{ij}^{01}} \right) + \sum_{j \neq i}^n \sum_{k \neq i, j}^n \text{cov} \left( a_{ij} Y_j \log \frac{\pi_{ij}^{11}}{\pi_{ij}^{01}}, a_{ik} Y_k \log \frac{\pi_{ik}^{11}}{\pi_{ik}^{01}} \right) \\
 &= \sum_{j \neq i}^n \left\{ \text{E} \left( a_{ij} Y_j \log^2 \frac{\pi_{ij}^{11}}{\pi_{ij}^{01}} \right) - \text{E}^2 \left( a_{ij} Y_j \log \frac{\pi_{ij}^{11}}{\pi_{ij}^{01}} \right) \right\} \\
 & \quad + \sum_{j \neq i}^n \sum_{k \neq i, j}^n \left\{ \text{E} \left( a_{ij} a_{ik} Y_j Y_k \log \frac{\pi_{ij}^{11}}{\pi_{ij}^{01}} \log \frac{\pi_{ik}^{11}}{\pi_{ik}^{01}} \right) \right. \\
 & \quad \left. - \text{E} \left( a_{ij} Y_j \log \frac{\pi_{ij}^{11}}{\pi_{ij}^{01}} \right) \text{E} \left( a_{ik} Y_k \log \frac{\pi_{ik}^{11}}{\pi_{ik}^{01}} \right) \right\} \\
 &= \sum_{j \neq i}^n \text{E} \left[ \log^2 \frac{\pi_{ij}^{11}}{\pi_{ij}^{01}} \{ \pi_{ij}^{11} p_i + \pi_{ij}^{01} (1 - p_i) \} p_j \right] \\
 & \quad - \sum_{j \neq i}^n \text{E} \left[ \log \frac{\pi_{ij}^{11}}{\pi_{ij}^{01}} \{ \pi_{ij}^{11} p_i + \pi_{ij}^{01} (1 - p_i) \} p_j \right]^2 \\
 & \quad + \sum_{j \neq i}^n \sum_{k \neq i, j}^n \text{E} \left[ \log \frac{\pi_{ij}^{11}}{\pi_{ij}^{01}} \log \frac{\pi_{ik}^{11}}{\pi_{ik}^{01}} \{ \pi_{ij}^{11} \pi_{ik}^{11} p_i + \pi_{ij}^{01} \pi_{ik}^{01} (1 - p_i) \} p_j p_k \right] \\
 & \quad - \sum_{j \neq i}^n \sum_{k \neq i, j}^n \text{E} \left[ \log \frac{\pi_{ij}^{11}}{\pi_{ij}^{01}} \{ \pi_{ij}^{11} p_i + \pi_{ij}^{01} (1 - p_i) \} p_j \right] \\
 & \quad \text{E} \left[ \log \frac{\pi_{ik}^{11}}{\pi_{ik}^{01}} \{ \pi_{ik}^{11} p_i + \pi_{ik}^{01} (1 - p_i) \} p_k \right].
 \end{aligned}$$

According to our assumptions, the order of the first two terms is  $O(1)$  and the order of the last two terms is  $O(1)$ . As a result,  $\text{var}(Q_1) = O(1)$ . Similarly, it can also be obtained that  $\text{var}(Q_j) = O(1)$  for  $2 \leq j \leq 4$ .

STEP 2. Analogous to the first step, we only need to evaluate the order of  $\text{var}(Q_5)$ . By Taylor's expansion, for any  $i$  and  $j$ , we have

$$\log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} = \log \left( 1 + \frac{\pi_{ij}^{01} - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \right) \approx \frac{\pi_{ij}^{01} - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} = O_p(n^{-1}).$$

Next, we need to compute the variance of  $Q_5$  as  $\text{var}(Q_5) =$

$$\begin{aligned}
 & \text{var} \left( \sum_{j \neq i}^n Y_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \right) + \text{var} \left( \sum_{j \neq i}^n a_{ij} Y_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \right) \\
 & - 2 \text{cov} \left( \sum_{j \neq i}^n Y_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}}, \sum_{j \neq i}^n a_{ij} Y_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \right)
 \end{aligned}$$

$$\begin{aligned}
 &= \sum_{j \neq i}^n \text{var} \left( Y_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \right) + \sum_{j \neq i}^n \sum_{k \neq i, j}^n \text{cov} \left( Y_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}}, Y_k \log \frac{1 - \pi_{ik}^{11}}{1 - \pi_{ik}^{01}} \right) \\
 &\quad + \sum_{j \neq i}^n \text{var} \left( a_{ij} Y_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \right) \\
 &\quad + \sum_{j \neq i}^n \sum_{k \neq i, j}^n \text{cov} \left( a_{ij} Y_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}}, a_{ik} Y_k \log \frac{1 - \pi_{ik}^{11}}{1 - \pi_{ik}^{01}} \right) \\
 &\quad - 2 \sum_{j \neq i}^n \text{cov} \left( Y_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}}, a_{ij} Y_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \right) \\
 &\quad - 2 \sum_{j \neq i}^n \sum_{k \neq i, j}^n \text{cov} \left( Y_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}}, a_{ik} Y_k \log \frac{1 - \pi_{ik}^{11}}{1 - \pi_{ik}^{01}} \right).
 \end{aligned}$$

Similar to step 1, some algebraic calculation shows that the order of the first term is

$$\begin{aligned}
 \sum_{j \neq i}^n \text{var} \left( Y_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \right) &= \sum_{j \neq i}^n \text{E} \left( p_j \log^2 \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \right) \\
 &\quad - \sum_{j \neq i}^n \text{E} \left( p_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \right)^2 = O(n^{-1}),
 \end{aligned}$$

and the second term is

$$\begin{aligned}
 &\sum_{j \neq i}^n \sum_{k \neq i, j}^n \text{cov} \left( Y_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}}, Y_k \log \frac{1 - \pi_{ik}^{11}}{1 - \pi_{ik}^{01}} \right) \\
 &= \sum_{j \neq i}^n \sum_{k \neq i, j}^n \text{E} \left( p_j p_k \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \log \frac{1 - \pi_{ik}^{11}}{1 - \pi_{ik}^{01}} \right) \\
 &\quad - \sum_{j \neq i}^n \sum_{k \neq i, j}^n \text{E} \left( p_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \right) \text{E} \left( p_k \log \frac{1 - \pi_{ik}^{11}}{1 - \pi_{ik}^{01}} \right).
 \end{aligned}$$

Because  $\log(1 - \pi_{ij}^{11})/(1 - \pi_{ij}^{01}) = \text{E}(\log(1 - \pi_{ij}^{11})/(1 - \pi_{ij}^{01})) + \log(1 - \pi_{ij}^{11})/(1 - \pi_{ij}^{01}) - \text{E}(\log(1 - \pi_{ij}^{11})/(1 - \pi_{ij}^{01}))$ , where  $\text{E}(\log(1 - \pi_{ij}^{11})/(1 - \pi_{ij}^{01})) = O(n^{-1})$  and  $\log(1 - \pi_{ij}^{11})/(1 - \pi_{ij}^{01}) - \text{E}(\log(1 - \pi_{ij}^{11})/(1 - \pi_{ij}^{01})) = o_p(n^{-1})$ . At the same time,  $E(p_j p_k) - E(p_j)E(p_k) = 0$  due to the independence of  $p_j$  and  $p_k$ . Hence, after some calculation, we have

$$\text{E} \left( p_j p_k \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \log \frac{1 - \pi_{ik}^{11}}{1 - \pi_{ik}^{01}} \right) - \text{E} \left( p_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \right) \text{E} \left( p_k \log \frac{1 - \pi_{ik}^{11}}{1 - \pi_{ik}^{01}} \right) = o(n^{-2}).$$

Thus, the order of the second term is  $o(1)$ . Similar to the calculation of the first term and the second term, we have the orders of the third term to sixth term are  $O(n^{-2})$ ,  $O(n^{-2})$ ,  $O(n^{-2})$  and  $O(n^{-1})$ . Hence the order of all terms is  $o(1)$ . As a result, we have  $\text{var}(Q_5) = o(1)$ . It can also be derived that  $\text{var}(Q_j) = o(1)$  for  $6 \leq j \leq 8$  in a similar way.

STEP 3. Since  $X_{is}$  are independent and identically distributed, and the dimension of predictor (i.e.,  $p$ ) is fixed, we have that  $\text{var}(X_i^\top \beta) = \beta^\top \text{var}(X_i) \beta = O(1)$ .

Combining the results in the above three steps, we can obtain the following two results, (1)  $\max_{5 \leq j \leq 8} \{\text{var}(Q_j)\} / \min_{1 \leq j \leq 4} \{\text{var}(Q_j)\} \rightarrow 0$  and (2)  $\max_{5 \leq j \leq 8} \{\text{var}(Q_j)\} / \text{var}(X_i^\top \beta) \rightarrow 0$ . This completes the proof of Theorem 1.

**Appendix B: Proof of Theorem 2**

The proof of Theorem 2 consists of three steps. In step 1, we show the variances of  $Q_j$  ( $1 \leq j \leq 4$ ) terms. In Step 2, we compute the variances of  $Q_j$  ( $5 \leq j \leq 8$ ) terms. In the last step, we derive the variance of  $X_i^\top \beta$ .

STEP 1. In the first step, we evaluate the order of  $\text{var}(Q_1)$ , and the result can be generalized to  $Q_2$  to  $Q_4$  easily. By assumption (2.5) and  $\gamma < 1$ , we have  $\pi_{ij}^{11} = O_p(n^{-\gamma})$ ,  $\pi_{ij}^{01} = O_p(n^{-\gamma})$  and  $\log(\pi_{ij}^{11} / \pi_{ij}^{01}) = O_p(1)$  for any  $i$  and  $j$ . Next, we evaluate the variance of  $Q_1$  as follows, i.e.,  $\text{var}(Q_1) =$

$$\begin{aligned} & \sum_{j \neq i}^n \text{var} \left( a_{ij} Y_j \log \frac{\pi_{ij}^{11}}{\pi_{ij}^{01}} \right) + \sum_{j \neq i}^n \sum_{k \neq i, j}^n \text{cov} \left( a_{ij} Y_j \log \frac{\pi_{ij}^{11}}{\pi_{ij}^{01}}, a_{ik} Y_k \log \frac{\pi_{ik}^{11}}{\pi_{ik}^{01}} \right) \\ &= \sum_{j \neq i}^n \left\{ \text{E} \left( a_{ij} Y_j \log^2 \frac{\pi_{ij}^{11}}{\pi_{ij}^{01}} \right) - \text{E}^2 \left( a_{ij} Y_j \log \frac{\pi_{ij}^{11}}{\pi_{ij}^{01}} \right) \right\} \\ & \quad + \sum_{j \neq i}^n \sum_{k \neq i, j}^n \left\{ \text{E} \left( a_{ij} a_{ik} Y_j Y_k \log \frac{\pi_{ij}^{11}}{\pi_{ij}^{01}} \log \frac{\pi_{ik}^{11}}{\pi_{ik}^{01}} \right) \right. \\ & \quad \left. - \text{E} \left( a_{ij} Y_j \log \frac{\pi_{ij}^{11}}{\pi_{ij}^{01}} \right) \text{E} \left( a_{ik} Y_k \log \frac{\pi_{ik}^{11}}{\pi_{ik}^{01}} \right) \right\} \\ &= \sum_{j \neq i}^n \text{E} \left[ \log^2 \frac{\pi_{ij}^{11}}{\pi_{ij}^{01}} \{ \pi_{ij}^{11} p_i + \pi_{ij}^{01} (1 - p_i) \} p_j \right] \\ & \quad - \sum_{j \neq i}^n \text{E} \left[ \log \frac{\pi_{ij}^{11}}{\pi_{ij}^{01}} \{ \pi_{ij}^{11} p_i + \pi_{ij}^{01} (1 - p_i) \} p_j \right]^2 \end{aligned}$$

$$\begin{aligned}
 & + \sum_{j \neq i}^n \sum_{k \neq i, j}^n \mathbb{E} \left[ \log \frac{\pi_{ij}^{11}}{\pi_{ij}^{01}} \log \frac{\pi_{ik}^{11}}{\pi_{ik}^{01}} \{ \pi_{ij}^{11} \pi_{ik}^{11} p_i + \pi_{ij}^{01} \pi_{ik}^{01} (1 - p_i) \} p_j p_k \right] \\
 & - \sum_{j \neq i}^n \sum_{k \neq i, j}^n \mathbb{E} \left[ \log \frac{\pi_{ij}^{11}}{\pi_{ij}^{01}} \{ \pi_{ij}^{11} p_i + \pi_{ij}^{01} (1 - p_i) \} p_j \right] \\
 & \mathbb{E} \left[ \log \frac{\pi_{ik}^{11}}{\pi_{ik}^{01}} \{ \pi_{ik}^{11} p_i + \pi_{ik}^{01} (1 - p_i) \} p_k \right].
 \end{aligned}$$

According to our assumptions, the order of the first two terms is  $O(n^{1-\gamma})$  and the order of the last two terms is  $O(n^{2-2\gamma})$ . As a result,  $\text{var}(Q_1) = O(n^{2-2\gamma})$ . Similarly, it can also be obtained that  $\text{var}(Q_j) = O(n^{2-2\gamma})$  for  $2 \leq j \leq 4$ .

STEP 2. Analogous to the first step, we only need to evaluate the order of  $\text{var}(Q_5)$ . By Taylor’s expansion, for any  $i$  and  $j$ , we have

$$\log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} = \log \left( 1 + \frac{\pi_{ij}^{01} - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \right) \approx \frac{\pi_{ij}^{01} - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} = O_p(n^{-\gamma}).$$

Next, we need to compute the variance of  $Q_5$  as  $\text{var}(Q_5) =$

$$\begin{aligned}
 & \text{var} \left( \sum_{j \neq i}^n Y_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \right) + \text{var} \left( \sum_{j \neq i}^n a_{ij} Y_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \right) \\
 & - 2 \text{cov} \left( \sum_{j \neq i}^n Y_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}}, \sum_{j \neq i}^n a_{ij} Y_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \right) \\
 & = \sum_{j \neq i}^n \text{var} \left( Y_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \right) + \sum_{j \neq i}^n \sum_{k \neq i, j}^n \text{cov} \left( Y_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}}, Y_k \log \frac{1 - \pi_{ik}^{11}}{1 - \pi_{ik}^{01}} \right) \\
 & + \sum_{j \neq i}^n \text{var} \left( a_{ij} Y_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \right) + \sum_{j \neq i}^n \sum_{k \neq i, j}^n \text{cov} \left( a_{ij} Y_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}}, a_{ik} Y_k \log \frac{1 - \pi_{ik}^{11}}{1 - \pi_{ik}^{01}} \right) \\
 & - 2 \sum_{j \neq i}^n \text{cov} \left( Y_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}}, a_{ij} Y_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \right) \\
 & - 2 \sum_{j \neq i}^n \sum_{k \neq i, j}^n \text{cov} \left( Y_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}}, a_{ik} Y_k \log \frac{1 - \pi_{ik}^{11}}{1 - \pi_{ik}^{01}} \right).
 \end{aligned}$$

Similar to step 1, some algebraic calculation shows that the order of the first term is

$$\sum_{j \neq i}^n \text{var} \left( Y_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \right) = \sum_{j \neq i}^n \mathbb{E} \left( p_j \log^2 \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \right)$$

$$- \sum_{j \neq i}^n \mathbb{E} \left( p_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \right)^2 = O(n^{1-2\gamma}),$$

and the second term is

$$\begin{aligned} & \sum_{j \neq i}^n \sum_{k \neq i, j}^n \text{cov} \left( Y_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}}, Y_k \log \frac{1 - \pi_{ik}^{11}}{1 - \pi_{ik}^{01}} \right) \\ &= \sum_{j \neq i}^n \sum_{k \neq i, j}^n \mathbb{E} \left( p_j p_k \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \log \frac{1 - \pi_{ik}^{11}}{1 - \pi_{ik}^{01}} \right) \\ & \quad - \sum_{j \neq i}^n \sum_{k \neq i, j}^n \mathbb{E} \left( p_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \right) \mathbb{E} \left( p_k \log \frac{1 - \pi_{ik}^{11}}{1 - \pi_{ik}^{01}} \right). \end{aligned}$$

Because  $\log(1 - \pi_{ij}^{11})/(1 - \pi_{ij}^{01}) = \mathbb{E}(\log(1 - \pi_{ij}^{11})/(1 - \pi_{ij}^{01})) + \log(1 - \pi_{ij}^{11})/(1 - \pi_{ij}^{01}) - \mathbb{E}(\log(1 - \pi_{ij}^{11})/(1 - \pi_{ij}^{01}))$ , where  $\mathbb{E}(\log(1 - \pi_{ij}^{11})/(1 - \pi_{ij}^{01})) = O(n^{-\gamma})$  and  $\log(1 - \pi_{ij}^{11})/(1 - \pi_{ij}^{01}) - \mathbb{E}(\log(1 - \pi_{ij}^{11})/(1 - \pi_{ij}^{01})) = o_p(n^{-\gamma})$ . At the same time,  $E(p_j p_k) - E(p_j)E(p_k) = 0$  due to the independence of  $p_j$  and  $p_k$ . Hence, after some calculation, we have

$$\begin{aligned} & \mathbb{E} \left( p_j p_k \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \log \frac{1 - \pi_{ik}^{11}}{1 - \pi_{ik}^{01}} \right) - \mathbb{E} \left( p_j \log \frac{1 - \pi_{ij}^{11}}{1 - \pi_{ij}^{01}} \right) \mathbb{E} \left( p_k \log \frac{1 - \pi_{ik}^{11}}{1 - \pi_{ik}^{01}} \right) \\ &= o(n^{-2\gamma}). \end{aligned}$$

Thus, the order of the second term is  $o(n^{2-2\gamma})$ . Similar to the calculation of the first term and the second term, we have the orders of the third term to sixth term are  $O(n^{1-3\gamma})$ ,  $O(n^{2-4\gamma})$ ,  $O(n^{1-3\gamma})$  and  $O(n^{2-3\gamma})$ . Hence the order of all terms is  $o(n^{2-2\gamma})$ . As a result, we have  $\text{var}(Q_5) = o(n^{2-2\gamma})$ . It can also be derived that  $\text{var}(Q_j) = o(n^{2-2\gamma})$  for  $6 \leq j \leq 8$  in a similar way.

STEP 3. As proved in Appendix A, we know that  $\text{var}(X_i^\top \beta) = O(1)$ .

Combining the results derived in the above three steps, we obtain the conclusions (1)  $\max_{5 \leq j \leq 8} \{\text{var}(Q_j)\} / \min_{1 \leq j \leq 4} \{\text{var}(Q_j)\} \rightarrow 0$  and (2)  $\text{var}(X_i^\top \beta) / \min_{1 \leq j \leq 4} \{\text{var}(Q_j)\} \rightarrow 0$ , as  $n \rightarrow \infty$ . This completes the proof of Theorem 2.

### References

Arabshahi, F., Huang, F., Anandkumar, A., Butts, C. T. and Fitzhugh, S. M. (2015). Are you going to the party: depends, who else is coming?: learning hidden group dynamics via conditional latent tree models. *International Conference on Data Mining*, 697–702.

Almquist, Z. W. and Butts, C. T. (2014). Logistic network regression for scalable analysis of networks with joint edge/vertex dynamics. *Sociological Methodology* **44**, 273–321.

Buhlmann, P. and Yu, B. (2003). Boosting with the L2 loss: regression and classification.

- Journal of the American Statistical Association* **98**, 324–340.
- Carrington, P. J., Scott, J. and Wasserman, S. (2005). *Models and Methods in Social Network Analysis*. Cambridge University Press.
- Erdős, P. and Rényi, A. (1959). Random graphs. *Publicationes Mathematicae* **6**, 290–297.
- Frank, O. and Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association* **81**, 832–842.
- Hand, D. J. and Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **160**, 523–541.
- 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 (Statistics in Society)* **170**, 301–354.
- Hastie, T., Tibshirani, R. and Friedman, J. (2001). *The Elements of Statistical Learning*. Springer.
- 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.
- 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.
- Holland, P. W., Laskey, K. B. and Leinhardt, S. (1983). Stochastic blockmodels: first steps. *Social Networks* **5**, 109–137.
- Hosmer, Jr, D. W., Lemeshow, S. and Sturdivant, R. X. (2013). *Applied Logistic Regression*. John Wiley and Sons.
- Kaufman, L. and Rousseeuw, P. J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley and Sons.
- Kolaczyk, E. D. (2009). *Statistical Analysis of Network Data*, Springer.
- Kolaczyk, E. D. and Csárdi, G. (2014). *Statistical Analysis of Network Data with R*. Springer.
- Lee, L. F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* **72**, 1899–1925.
- Lindquist, M. and Mckeague, I. (2012). Logistic regression with brownian-like predictors. *Journal of the American Statistical Association* **104**, 1575–1585.
- Lusher, D., Koskinen, J. and Robins, G. (2012). *Exponential Random Graph Models for Social Networks: Theory, Methods and Applications*. Cambridge University Press.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Model* (2 edition). Chapman and Hall.
- Michie, D., Spiegelhalter, D. J. and Taylor, C. C. (1994). Machine learning, neural and statistical classification. *Technometrics* **37**, 459.
- Nowicki, K. and Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association* **96**, 1077–1087.
- Polson, N., Scott, J. and Windle, J. (2013). Bayesian inference for logistic models using pólygamma latent variables. *Journal of the American Statistical Association* **108**, 1339–1349.
- Robins, G., Pattison, P. and Elliott, P. (2001). Network models for social influence processes. *Psychometrika* **66**, 161–189.
- Robins, G., Pattison, P. and Wasserman, S. (1999). Logit models and logistic regressions for social networks: III. valued relations. *Psychometrika* **64**, 371–394.

- Robins, G., Snijders, T., Wang, P., Handcock, M. and Pattison, P. (2007). Recent developments in exponential random graph ( $p^*$ ) models for social networks. *Social Networks* **29**, 192–215.
- Sarkar, P. and Moore, A. W. (2005). Dynamic social network analysis using latent space models. *ACM SIGKDD Explorations Newsletter* **7**, 31–40.
- Schweinberger, M. (2011). Instability, sensitivity and degeneracy of discrete exponential families. *Journal of the American Statistical Association* **106**, 1361–1370.
- Sewell, D. K. and Chen, Y. (2015). Latent space models for dynamic networks. *Journal of the American Statistical Association* **110**, 1646–1657.
- Strauss, D. and Ikeda, M. (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association* **85**, 204–212.
- Wang, Y. J. and Wong, G. Y. (1987). Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association* **82**, 8–19.
- Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press.
- Wasserman, S. and Pattison, P. (1996). Logit models and logistic regressions for social networks: I. an introduction to Markov graphs and  $p$ -star. *Psychometrika* **61**, 401–425.
- Wasserman, S. and Pattison, P. (1999). Logit models and logistic regressions for social networks: II. multivariate relations. *British Journal of Mathematical and Statistical Psychology* **52**, 169–193.
- Wu, Y. and Liu, Y. (2007). Robust truncated-hinge-loss support vector machines. *Journal of the American Statistical Association* **102**, 974–983.
- Zhang, H. H. (2006). Variable selection for support vector machines via smoothing spline ANOVA. *Statistica Sinica* **16**, 659–674.
- Zou, H., Zhu, J. and Hastie, T. (2008). New multiclass boosting algorithms based on multiclass Fisher-consistent losses. *The Annals of Applied Statistics* **2**, 1290–1306.

School of Mathematics and Statistics, Northeast Normal University, 5268 Renmin Street, Changchun, 130024, Jilin, P. R. China.

E-mail: zhangx264@nenu.edu.cn

School of Statistics and Mathematics, Central University of Finance and Economics, 39 Xueyuan South Road, Beijing, 100081, P. R. China.

E-mail: panrui\_cufe@126.com

School of Economics, Northeast Normal University, 2555 Jingyue Street, Changchun, 130117, Jilin, P. R. China.

E-mail: guangy599@nenu.edu.cn

School of Data Science, Fudan University, Zibin Courtyard, 220 Handan Road, Shanghai, 200433, P. R. China.

E-mail: xueningzhu@fudan.edu.cn

Department of Business Statistics and Econometrics, Guanghua School of Management, Peking University, Beijing, 100871, P. R. China.

E-mail: hansheng@pku.edu.cn

(Received July 2016; accepted May 2018)