Statistics and Probability Letters 138 (2018) 27–30

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

A note on a network model with degree heterogeneity and homophily

Liju Su [1], Xiaodi Qian *, Ting Yan

Central China Normal University, China

ARTICLE INFO

Article history:
Received 4 June 2017
Received in revised form 19 January 2018
Accepted 10 February 2018
Available online 24 February 2018

MSC:
62E20
62F12

Keywords:
Central limit theorem
Maximum likelihood estimation
Network data

ABSTRACT

In this note, we establish a central limit theorem for the maximum likelihood estimator of the degree parameter in a network model with degree heterogeneity and homophily when the number of nodes goes to infinity.

1. Introduction

Two types of features commonly exist in social and econometric network data: degree heterogeneity and homophily. The first characterizes the variation in the node degrees, where some nodes have many links and some others have few links (e.g., Barabási and Albert, 1999). The second characterizes the phenomenon that the nodes tend to form connections with those like themselves (e.g., McPherson et al., 2001). To model how the degree heterogeneity and homophily affect the link formation, Graham (2017) introduces a network model with "fixed effects", which generalizes the β-model (Chatterjee et al., 2011) to allow homophily.

Graham's model can be described as follows. Consider an undirected graph $\mathcal{G}_n$ on $n \geq 2$ agents labeled by $1, \ldots, n$. Let $a_{ij} \in \{0, 1\}$ be an indicator of whether there is a link between agents i and j. That is, if there is a link between i and j, then $a_{ij} = 1$; otherwise, $a_{ij} = 0$. Denote $A = (a_{ij})_{n\times n}$ as the symmetric adjacency matrix of $\mathcal{G}_n$. We assume that there are no self-loops, i.e., $a_{ii} = 0$. The model postulates that a_{ij}'s are independently distributed by Bernoulli distributions, in which a link connects agents i and j with probability

$$P(a_{ij} = 1) = \frac{\exp(Z_{ij}^{\top}\boldsymbol{\gamma} + \beta_i + \beta_j)}{1 + \exp(Z_{ij}^{\top}\boldsymbol{\gamma} + \beta_i + \beta_j)}. \quad (1)$$

In the above equation, Z_{ij} is the dyad-level vector between agents i and j constructed by vectors of agent-level attributes X_i, i.e., $Z_{ij} = g(X_i, X_j)$ for some symmetric function $g(\cdot, \cdot)$. For example, we can use $g(X_i, X_j) = \|X_i - X_j\|_1$ to measure the

* Corresponding author.
E-mail addresses: lijusu@mails.ccnu.edu.cn (L. Su), qianxiaodi@mails.ccnu.edu.cn (X. Qian), tingyanty@mail.ccnu.edu.cn (T. Yan).
[1] Department of Statistics, Central China Normal University, Wuhan, 430079, China.

https://doi.org/10.1016/j.spl.2018.02.046
0167-7152/

similarity. The parameter $\boldsymbol{\gamma}$ quantifies the magnitude of the homophily, while the parameter β_i quantifies the strength of agent i to participate in network connection.

The asymptotic property of model (1) is nonstandard since the number of parameters increases as the number of nodes grows and only one network is observed. Graham (2017) proposes two types of estimators for the homophily parameter – the tetrad logit estimator and the maximum likelihood estimator (MLE) – and establishes their asymptotic distributions. However, the asymptotic distribution of the MLE for the degree parameter has not been explored in his paper. In this note, we further establish the central limit theorem for the MLE of the degree parameter.

For the remainder of this paper, we proceed as follows. We present the central limit theorem in Section 2. We evaluate the asymptotic result by simulations in Section 3. Some further discussion is provided in Section 4. All proofs are relegated to the online supplementary material.

2. Main results

Let $\mathbf{d} = (d_1, \ldots, d_n)^\top$ be the degree sequence of $\mathcal{G}_n$, where $d_i = \sum_j a_{ij}$. Given the observed network $\mathcal{G}_n$, the log-likelihood for model (1) is

$$\ell(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \sum_{i<j} Z_{ij}^\top \boldsymbol{\gamma} a_{ij} + \sum_i \beta_i d_i - \sum_{i<j} \log(1 + e^{Z_{ij}^\top \boldsymbol{\gamma} + \beta_i + \beta_j}). \tag{2}$$

Following Graham (2017), we assume that $\boldsymbol{\gamma} \in \Gamma$ and $\boldsymbol{\beta} \in \Xi$, where $\Gamma \subset \mathbb{R}^p$ with a fixed dimension p and $\Xi \subset \mathbb{R}^n$ are compact subsets. For the convenience of notation, we replace the condition $\boldsymbol{\beta} \in \Xi$ by $\|\boldsymbol{\beta}\|_\infty \le L$ where L is a fixed constant. Assume that $\mathbb{E}[\ell(\boldsymbol{\gamma}, \boldsymbol{\beta}) | Z, \boldsymbol{\gamma}^*, \boldsymbol{\beta}^*]$ is uniquely maximized at $\boldsymbol{\gamma} = \boldsymbol{\gamma}^*$ and $\boldsymbol{\beta} = \boldsymbol{\beta}^*$, where $\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*$ denote the true parameters under which the data are generated. We further assume that all Z_{ij}'s lie in one compact subset. Since $\boldsymbol{\gamma}$ and Z_{ij}'s are all in compact subsets, they are bounded such that we have

$$|Z_{ij}^\top \boldsymbol{\gamma}| \le \kappa, \quad 1 \le i \ne j \le n, \tag{3}$$

where κ is a constant. Following Graham (2017), we study the restricted joint maximum likelihood estimation for $\boldsymbol{\gamma} \in \Gamma$ and $\|\boldsymbol{\beta}\|_\infty \le L$ and define the MLE as

$$(\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\beta}}) = \mathop{\operatorname{argmax}}_{\boldsymbol{\gamma} \in \Gamma, \|\boldsymbol{\beta}\|_\infty \le L} \ell(\boldsymbol{\gamma}, \boldsymbol{\beta}),$$

where $\widehat{\boldsymbol{\gamma}} = (\hat{\gamma}_1, \ldots, \hat{\gamma}_p)$ and $\widehat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_n)$ are the MLEs of $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$, respectively. Then the likelihood equations are

$$\begin{aligned} d_i &= \sum_{j \ne i} \frac{e^{Z_{ij}^\top \boldsymbol{\gamma} + \beta_i + \beta_j}}{1 + e^{Z_{ij}^\top \boldsymbol{\gamma} + \beta_i + \beta_j}}, \quad i = 1, \ldots, n, \\ \sum_{i<j} Z_{ij} a_{ij} &= \sum_{i<j} \frac{Z_{ij} e^{Z_{ij}^\top \boldsymbol{\gamma} + \beta_i + \beta_j}}{1 + e^{Z_{ij}^\top \boldsymbol{\gamma} + \beta_i + \beta_j}}. \end{aligned} \tag{4}$$

We use superscript "0" to denote the interior of a set. If $\boldsymbol{\gamma} \in \Gamma^0$ and $\boldsymbol{\beta} \in \Xi^0$, then the solution to the above system of equations is precisely the MLE.

Denote the covariance matrix of $\boldsymbol{d} = (d_1, \ldots, d_n)$ by $V_n = (v_{ij})_{n \times n}$, where

$$v_{ij} = \frac{e^{Z_{ij}^\top \boldsymbol{\gamma} + \beta_i + \beta_j}}{(1 + e^{Z_{ij}^\top \boldsymbol{\gamma} + \beta_i + \beta_j})^2}, \quad v_{ii} = \sum_{j \ne i} v_{ij} \ \ (i, j = 1, \ldots, n; i \ne j). \tag{5}$$

It is also the Fisher information matrix for $\boldsymbol{\beta}$. Let $\hat{v}_{ii}$ be the estimator of v_{ii} by replacing $\boldsymbol{\gamma}$ with $\widehat{\boldsymbol{\gamma}}$ and keeping $\boldsymbol{\beta}$ unchanged. The central limit theorem is stated below.

Theorem 1. *Assume that $\boldsymbol{\gamma}^* \in \Gamma$ and $A \sim \mathbb{P}_{\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*}$, where $\mathbb{P}_{\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*}$ denotes the probability distribution (1) on A under the parameters $\boldsymbol{\gamma}^*$ and $\boldsymbol{\beta}^*$. If $\|\boldsymbol{\beta}^*\|_\infty \le L$, then for any fixed $r \ge 1$, as $n \to \infty$, the vector $\{\hat{v}_{11}^{1/2}(\hat{\beta}_1 - \beta_1^*), \ldots, \hat{v}_{rr}^{1/2}(\hat{\beta}_r - \beta_r^*)\}$ converges in distribution to the r-dimensional standardized multivariate normal distribution.*

Remark 1. Let $\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma}^*)$ be the profile MLE of $\boldsymbol{\beta}$, i.e.,

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma}^*) = \arg\max_{\boldsymbol{\beta} \in \Xi} \ell(\boldsymbol{\gamma}^*, \boldsymbol{\beta}).$$

In Lemma 6 of the supplementary material in Graham (2017), Graham obtains the central limit theorem for the vector $(\hat{\beta}_1(\boldsymbol{\gamma}^*), \ldots, \hat{\beta}_r(\boldsymbol{\gamma}^*))^\top$. Here, we obtain the asymptotic distribution of the MLE $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\gamma}})$, which is different.

Remark 2. By Theorem 1, for any fixed i, the convergence rate of $\hat{\beta}_i$ is $1/v_{i,i}^{1/2}$ in the magnitude of $n^{-1/2}$ when $\boldsymbol{\gamma}^*$ and $\boldsymbol{\beta}^*$ are bounded.