Statistics and Probability Letters 119 (2016) 163–169

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Asymptotics of score test in the generalized β-model for networks

Ting Yan [a], Yunpeng Zhao [b,*]

[a] Department of Statistics and Hubei Key Laboratory of Mathematical Sciences, Central China Normal University, Wuhan, 430079, China

[b] Department of Statistics, George Mason University, VA 22030-4444, USA

ARTICLE INFO

Article history:
Received 17 March 2016
Received in revised form 23 July 2016
Accepted 24 July 2016
Available online 9 August 2016

MSC:
62E20
62F05

Keywords:
Central limit theorem
Network models
Score test

ABSTRACT

We show that the scaled score test statistic under a simple null in the generalized β-model for undirected networks asymptotically follows standard normal distribution when the number of network vertices goes to infinity.

1. Introduction

Pairwise relationships among a set of objects (e.g., friendships between individuals, citations between papers) can be conveniently represented in a network. Many statistical models have been developed to describe the generation mechanism of networks (see Bhattacharyya and Bickel, 2015; Goldenberg et al., 2010; Matias and Robin, 2014 for recent review), with applications in biology, computer science, social sciences and many other areas. Meanwhile, the non-standard structure of network data poses new challenge for statistical inference since typically, only one realized network is available (Fienberg, 2012).

The degrees of network vertices are one of the most important network statistics and provide important insights to understand more complex network structures such as the "small-world phenomenon" in Chung and Lu (2002) and some refined network statistics (e.g., "alternating k-star statistic") developed by Snijders et al. (2006). One natural approach for modeling the degrees is to put them as sufficient statistics for exponential family distributions on graphs according to the Koopman–Pitman–Darmois theorem (Koopman, 1936; Pitman, 1936; Darmois, 1935) or the principle of maximum entropy. Chatterjee et al. (2011) coined this model as β-model and proved the uniform consistency of the maximum likelihood estimator (MLE); Yan and Xu (2013) established its asymptotic normality. Hillar and Wibisono (2013) generalized the β-model to weighted edges and also proved that the MLE is uniformly consistent.

In this note we further establish the asymptotic normality of one of the score test statistics under the generalized β-model for the finite discrete weighted edges, i.e., $(2n)^{-1/2}(T-n) \xrightarrow{D} N(0,1)$ as n goes to infinity, where T is the score test

* Corresponding author.
E-mail addresses: tingyanty@mail.ccnu.edu.cn (T. Yan), yzhao15@gmu.edu (Y. Zhao).

http://dx.doi.org/10.1016/j.spl.2016.07.022

statistic under a simple null and n is the number of vertices of the network. In this model, each vertex is assigned with one parameter to measure its strength to participate in network connection. As n increases, the number of parameter diverges. It makes the asymptotic inference nonstandard. To the best of our knowledge, this is the first asymptotic result for the score test with a diverging number of parameters in network models as n the number of vertices of the network goes to infinity. A key step in the proof is that we use the approximate inverse of the Fisher information matrix in Yan and Xu (2013) to obtain the approximately explicit expression for T. Another technical step is to construct a dependency graph to obtain the asymptotic distribution for the weighted quadratic sum of the centered degrees.

The rest of the paper is organized as follows. In Section 2, we lay out the main result, i.e., the asymptotic distribution of our score test statistic in the generalized β-model. Simulation studies are provided in Section 3. Section 4 concludes with some discussion and future work. All proofs are given in Appendix A.

2. Main results

Consider an undirected graph $\mathcal{G}_n$ on n vertices labeled by $1, \ldots, n$ with no self-loops. Let a_{ij} be the weight of edge (i, j) taking q discrete values from the set $\Omega = \{0, \ldots, q-1\}$, where q is assumed to be fixed. The β-model is only involved with the binary edges (i.e., $q = 2$). Here, we consider a generalized β-model for finite discrete weighted edges proposed by Hillar and Wibisono (2013). Rinaldo et al. (2013) defined a different version of generalized β-model by assuming that a_{ij} comes from n Bernoulli trials, which we did not consider here. Define the degree of vertex i by $d_i = \sum_{j\neq i} a_{ij}$ and the degree sequence of $\mathcal{G}_n$ by $\mathbf{d} = (d_1, \ldots, d_n)^T$. The probability density function of $\mathcal{G}_n$ under the generalized β-model is

$$p_\beta(\mathcal{G}_n) = \exp\big(\boldsymbol{\beta}^\top \mathbf{d} - z(\boldsymbol{\beta})\big), \tag{1}$$

where $z(\boldsymbol{\beta})$ is the normalizing constant. The parameters $\beta_1, \ldots, \beta_n$ measure the strength of each vertex participating in network connection. It can be obtained from (1) that the edges a_{ij}s for all $1 \le i < j \le n$ are mutually independent with the probability:

$$P(a_{ij} = a) = \frac{e^{a(\beta_i+\beta_j)}}{\sum_{k=0}^{q-1} e^{k(\beta_i+\beta_j)}}, \quad a = 0, 1, \ldots, q-1.$$

When $q = 2$, it reduces to the β-model.

In general, only one realization of a random network is observed. Based on a single observed network, the log-likelihood function is $\ell(\boldsymbol{\beta}) = \boldsymbol{\beta}^\top \mathbf{d} - z(\boldsymbol{\beta})$. The solution to $\nabla z(\boldsymbol{\beta}) = \mathbf{d}$ is the MLE of $\boldsymbol{\beta}$ and $E_\beta(\mathbf{d}) = \nabla z(\boldsymbol{\beta})$ by the property of exponential family (Brown, 1986)]. For convenience, we will suppress the subscript β hereafter. Let $V_n = (v_{ij})_{i,j=1,\ldots,n}$ be the Fisher information matrix of the parameters $\beta_1, \ldots, \beta_n$, which is also the covariance matrix of $\mathbf{d}$. By Yan et al. (2015), we have

$$v_{ij} = \frac{\sum_{0\le k<l\le q-1} (k-l)^2 e^{(k+l)(\beta_i+\beta_j)}}{\left(\sum_{a=0}^{q-1} e^{a(\beta_i+\beta_j)}\right)^2}, \quad j \neq i, \ v_{ii} = \sum_{j\neq i} v_{ij}.$$

V_n is the diagonally dominant matrix with nonnegative entries. Let $U(\boldsymbol{\beta})$ be the score function of the log-likelihood $\ell(\boldsymbol{\beta})$:

$$U(\boldsymbol{\beta}) = \frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{d} - E(\mathbf{d}),$$

and define

$$T_n(\boldsymbol{\beta}) = U^\top(\boldsymbol{\beta})[V_n(\boldsymbol{\beta})]^{-1}U(\boldsymbol{\beta}).$$

Then $T_n(\boldsymbol{\beta}_0)$ is the score test statistic under the simple null: $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$. We will investigate the asymptotic distribution of $T_n(\boldsymbol{\beta})$ for a general $\boldsymbol{\beta}$ as n goes to infinity. We use the notations T_n, V_n, U instead of $T_n(\boldsymbol{\beta}), V_n(\boldsymbol{\beta}), U(\boldsymbol{\beta})$ hereafter for convenience.

The dimensions of U and V_n will increase with n, which makes the study of the asymptotic distribution of T_n difficult. In order to obtain the asymptotic distribution of T_n, we need to handle the inverse of V_n with a large dimension. Since V_n^{-1} does not have a closed form, we use its approximation $S_n = \mathrm{diag}(1/v_{11}, \ldots, 1/v_{nn})$ proposed by Yan et al. (2015), whose approximate error is given in Proposition 1 of their paper. As a result, T_n is divided into the sum of two parts: the quadratic sum $\sum_i (d_i - E(d_i))^2/v_{ii}$ and a remainder. Therefore, the object is to derive the asymptotic distribution of $\sum_i (d_i - E(d_i))^2/v_{ii}$ and bound the remainder. For any fixed k, the vector $((d_{i_1} - E(d_{i_1}))/v_{i_1,i_1}^{1/2}, \ldots, (d_{i_k} - E(d_{i_k}))/v_{i_k,i_k}^{1/2})$ is asymptotically independent standard normal as shown in Proposition 2 of Yan et al. (2015). In order to obtain the asymptotic distribution of $\sum_i (d_i - E(d_i))^2/v_{ii}$, we consider a more general weighted quadratic sum $\sum_i c_i(d_i - E(d_i))^2$. Let $\tilde{a}_{ij} = a_{ij} - E(a_{ij})$ be the centered random variable of a_{ij} and define $\tilde{a}_{ii} = 0$ for all $i = 1, \ldots, n$. By noting that $Var(\mathbf{d}) = V_n$, the expectation and variance of this sum are given below, the calculation of which is given in the supplementary material (see Appendix B).

Property 1. (1) $E[\sum_i c_i(d_i - E(d_i))^2] = \sum_i c_i v_{ii}$.
(2) *Let* $u_{ij} = Var(\tilde{a}_{ij})^2$. *Then*

$$Var\left[\sum_i c_i(d_i - E(d_i))^2\right] = \sum_i c_i^2(2v_{ii}^2 + \sum_{j\neq i} u_{ij}) + 2\sum_{i<j} c_ic_ju_{ij}. \tag{2}$$

The central limit theorem for the weighted quadratic sum $\sum_i c_i(d_i - E(d_i))^2$ is stated below, whose proof is given in Appendix A.1 by constructing a dependency graph.

Theorem 1. *Assume that* $\rho_1 \le c_i \le \rho_2$ *for all* $i = 1, \dots, n$, *and* $v_* \le v_{ij} \le v_{**}$ *for all* $i \neq j$, *and* $\min_{i,j} u_{i,j} \ge u_*$. *If the following holds:*

$$\rho_2^6 v_{**}/\rho_1^5 v_*^{5/2} = o(n^{1/2}), \quad u_*/v_* = o(n), \tag{3}$$

then the weighted sum $\sum_i c_i(d_i - E(d_i))^2$ *is asymptotically normally distributed with mean* $\sum_i c_iv_{ii}$ *and variance* $\sum_i c_i^2(\sum_{j\neq i} u_{ij} + 2v_{ii}^2) + 2\sum_{1\le i<j\le n} c_ic_ju_{ij}$.

Remark 1. Ouadah et al. (2015) introduce the degree mean square statistic $\frac{1}{n}\sum_i(d_i - E(d_i))^2$ under the heterogeneous Erdős–Rényi model, which assumes that all edges are independent Bernoulli random variables with their respective successful probabilities. By taking $c_i = 1$, it immediately follows from Theorem 1 that the central limit theorem for the degree mean square statistic holds.

If $v_* \le v_{ij} \le v_{**}$ and $u_{**}/v_*^2 = o(n)$ with $u_{ij} \le u_{**}$ for all $i \neq j$, then $1/v_{**} \le (n-1)/v_{ii} \le 1/v_*$ and

$$\lim_{n\to\infty} \frac{1}{2n}\left(\sum_i \frac{1}{v_{ii}^2}(2v_{ii}^2 + \sum_{j\neq i} u_{ij}) + 2\sum_{i\neq j} \frac{u_{ij}}{v_{ii}v_{jj}}\right) = 1.$$

Taking $\rho_1 = 1/v_{**}$ and $\rho_2 = 1/v_*$ and $c_i = (n-1)/v_{ii}$ in Theorem 1, it follows that

Corollary 1. *If* $v_{**}^6/v_*^{17/2} = o(n^{1/2})$ *and* $u_{**}/v_*^2 = o(n)$, *then* $\{\sum_{i=1}^n [d_i - E(d_i)]^2/v_{ii} - n\}/(2n)^{1/2}$ *converges in distribution to the standard normal distribution.*

Let $Q_n := 2\max_i |\beta_i|$. By Yan et al. (2015) (see inequalities [3] and [4] in their paper), the variance v_{ij} of a_{ij} satisfies:

$$\frac{1}{2(1+e^{Q_n})} \le v_{ij} \le \frac{q^2}{2}, \quad 1 \le i \neq j \le n. \tag{4}$$

Since $\max_{i,j} a_{ij} \le q-1$, u_{**} is bounded by a constant. Applying Corollary 1, we have:

Corollary 2. *If* $e^{17Q_n/2} = o(n^{1/2})$, *then* $\{\sum_i [d_i - E(d_i)]^2/v_{ii} - n\}/(2n)^{1/2}$ *converges in distribution to the standard normal distribution.*

The central limit theorem for the score test statistic is stated as follows.

Theorem 2. *If* $e^{17Q_n/2} = o(n^{1/2})$, *then the score test statistic* T_n *is asymptotically normally distributed in the sense that*

$$\frac{T_n - n}{\sqrt{2n}} \xrightarrow{D} N(0,1), \quad \text{as } n \to \infty. \tag{5}$$

3. Simulations

In this section, we will evaluate the asymptotic result in Theorem 2 through numerical simulations. The quantile–quantile (QQ) plots of $(T_n - n)/(2n)^{1/2}$ are shown in Fig. 1. Following Yan and Xu (2013), the parameter settings in simulation studies are listed as follows. Let $\beta_i = iM/n$ and qth count of the possible values for edges be 3 for simplicity; A variety of M and n are chosen: $M = 0, \log(\log n), (\log n)^{1/2}, \log n$, and $n = 100, 200, 1000$.

The QQ plots in Fig. 1 are based on 10,000 repetitions for each scenario. The horizontal and vertical axes are the empirical and theoretical quantiles, respectively. The gray lines correspond to $y = x$. The QQ plots for $n = 200$ are similar to those for $n = 100$ and not shown here to save space. We also do not show the pictures for $M = 0$ which are similar to $M = \log(\log n)$. In Fig. 1, when $M = \log(\log n), (\log(n))^{1/2}$, the sample quantiles deviate from the theoretical ones slightly. This phenomenon disappears when n increases to 1000 and the sample quantiles coincide with the theoretical ones very well. On the other hand, when $M = \log(n)$, the sample quantile evidently deviates from the theoretical one for both $n = 100$ and $n = 1000$.

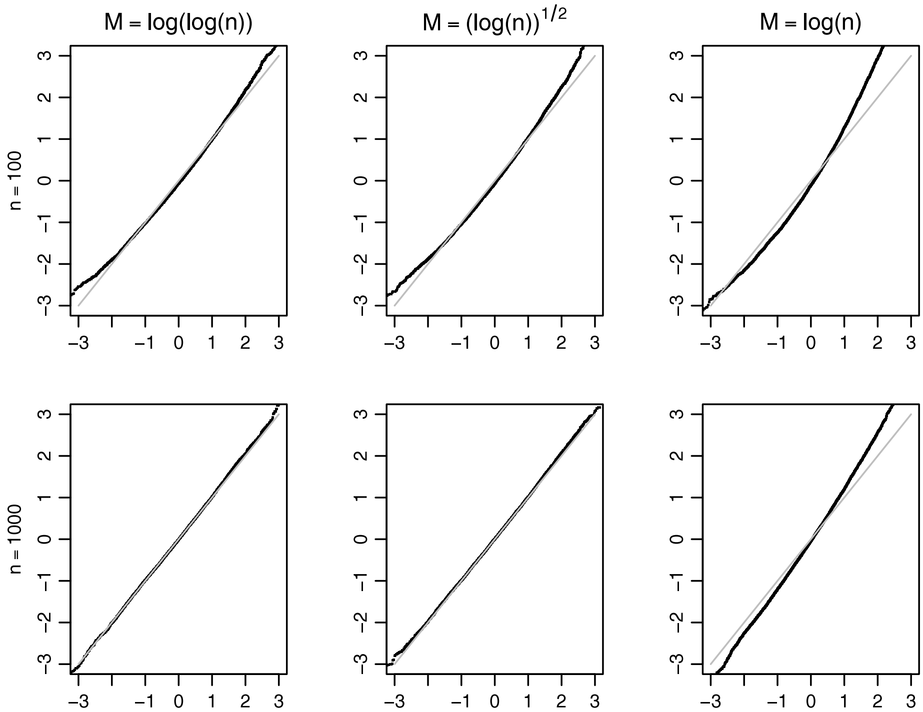

Fig. 1. The QQ plots of $(T - n)/(2n)^{1/2}$.

Table 1
Estimated coverage probabilities of T.

n	$M = 0$	$M = \log(\log n)$	$M = (\log n)^{1/2}$	$M = \log(n)$
100	95.28	94.76	94.42	88.38
200	95.09	94.83	94.70	88.99
1000	94.98	94.93	94.75	89.61

The 95% coverage frequencies for T lying in the interval $[n - z_\alpha(2n)^{1/2}, n + z_\alpha(2n)^{1/2}]$ are reported in Table 1, where z_α is the α-quantile of the standard normality and $\alpha = 0.975$. From this table, we can see that the estimated coverage frequencies are close to the target level 95% when $M \le (\log n)^{1/2}$ while those are less than 95% when $M = \log(n)$. In both the QQ plot and the table, when $M \le (\log n)^{1/2}$, the asymptotic approximation is very good. It may show that there exists room for improvement on the increasing rate of Q_n in Theorem 2. On the other hand, the asymptotic approximation is not good when $M = \log(n)$. This shows that it is necessary to restrict the increasing rate of Q_n in order to guarantee the asymptotic normality for the score test statistic.

4. Summary and discussion

We have established the central limit theorem for one of the score test statistics in the generalized β-model when the number of parameters goes to infinity. Along the way, we derive the central limit theorem for the weighted quadratic sum of degrees. If a_{ij}'s, $1 \le i < j \le n$ are mutually independent and bounded, then Theorem 2 still holds. The simulation results show that even if $Q_n = (\log n)^{1/2}$, the asymptotic approximation is still very good. It would be of interest to investigate whether the condition imposed on Q_n in Theorem 2 could be relaxed.

Hillar and Wibisono (2013) considered two other types of weights (i.e., continuous and infinite discrete). In both cases, condition (6) holds no longer since it is an unbounded random variable. It is interesting to seek for other conditions to replace (6) so that the unbounded random variables can be covered. We note that the result may be extended to more general scenarios. For a class of n-parameter network models with the generalized β-model as a special case, Yan et al. (2016) have obtained the consistency and asymptotic normality of the moment estimator. It could be expected that there may be similar results in the model considered in Yan et al. (2016) as well as other degree-based models (Chung and Lu, 2002; Perry and Wolfe, 2012).

Acknowledgments

We are very grateful to an anonymous referee and the Editor for their valuable comments that have greatly improved the manuscript. Yan's research is partially supported by the National Natural Science Foundation of China (No. 1140239).

Appendix A

We introduce one useful theorem which will be used in the proofs. Let $\{X_i\}_{i\in\mathcal{I}}$ be a family of random variables (defined on a common probability space). *A dependency graph* for $\{X_i\}$ is any graph L with vertex set $V(L)=\mathcal{I}$ such that if A and B are two disjoint subsets of $\mathcal{I}$ with $e_L(A,B)=0$, then the families $\{X_i\}_{i\in A}$ and $\{X_i\}_{i\in B}$ are mutually independent, where $e_L(A,B)$ denotes the number of edges connecting one vertex in A and another in B. For any integer $r\ge 1$ and $i_1,\ldots,i_r\in\mathcal{I}$, denote by

$$\overline{N}_L(i_1,\ldots,i_r)=\bigcup_{k=1}^{r}\{j\in\mathcal{I}: j=i_k \text{ or } j \text{ is adjacent to } i_k \text{ in } L\}$$

the closed neighborhood of $\{i_1,\ldots,i_r\}$ in L. If the family $\{X_i\}_{i\in\mathcal{I}}$ can construct a dependency graph, then the normalized sum $\sum_i X_i$ converges to the standard normal distribution (Mikhailov, 1991). Here we use an improved version, i.e., Theorem 6.21 in Janson et al. (2000) stated as one theorem below.

Theorem 3 (Janson et al., 2000). *Suppose that $\{S_n\}_{n=1}^{\infty}$ is a sequence of random variables such that $S_n=\sum_{\alpha\in\mathcal{A}_n}X_{n\alpha}$, where for each n, $\{X_{n\alpha}\}_{\alpha\in\mathcal{A}_n}$ is a family of random variables with dependency graph L_n. Suppose further that there exist numbers M_n, K_n and B_r such that $\sum_{\alpha\in\mathcal{A}_n}E|X_{n\alpha}|\le M_n$ and, for every n and $r\ge 1$, and $\alpha_1,\ldots,\alpha_r\in\mathcal{A}_n$,*

$$\sum_{\alpha\in\overline{N}_{L_n}(\alpha_1,\ldots,\alpha_r)} E(|X_{n\alpha}||X_{n\alpha_1},\ldots,X_{n\alpha_r})\le B_rK_n. \tag{6}$$

Let $\sigma_n^2=Var(S_n)$. If $M_nK_n^{s-1}/\sigma_n^s\to 0$ for some real $s>2$, then $\sigma_n^{-1}(S_n-E(S_n))$ converges to the standard normal distribution.

A.1. Proof of Theorem 1

Proof of Theorem 1. Let $\tilde{a}_{ij}=a_{ij}-E(a_{ij})$, $z_{ij}=(c_i+c_j)\tilde{a}_{ij}^2$, $z_{ijk}=2c_i\tilde{a}_{ij}\tilde{a}_{ik}$, $\Gamma_1=\{(i,j):1\le i<j\le n\}$ and $\Gamma_2=\{(i,j,k):1\le i\le n;1\le j<k\le n;j,k\neq i\}$. Denote the set of all z_{ij}s with $(i,j)\in\Gamma_1$ and all z_{ijk}s with $(i,j,k)\in\Gamma_2$ by $\{X_{n\alpha}\}_{\alpha\in\mathcal{A}_n}$, where $\mathcal{A}_n=\Gamma_1\bigcup\Gamma_2$. For simplicity, we omit the subscript n in $X_{n\alpha}$. Then the cardinality of $\mathcal{A}_n$ is:

$$|\mathcal{A}_n|=\left|\Gamma_1\bigcup\Gamma_2\right|=\binom{n}{2}+3\binom{n}{3}=\frac{1}{2}n(n-1)^2. \tag{7}$$

It is easy to verify that

$$\sum_{i=1}^{n}c_i(d_i-E(d_i))^2=\sum_{1\le i<j\le n}(c_i+c_j)\tilde{a}_{ij}^2+\sum_{1\le j<k\le n,k,j\neq i}2c_i\tilde{a}_{ij}\tilde{a}_{ik}=\sum_{\alpha\in\mathcal{A}_n}X_\alpha.$$

We start by outlining the idea of the proof: We will use Theorem 3 to obtain the limiting distribution of the above summation by constructing a dependency graph L_n. The graph L_n is made up of a set of vertices $\{\alpha:\alpha\in\mathcal{A}_n\}$ and edges by connecting every pair of vertices α and β such that X_α and X_β share at least one common random variable $\tilde{a}_{kl}\in\{\tilde{a}_{ij}:1\le i<j\le n\}$. It is clearly a dependency graph for $\{X_\alpha\}_{\alpha\in\mathcal{A}_n}$. According to the definition of X_α, for different $\alpha,\beta\in\mathcal{A}_n$, X_α and X_β share at most one common random variable.

We verify the conditions of Theorem 3 as follows. Recall the notations used in Theorem 1: $V_n=(v_{ij})$ is the Fisher information matrix. v_* and v_{**} are the lower and upper bound of v_{ij} for $i\neq j$. ρ_1 and ρ_2 are the lower and upper bound of c_i. $u_{ij}=Var(\tilde{a}_{ij})^2$ and u_* is the lower bound of u_{ij}. Then,

$$E(\tilde{a}_{ij}^2)=v_{ij}\le v_{**},\qquad E|\tilde{a}_{ij}\tilde{a}_{ik}|\le\frac{1}{2}E(\tilde{a}_{ij}^2+\tilde{a}_{ik}^2)=\frac{1}{2}(v_{ij}+v_{ik})\le v_{**},$$

and $\rho_1\le c_i\le\rho_2$. Therefore, $\sum_{\alpha\in\mathcal{A}_n}E|X_\alpha|\le 2\rho_2v_{**}|\mathcal{A}_n|$. By equality (7), $|\mathcal{A}_n|=\frac{1}{2}n(n-1)^2$, so M_n in Theorem 3 can be chosen as $n^3\rho_2v_{**}$.

Next, we will calculate the maximum value of the cardinality of $\overline{N}_{L_n}(\alpha_1,\ldots,\alpha_r)$. We say that T_α is an edge with vertices i and j if $X_\alpha=z_{ij}$; T_α is a two-path with vertices i,j and k if $X_\alpha=z_{ijk}$. Then there is a one-to-one mapping between X_α and T_α. Suppose that $\alpha_1,\ldots,\alpha_r\in\mathcal{A}_n$ are given. Consider the union $\bigcup_{i=1}^{r}T_{\alpha_i}$, whose vertices construct a vertex set V. Let K_V be the complete graph on V. Note that K_V has at most $3r$ vertices, and the cardinality of $\overline{N}_{L_n}(\alpha_1,\alpha_2)\bigcap\{\alpha:T_\alpha\subset K_V\}$ is not more than $2\cdot(3r)^2=18r^2$. Moreover, each T_β for $\beta\in\overline{N}_{L_n}(\alpha_1,\ldots,\alpha_r)\setminus\{\alpha:T_\alpha\subset K_V\}$ is such a path that one edge is in K_V

and another one is not, hence the number of such β is less than $8n$. Recall that a_{ij} takes values from the set $\{0, 1, \ldots, q-1\}$ and q is a fixed integer. Since $|\tilde{a}_{ij}| \le (q-1)^2$, we have that for any $\alpha, \alpha_1, \ldots, \alpha_r \in \mathcal{A}_n$,

$$E[|X_\alpha||X_{\alpha_1}, \ldots, X_{\alpha_r}] \le 2(q-1)^2,$$

such that

$$\begin{aligned} \sum_{\alpha \in \overline{N}_{L_n}(\alpha_1,\ldots,\alpha_r)} E[|X_\alpha||X_{\alpha_1}, \ldots, X_{\alpha_r}] &\le 2(q-1)^2 |\overline{N}_{L_n}(\alpha_1, \ldots, \alpha_r)| \\ &\le 2(q-1)^2(18r^2 + 8n)\rho_2 \\ &< [2(q-1)^2(8 + 18r^2)]\rho_2 n. \end{aligned}$$

So we could choose B_r in Theorem 3 as $2(q-1)^2(8+18r^2)$ and K_n as $n\rho_2$. This shows that condition (6) holds.

By equality (2), we have

$$Var\left[\sum_i c_i \tilde{d}_i^2\right] \ge n\rho_1^2[(n-1)u_* + 2(n-1)^2 v_*] + n(n-1)\rho_1^2 u_*.$$

It follows that for $s = 5$,

$$\begin{aligned} \frac{M_n K_n^{5-1}}{(Var(Z_n))^{5/2}} &\le \frac{n^3 \rho_2 v_{**} \cdot (\rho_2 n)^5}{[\rho_1^2(nu_* + 2n(n-1)^2 v_* + n(n-1)u_*)]^{5/2}} \\ &= O\left(\frac{\rho_2^6 v_{**}}{\rho_1^5 v_*^{5/2} n^{1/2}} \times \frac{1}{\left(2 + \frac{u_*}{nv_*} + \frac{u^*}{nv_*}\right)^{5/2}}\right) = o(1). \end{aligned}$$

This complete the proof. □

A.2. Proof of Theorem 2

Proof of Theorem 2. Let $S_n = \text{diag}(1/v_{11}, \ldots, 1/v_{nn})$ and $W_n = V_n^{-1} - S_n$. By Corollary 2, it is sufficient to show

$$\frac{(\mathbf{d} - E(\mathbf{d}))^\top W_n (\mathbf{d} - E(\mathbf{d}))}{\sqrt{2n}} = o_p(1). \tag{8}$$

in order to prove Theorem 2.

First, we have

$$\begin{aligned} E[\{\mathbf{d} - E(\mathbf{d})\}^T W_n \{\mathbf{d} - E(\mathbf{d})\}] &= E[tr(\{\mathbf{d} - E(\mathbf{d})\}^T W_n \{\mathbf{d} - E(\mathbf{d})\})] \\ &= tr\{W_n E[\{\mathbf{d} - E(\mathbf{d})\}\{\mathbf{d} - E(\mathbf{d})\}^T]\} \\ &= tr(I_n - S_n V_n) = n - \sum_{i=1}^n \frac{v_{ii}}{v_{ii}} = 0. \end{aligned}$$

For simplifying the notation, we denote $\tilde{d}_i = d_i - E(d_i)$. Therefore, it is sufficient to show

$$\frac{\text{Var}\left(\sum_{i,j=1}^n \tilde{d}_i w_{ij} \tilde{d}_j\right)}{2n} = o(1).$$

in order to prove (8). There are four cases for calculating the covariance $Cov(\tilde{d}_i\tilde{d}_j, \tilde{d}_h\tilde{d}_g)$.

Case 1: $i = j = h = g$. Similar to the calculations of (2), we have $Cov\{\tilde{d}_i^2, \tilde{d}_i^2\} = 2v_{ii}^2 + \sum_{j\ne i} u_{ij}$.

Case 2: Three indices among the four indices i, j, h, g are the same (e.g. $j = h = g$). In order to guarantee non-zero covariance of $\tilde{a}_{i\alpha}\tilde{a}_{j\beta}$ and $\tilde{a}_{jk}\tilde{a}_{jl}$ with $i \ne j$, $(\tilde{a}_{i\alpha}, \tilde{a}_{j\beta})$ and $(\tilde{a}_{jk}, \tilde{a}_{jl})$ must have at least one common element. By noting that $E(\tilde{a}_{ij}) = 0$ and $\tilde{a}_{ij}$, $1 \le i < j \le n$ are mutually independent, the only case for non-zero covariance is $\alpha = k = l = j$. Therefore, we have

$$Cov(\tilde{d}_i\tilde{d}_j, \tilde{d}_j^2) = Cov\left(\sum_{\alpha\ne i, \beta\ne j} \tilde{a}_{i\alpha}\tilde{a}_{j\beta}, \sum_{k\ne j, l\ne j} \tilde{a}_{jk}\tilde{a}_{jl}\right) = Cov(\tilde{a}_{ij}\tilde{a}_{ji}, \tilde{a}_{ji}\tilde{a}_{ji}).$$

There are similar arguments for the other two cases below.

Case 3: Two indices among the four indices i, j, h, g are the same (e.g. $i = j$ or $j = h$).

$$Cov(\tilde{d}_i^2, \tilde{d}_g\tilde{d}_h) = 2Cov(\tilde{a}_{ig}\tilde{a}_{ih}, \tilde{a}_{gi}\tilde{a}_{hj}),$$
$$Cov(\tilde{d}_i\tilde{d}_j, \tilde{d}_j\tilde{d}_g) = Cov(\tilde{a}_{ij}\tilde{a}_{jg}, \tilde{a}_{ji}\tilde{a}_{gj}) + \sum_{\beta\neq i} Cov(\tilde{a}_{ig}\tilde{a}_{j\beta}, \tilde{a}_{j\beta}\tilde{a}_{gi}).$$

Case 4: All four indices i, j, h, g are different.

$$Cov(\tilde{d}_i\tilde{d}_j, \tilde{d}_h\tilde{d}_g) = Cov(\tilde{a}_{ih}\tilde{a}_{jg}, \tilde{a}_{hi}\tilde{a}_{gj}) + Cov(\tilde{a}_{ig}\tilde{a}_{jh}, \tilde{a}_{hj}\tilde{a}_{gi}).$$

Let $g_{ijhg} = Cov(\tilde{d}_i w_{ij}\tilde{d}_j, \tilde{d}_h w_{hg}\tilde{d}_g)$. By noting that $|\tilde{a}_{ij}| \le q$, for different i, j, h, g,

$$|g_{iiii}| \le w_{ii}^2(2(n-1)^2 + (n-1))q^4, \qquad |g_{ijjj}| \le q^4|w_{ij}w_{jj}|, \qquad |g_{iigh}| \le 3(n-1)q^4|w_{ii}w_{hg}|,$$
$$|g_{iijg}| \le 3(n-1)q^4|w_{ii}w_{jg}|, \qquad |g_{ijhg}| \le |w_{ij}w_{hg}|q^4.$$

By Lemma 1 in the supplementary material and inequality (4), we have

$$\|W_n\| = O\left(\frac{M^2}{m^3n^2}\right) = O\left(\frac{(1+e^{Q_n})^3}{n^2q^4}\right).$$

Consequently, if $e^{Q_n} = o(n^{1/6})$, then we have

$$\frac{\text{var}[\{\mathbf{d}_n - E(\mathbf{d}_n)\}^\top W_n\{\mathbf{d}_n - E(\mathbf{d}_n)\}]}{2n} \le \frac{1}{2n}O\left(\left(\frac{(1+e^{Q_n})^3}{n^2q^4}\right)^2\right) \times \left[n(2n^2+n) + \binom{n}{2} + \binom{n}{3}\cdot 3n + 2\binom{n}{4}\right]\frac{1}{q^4}$$
$$= O\left(\frac{e^{6Q_n}}{n}\right) = o(1).$$

The condition $e^{17Q_n/2} = o(n^{1/2})$ in Theorem 2 implies that $e^{Q_n} = o(n^{1/17})$. Therefore, equality (8) holds. This completes the proof. □

Appendix B. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.spl.2016.07.022.

References

Bhattacharyya, S., Bickel, P.J., 2015. Spectral clustering and block models: a review and a new algorithm. Preprint. Available at http://arxiv.org/abs/1508.01819.

Brown, L.D., 1986. Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory. In: Institute of Mathematical Statistics Lecture Notes – Monograph Series, vol. 9. IMS, Hayward, CA.

Chatterjee, S., Diaconis, P., Sly, A., 2011. Random graphs with a given degree sequence. Ann. Appl. Probab. 21, 1400–1435.

Chung, F., Lu, L., 2002. The average distances in random graphs with given expected degrees. Proc. Natl. Acad. Sci. USA 99, 15879–15882.

Darmois, G., 1935. Sur les lois de probabilité à estimation exhaustiv. C. R. Acad. Sci., Paris 260, 1265–1266.

Fienberg, S.E., 2012. A brief history of statistical models for network analysis and open challenges. J. Comput. Graph. Statist. 21, 825–839.

Goldenberg, A., Zheng, A.X., Fienberg, S.E., Airoldi, E.M., 2010. A survey of statistical network models. Found. Trends Mach. Learn. 2, 129–233.

Hillar, C., Wibisono, A., 2013. Maximum entropy distributions on graphs. Preprint. Available at http://arxiv.org/abs/1301.3321.

Koopman, L.H., 1936. On distributions admitting a sufficient statistic. Trans. Amer. Math. Soc. 39, 399–409.

Janson, S., Luczak, T., Ruciński, A., 2000. Random Graphs. Wiley-Interscience, New York.

Matias, C., Robin, S., 2014. Modeling heterogeneity in random graphs through latent space models: a selective review. ESAIM Proc. Surv. 47, 55–74.

Mikhailov, V.G., 1991. On a theorem of Janson. Teor. Veroyatn. Primen. 36, 168–170 (in Russian); an English translation appears in Theory Probab. Appl. 36, 173–176.

Ouadah, S., Robin, S., Latouche, P., 2015. A degree-based goodness-of-fit test for heterogeneous random graph models. Preprint. Available at http://arxiv.org/abs/1507.08140.

Perry, P.O., Wolfe, P.J., 2012. Null models for network data. Preprint. Available at http://arxiv.org/abs/1201.5871.

Pitman, E.J.G., 1936. Sufficient statistics and intrinsic accuracy. Proc. Camb. Phil. Soc. 32, 567–579.

Rinaldo, A., Petrović, S., Fienberg, S.E., 2013. Maximum likelihood estimation in the β-model. Ann. Statist. 41, 1085–1110.

Snijders, T.A.B., Pattison, P.E., Robins, G.L., Handcock, M.S., 2006. New specifications for exponential random graph models. Sociol. Methodol. 36, 99–153.

Yan, T., Qin, H., Wang, H., 2016. Asymptotics in undirected random graph models parameterized by the strengths of vertices. Statist. Sinica 26, 273–293.

Yan, T., Xu, J., 2013. A central limit theorem in the β-model for undirected random graphs with a diverging number of vertices. Biometrika 100, 519–524.

Yan, T., Zhao, Y., Qin, H., 2015. Asymptotic normality in the maximum entropy models on graphs with an increasing number of parameters. J. Multivariate Anal. 133, 61–76.