Undirected network models with degree heterogeneity and homophily

Ting Yan*
Central China Normal University

Abstract

The degree heterogeneity and homophily are two typical features in network data. In this paper, we formulate a general model for undirected networks with these two features and present the moment estimation for inferring the degree and homophily parameters. Our model only specifies a marginal distribution on each edge in weighted or unweighted graphs and admits the non-independent dyad structures unlike previous works that assume independent dyads. We establish a unified theoretical framework under which the consistency of the moment estimator hold as the size of networks goes to infinity. We also derive its asymptotic representation that can be used to characterize its limiting distribution. The asymptotic representation of the estimator of the homophily parameter contains a bias term. Accurate inference necessitates bias-correction. Several applications are provided to illustrate the unified theoretical result.

Key words: Asymptotical representation; Consistency; Moment estimation; Network data

Mathematics Subject Classification: 62F12, 91D30.

1 Introduction

Networks/graphs provide a natural way to represent many complex interactive behaviors among a set of actors, where each node represents an actor and an edge exists between two nodes if the two corresponding actors interact in some way. The types of interactions could be friendships between peoples, follow between users in social media such as Twitter,

*Department of Statistics, Central China Normal University, Wuhan, 430079, China. Email: tingyanty@mail.ccnu.edu.cn.
citations between papers, hyperlinks between webs and so on. The edges can be directed or undirected, binary (when each edge is either present or absent) or weighted (when a discrete value is recorded for each observed edge). With the demand of research for a variety of purposes, more and more network data sets are collected and stored. At the same time, statistical network analysis have made great process in recent years and many approaches are developed; see Robins et al. (2007), Goldenberg et al. (2010), Fienberg (2012) for some recent reviews. The book by Kolaczyk (2009) provides a comprehensive description on statistical analyses of network data.

One of the most important features of network data is the degree heterogeneity that characterizes the variation in the node degrees. As an example in a well-known yeast dataset [von Mering et al. (2002)] available at the R package “igraphdata”, the node degree varies from the minimum value 1 to the maximum value 118 in its largest connected subgraph that has 2375 nodes. To model the degree heterogeneity, a class of node-parameter models are proposed, in which each node degree is attached to one parameter. Holland and Leinhardt (1981) is generally acknowledged as the first one to model the degree variation, who proposed the $p_1$ model in which the bi-degrees of nodes and the number of reciprocated dyads form the sufficient statistics for the exponential distribution on directed graphs. Other node-parameter models include the Chung-Lu model [Chung and Lu (2002)] with the expected degrees as the parameters, the $\beta$-model [Chatterjee et al. (2011); Blitzstein and Diaconis (2011); Park and Newman (2004)], null models [Perry and Wolfe (2012)] and maximum entropy models for weighted graphs [Hillar and Wibisono (2013)]. In these models, the number of parameters increases as the network size grows such that asymptotic inference is nonstandard. The theoretical properties of the maximum likelihood estimators (MLEs) or moment estimators have been derived until recently [e.g., Chatterjee et al. (2011); Rinaldo et al. (2013); Hillar and Wibisono (2013); Yan et al. (2016a,b); Yan and Xu (2013); Yan et al. (2015)]. In particular, Yan et al. (2016b) establish a unified theoretical framework for this class of models under which the consistency and asymptotical normality of the estimator hold.

Another important feature commonly existing in social and econometric network data is the homophily on individual-level attributes—a phenomenon that the individuals tend to form connections with those like themselves [e.g., McPherson et al. (2001); Kossinets and Watts (2006); Currrarini et al. (2009)]. The individual attributes may be immutable characteristics such as racial and ethnic groups and ages; it also may be mutable characteristics such as places living, occupations, levels of affluence, and interests. The presence of homophily has important implications on the network formation process. In one hand, it produces preferential selection—individuals are more easily interact with those with similar characteristics. In the other hand, the existing links create social influence: people may modify their behaviors to bring them more closely into alignment with the behaviors
of their associated ones.

The link formation is not only decided by the homophily effect but also the degree effect. As given in a toy example with a strong taste of homophily in Graham (2017), neglecting the degree effect might incorrectly conclude that preferences are heterophilic. To simultaneously model these two features, Graham (2017) proposes a link surplus model in which a link between two nodes is present only if the sum of a degree component and a homophily component exceeds a latent random variable drawn from the logistic distribution. By using a fixed effects method that treats the degree parameters as fixed values, he derives the consistency and asymptotic normal distribution of the MLE of the homophily parameter. Dzemski (2017) and Yan et al. (2018) derive the consistency and asymptotic distribution of the MLE in the directed link surplus model in which the latent random variables are drawn from the bivariate normal distribution and the logistic distribution, respectively. If the focus is only about the homophily parameter, then the conditional method can be used to eliminate the degree parameters in the case of logistic distribution [Graham (2017); Jochmans (2017)]. Another way to address the degree parameters is to treat them as the random effects and inference are performed by using Bayesian methods [e.g., van Duijn et al. (2004); Krivitsky et al. (2009); Hoff (2005)]. In contrast with the random effects method, the joint distribution of the degree heterogeneity and homophily component is left unrestricted in the fixed effects method.

The contributions of this paper are as follows. First, we formulate a general network model with degree effects and homophily effects for weighted or unweighted graphs by only specifying the marginal distribution of each edge. In contrast with previous works [e.g., Graham (2017); Dzemski (2017); Yan et al. (2018)] that assume independent dyad structures, our model admit the dependent dyads and also generalized their models to weighted edges. Second, we establish a unified theoretical framework under which the consistency of the moment estimator hold as the number of nodes goes to infinity. Unlike Graham (2017) who works with the restricted MLE that restricts the maximal optimal problem of the likelihood function to a compact set, our estimator is left unrestricted. If the marginal distribution belongs to the exponential family, then the moment estimator is identical to the MLE. Moreover, our result is general, not restricted to a specified marginal distribution. Even if the model is misspecified, the estimator is still consistent as long as those conditions hold. We also derive its asymptotic representation that can be used to characterize its limiting distribution, from which the central limit theorem holds if the sum of the observed dyads converges in distribution to the normal distribution. The asymptotic representation of the estimator of the homophily parameter contains a bias term. Accurate inference necessitates bias-correction. Finally, the unified theoretical framework is illustrated by an application.

For the rest of the paper, we proceed as follows. In Section 2, we introduce the model.
In Section 3, we present the estimation method. In Section 4, we present the consistency and asymptotic normality of the moment estimator. In Section 5, we illustrate the main result by an application. We make the summary and further discussion in Section 6. All proofs are relegated into Section ??.

2 Model

Let $G_n$ be an undirected graph on $n \geq 2$ nodes labeled by “1, …, n”. Let $A = (a_{ij})_{n \times n}$ be the adjacency matrix of $G_n$, where $a_{ij}$ is the weight of the edge between nodes $i$ and $j$. We don’t consider the self-loops here, i.e., $a_{ii} = 0$. The graph $G_n$ may be weighted or unweighted. If the edge weight $a_{ij}$ is an indicator (present or absent), then $G_n$ is unweighted (or called a simple graph). If $a_{ij}$ takes values from a set of positive integers (e.g., the number of papers collaborated by authors $i$ and $j$ in coauthor networks), then the graph $G_n$ is weighted. Moreover, $a_{ij}$ could be continuous (e.g., the calling time between two peoples in telephone networks). Let $d_i = \sum_{j \neq i} a_{ij}$ be the degree of node $i$ and $d = (d_1, \ldots, d_n)^T$ be the degree sequence of the graph $G_n$. We also observe a vector $z_{ij}$, the covariate information attached to the edge between nodes $i$ and $j$. The covariate $z_{ij}$ can be formed according to the similarity or dissimilarity between node attributes $x_i$ and $x_j$. Specifically, $z_{ij}$ can be represented through a symmetric function $g(\cdot, \cdot)$ with $z_i$ and $z_j$ as its arguments. As an example if $x_{i1}$ and $x_{i2}$ are location coordinates, then $z_{ij} = [(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2]^{1/2}$, denoting the Euclidean distance between $i$ and $j$.

We mainly focus on network models with two typical network features: the degree heterogeneity and homophily. The first is measured by a set of unobserved degree parameters $\{\beta_i\}_{i=1}^n$ and the second by the regression coefficients $\gamma$ of the pairwise covariates. Instead of imposing a global probability distribution on the graph $G_n$, we only specify the marginal distribution on each edge induced from some global probability distribution on $G_n$. We assume that the marginal probability density function of the edge variables $a_{ij}$’s conditional on the unobserved degree effects and observed covariates has the following form:

$$a_{ij} = a(z_{ij}, \beta_i, \beta_j) \sim f(a(z_{ij}, \gamma, \beta_i, \beta_j)), \quad (1)$$

where $f$ is a known probability density function, $\beta_i$ is the degree parameter of node $i$ and $\gamma$ is a dim $\gamma$-dimensional coefficient for the covariate $z_{ij}$. The parameter $\beta_i$ is the intrinsic individual effect that reflects the node heterogeneity to participate in network connection. The common parameter $\gamma$ is exogenous, measuring the homophily effect. If $f(\cdot)$ is an increasing function on $\beta_i$, then those nodes having relatively large degree parameters will have more links than those nodes with low degree parameters when the homophily component is the same level. A larger homophily component $z_{ij}^T \gamma$ means a
larger homophily effect.

The use of the marginal distribution has two advantages. First, it admit the non-independent dyadic structures unlike the previous works [e.g., Graham (2017); Dzemski (2017); Yan et al. (2018)] in which the mutually independent dyad assumption on the set of random variables \{(a_{ij}, a_{ji})\}_{i<j} is made. This is due to that different global distributions may lead to the same marginal distribution. If all dyads are independent, then our model framework is the same as in these works. The illustrated examples are given below. Second, it is enough for inferring degree and homophily parameters only to specify the edge marginal distribution under which we use the moment estimation and a unified theory for different models is established. For exponential-family distributions, the moment equation is identical to the maximum likelihood equation. Two running examples for illustrating the model are below.

**Example 1.** (Binary weight) Let \(a_{ij}\) be the binary weight of edge \((i, j)\), i.e., \(a_{ij} \in \{0, 1\}\), and \(F\) a cumulative distribution function. The marginal distribution of \(a_{ij}\) is
\[
\mathbb{P}(a_{ij} = a) = F^a(\beta_i + \beta_j + z_{ij}^\top \gamma)(1 - F(\beta_i + \beta_j + z_{ij}^\top \gamma))^{1-a}, \ a = 0, 1.
\]

Two common examples for \(F(\cdot)\) are the logistic function \((F(x) = e^x(1 + e^x)^{-1})\) [e.g., Graham (2017)] and probit function \((F(x) = \Phi(x))\) [e.g., Yan (2018)]. Here, the convention \(\Phi(x)\) is the cumulative distribution function for the standard normal random variable. An important application of the probit distribution is to model the dependent edge. Yan (2018) assumes that a link forms according to the rule:
\[
a_{ij} = 1(\beta_i + \beta_j + z_{ij}^\top \gamma > U_{ij}),
\]
where a latent vector \((U_{12}, U_{13}, \ldots,)\) is generated from a multivariate normal distribution with mean zeros and a standard covariance matrix with the diagonal elements 1 and non-diagonal elements \(\rho_{ij}\). Then the probability of a link between \(i\) and \(j\) is \(\Phi(\beta_i + \beta_j + z_{ij}^\top \gamma)\).

**Example 2.** (Infinite discrete weight) Let \(a_{ij} \in \{0, 1, \ldots\}\). We can model the edge weight using Poisson distribution with mean \(\lambda = \exp(\beta_i + \beta_j + Z_{ij}^\top \gamma)\):
\[
\log P(a_{ij} = a) = a(\beta_i + \beta_j + Z_{ij}^\top \gamma) - \exp(\beta_i + \beta_j + Z_{ij}^\top \gamma) - \log k!.
\]

To establish a unified theoretical result, we need to make a basic model assumption.

**Basic model assumption.** We assume that the degree parameters enter the marginal probability density function \(f(\cdot)\) additively through \(\beta_i + \beta_j\). Further, the additive structure also applies to the homophily component. That is, \(a_{ij}|z_{ij}, \beta, \gamma \sim f(a|z_{ij}^\top \gamma + \beta_i + \beta_j)\).

The dependence of the distribution \(f(\cdot)\) on the parameters is through an index \(z_{ij}^\top \gamma + \beta_i + \beta_j\) as given in the above example. This is referred to as single index models in
economic literature. We focus on these additive models for computational tractability. However, the model developed in this paper can be easily adapted into the non-additive structure for both effects.

3 Estimation

Write $\mu(\cdot)$ as the expectation on the distribution $f(\cdot)$. Since the dependence of the expectation of $a_{ij}$ on parameters is only through $\pi_{ij} = z_{ij}^\top \gamma + \beta_i + \beta_j$. We can write $\mu(z_{ij}^\top \gamma + \beta_i + \beta_j)$ as the expectation of $a_{ij}$. For convenience, we denote $\mu_{ij}(\beta, \gamma)$ by $\mu_{ij}(\beta, \gamma)$. To infer the parameters, we use the moment estimation. The moment equations are as follows:

$$
d_i = \sum_{j \neq i} \mu_{ij}(\beta, \gamma), \quad i = 1, \ldots, n,
$$

$$
\sum_{i=1}^{n} \sum_{j=1, j<i}^{n} a_{ij} z_{ij} = \sum_{i=1}^{n} \sum_{j=1, j<i}^{n} z_{ij} \mu_{ij}(\beta, \gamma).
$$

(2)

The solution to the above equations are the moment estimator denoted by $(\hat{\beta}, \hat{\gamma})$.

Now we discuss some computational issues. When the number of nodes $n$ is small and $f$ is the binomial, Probit, or Poisson probability function or Gamma density function, we can simply use the R function “glm” to solve (2). For relatively large $n$, it might be not enough large memory to store the design matrix needed for $\beta$. In this case, we recommend the use of a two-step iterative algorithm by alternating between solving the first equation in (2) via the fixed point method in Chatterjee et al. (2011) and solving the second equation in (2) via an iteratively reweighted least squares method for generalized linear models [McCullagh and Nelder (1989)].

4 Asymptotic properties

In this section, we establish the consistency and asymptotically normal distribution of the moment estimator. We first give some notations. For a subset $C \subset \mathbb{R}^n$, let $C^0$ and $\overline{C}$ denote the interior and closure of $C$, respectively. For a vector $x = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n$, denote by $\|x\|$ for a general norm on vectors with the special cases $\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$ and $\|x\|_1 = \sum_i |x_i|$ for the $\ell_\infty$- and $\ell_1$-norm of $x$ respectively. When $n$ is fixed, all norms on vectors are equivalent. Let $B(x, \epsilon) = \{ y : \|x - y\|_\infty \leq \epsilon \}$ be an $\epsilon$-neighborhood of $x$. For an $n \times n$ matrix $J = (J_{i,j})$, let $\|J\|_\infty$ denote the matrix norm induced by the $\ell_\infty$-norm on vectors in $\mathbb{R}^n$, i.e.

$$
\|J\|_\infty = \max_{x \neq 0} \frac{\|Jx\|_\infty}{\|x\|_\infty} = \max_{1 \leq i \leq n} \sum_{j=1}^{n} |J_{i,j}|.
$$
\[ \|J\| = \max_{i,j} |J_{ij}| \] be the matrix maximum norm. We use the superscript \(^{**}\) to denote the true parameter under which the data are generated. When causing no confusion, we omit the superscript \(^{**}\). We define

\[ \kappa_n := \max_{i,j} \|z_{ij}\|_\infty. \tag{3} \]

Throughout the paper, we assume that \(\mu(\cdot)\) is a continuous function with the third derivative. When causing no confusion, we will simply write \(\mu_{ij}\) stead of \(\mu_{ij}(\beta, \gamma)\) for shorthand.

Let \(\pi_{ij} = \beta_i + \beta_j + z_{ij}^\top \gamma\). Write \(\mu', \mu'', \mu'''\) as the first, second, third derivative of \(\mu\) on \(\pi\).

When \(\beta \in B(\beta^*, \epsilon_{n1}), \gamma \in B(\gamma^*, \epsilon_{n2})\), we assume that there are three numbers \(b_{n1}, b_{n2}, b_{n3}\) such that

\[ \max_{i,j} |\mu'(\pi_{ij})| \leq b_{n1}, \quad \max_{i,j} |\mu''(\pi_{ij})| \leq b_{n2}, \quad \max_{i,j} |\mu'''(\pi_{ij})| \leq b_{n3}. \tag{4} \]

Let \(\epsilon_{n1}\) and \(\epsilon_{n2}\) be two small positive numbers. Under the above inequality, the following holds:

\[ \max_{i,j} \sup_{\beta \in B(\beta^*, \epsilon_{n1})} |\mu_{ij}(\beta, \gamma^*) - \mu_{ij}(\beta^*, \gamma^*)| \leq 2b_{n1} \|\beta - \beta^*\|_\infty, \tag{5} \]

\[ \max_{i,j} \sup_{\gamma \in B(\gamma^*, \epsilon_{n1})} |\mu_{ij}(\beta^*, \gamma) - \mu_{ij}(\beta^*, \gamma^*)| \leq b_{n1} \kappa_n \|\gamma - \gamma^*\|_1. \tag{6} \]

The notation \(\sum_{j<i}\) is a shorthand for \(\sum_{i=1}^{n} \sum_{j=1, j<i}^{n}\).

### 4.1 Consistency

To deduce the conditions of the consistency of the moment estimator, let us first define a system of functions based on the moment equations. Define

\[ F_i(\beta, \gamma) = d_i - \sum_{j=1, j\neq i}^{n} \mu_{ij}(\beta, \gamma), \quad i = 1, \ldots, n, \tag{7} \]

and \(F(\beta, \gamma) = (F_1(\beta, \gamma), \ldots, F_n(\beta, \gamma))^\top\). Further, we define \(F_{i,\gamma}(\beta)\) as the value of \(F_i(\beta, \gamma)\) for an arbitrarily fixed \(\gamma\) when only evaluated at \(\beta\) and \(F_{\gamma}(\beta) = (F_{1,\gamma}(\beta), \ldots, F_{n,\gamma}(\beta))^\top\).

Let \(\hat{\beta}_{\gamma}\) be the solution to \(F_{\gamma}(\beta) = 0\). Correspondingly, we define two functions for exploring the asymptotic behaviors of the estimator of the homophily parameter:

\[ Q(\beta, \gamma) = \sum_{j<i} z_{ij}(a_{ij} - \mu_{ij}(\beta, \gamma)), \tag{8} \]

\[ Q_{c}(\gamma) = \sum_{j<i} z_{ij}(a_{ij} - \mu_{ij}(\hat{\beta}_{\gamma}, \gamma)). \tag{9} \]
$Q_c(\gamma)$ could be viewed as the concentrated or profile function of the moment function $Q(\beta, \gamma)$ in which the degree parameter $\beta$ is concentrated out. It is clear that
\[
F(\hat{\beta}, \hat{\gamma}) = 0, \quad F(\hat{\gamma}) = 0, \quad Q(\hat{\beta}, \hat{\gamma}) = 0, \quad Q_c(\hat{\gamma}) = 0.
\]

If the moment estimator $(\hat{\beta}, \hat{\gamma})$ is consistent, then it is natural to require that the norm $\|F(\beta, \gamma)\|_\infty$ evaluated at the true parameters $\beta^*$ and $\gamma^*$ is small. This leads to our first condition.

**Condition 1.** For $F(\beta, \gamma)$ defined at (7), we require that
\[
\|F(\beta^*, \gamma^*)\|_\infty = O_p(h_{n1}\sqrt{n \log n}),
\]
where $h_{n1}$ is a scalar factor that may depend on the ranges of $\beta^*$ and $\gamma^*$.

Condition 1 requires that $F_i(\beta^*, \gamma^*)$ is in the order of $(n \log n)^{1/2}$. It can be verified as follows. If the sequence $\{a_{ij}\}_{j=1}^n$ is independent for any fixed $i$, then condition 1 holds in the light of Hoeffding’s inequality for bounded random variables or concentration inequality for sub-exponential random variables [e.g., Corollary 5.17 in Vershynin (2012)]. If $\{a_{ij}\}_{j=1}^n$ is weekly/negatively/positively dependent, there also exist exponential inequalities for the sum of $\{a_{ij}\}_{j=1}^n$ [e.g., Delyon (2009); Roussas (1996); Ioannides and Roussas (1999)]. These probability inequalities depend on the values of parameters that leads to the appearance of the additional factor $h_{n1}$. More specifically, $h_{n1}$ depends on $\|\beta^*\|_\infty$ and $\|\gamma^*\|$. If the latter are bounded by a constant, then $h_{n1}$ is also a constant, regardless of $n$.

Let $F(x) : \mathbb{R}^n \to \mathbb{R}^n$ be a function vector on $x \in \mathbb{R}^n$. We say that a Jacobian matrix $F'(x)$ with $x \in \mathbb{R}^n$ is lipschitz continuous on a convex set $D \subset \mathbb{R}^n$ if for any $x, y \in D$, there exists a constant $\lambda > 0$ such that for any vector $v \in \mathbb{R}^n$ the inequality
\[
\|F'(x)(v) - F'(y)(v)\|_\infty \leq \lambda \|x - y\|_\infty \|v\|_\infty
\]
holds. Our third condition requires that $F_{\gamma^*}(\beta)$ is Lipschitz continuous on $D$ containing $\beta^*$.

**Condition 2.** Let $D \subset \mathbb{R}^n$ be an open convex set containing the true point $\beta^*$. The Jacobian matrix $F_{\gamma}'(x)$ of $F_{\gamma}$ on $x$ is Lipschitz continuous on $D$ with the Lipschitz coefficient $h_{n2}$. Here, $h_{n2}$ may depend on $n$ but not depend on $D$.

We use the Newton iterative sequence to show the consistency. The Lipschitz continuous property of $F_{\gamma}$ is one of conditions to guarantee its convergence. If the common distributions [e.g., the logistic, Possion, probit distributions] are used, then Condition 2 hold as given in Yan et al. (2016b).
The Jacobian matrix $F'_\gamma(\beta)$ of $F_\gamma(\beta)$ on the parameter $\beta$ has a special structure that can be characterized in the form of a matrix class. Given $m, M > 0$, we say an $n \times n$ matrix $V = (v_{ij})$ belongs to the matrix class $L_n(m, M)$ if $V$ is a diagonally balanced matrix with positive elements bounded by $m$ and $M$, i.e.,

$$v_{ii} = \sum_{j=1, j\neq i}^{n} v_{ij}, \quad i = 1, \ldots, n,$$

$$m \leq v_{ij} \leq M, \quad i, j = 1, \ldots, n; i \neq j.$$  \hspace{1cm} (11)

It is easily checked that $F'_\gamma(\beta)$ or $-F'_\gamma(\beta)$ belongs to this matrix class. This special structure for the Jacobian matrix $F'_\gamma(\beta)$ makes it possible to prove the consistency of the estimator $\hat{\beta}$ through obtaining the convergence rate of the Newton iterative sequence. This in turn needs to approximate the inverse of $F'_\gamma(\beta)$ since its inverse does not have a close form. Yan et al. (2015) proposed to approximate the inverse $V^{-1}$ of $V$ by a diagonal matrix $S = \text{diag}(1/v_{11}, \ldots, 1/v_{nn})$ and obtain the upper bound of the approximate error. To characterize the lower bound and upper bound, we list the following condition.

**Condition 3.** Let $\epsilon_n$ be a small number. When $\beta \in B(\beta^*, \epsilon_n)$, there exist two positive numbers $m_n$ and $M_n$ such that $F'_{\gamma^*}(\beta) \in L_n(m_n, M_n)$ or $-F'_{\gamma^*}(\beta) \in L_n(m_n, M_n)$, where $m_n$ and $M_n$ depend on the range of $\beta^*$ and the pairwise covariates $z_{ij}$'s.

By inequality (4), $M_n = b_n$. For clarification, we keep both notations. The following lemma characterizes the upper bound of the error between $\hat{\beta}_\gamma$ and $\beta^*$.

**Lemma 1.** Under Conditions 1, 2, 3, if $\gamma \in B(\gamma^*, (\log n/n)^{1/2})$ and

$$M_n^2 (h_{n1} + b_{n1} \kappa_n)(\log n)^{1/2} \times \max \left\{1, \frac{M_n^2 h_{n2}}{nm_n^3} \right\} = o(1),$$

then with probability approaching one, $\hat{\beta}_\gamma$ exists and satisfies

$$\|\hat{\beta}_\gamma - \beta^*\|_\infty = O_p \left(\frac{M_n^2 (h_{n1} + b_{n1} \kappa_n)}{m_n^3} \sqrt{\frac{\log n}{n}} \right) = o_p(1).$$

To show the consistency of $\hat{\gamma}$, similar to Conditions 1 and 2, we need the following two conditions.

**Condition 4.** $\|Q(\beta^*, \gamma^*)\| = O(h_{n3} n^{3/2} \log n)$, where $h_{n3}$ is a scalar factor.

**Condition 5.** $Q_c(\gamma)$ is the Lipschitz continuous with the Lipschitz coefficient $n^2 h_{n4}$.

The asymptotic behavior of $\hat{\gamma}$ crucially depends on the Jacobian matrix $Q'_c(\gamma)$. The
expression for the derivative of $Q_c(\gamma)$ on $\gamma$ is
\[
\frac{\partial Q_c(\gamma)}{\partial \gamma} = \frac{\partial Q(\hat{\beta}, \gamma)}{\partial \gamma} - \frac{\partial Q(\hat{\beta}, \gamma)}{\partial \beta} \left[ \frac{\partial F(\hat{\beta}, \gamma)}{\partial \beta} \right]^{-1} \frac{\partial F(\hat{\beta}, \gamma)}{\partial \gamma},
\]
where $\frac{\partial Q(\hat{\beta}, \gamma)}{\partial \gamma}$ denotes the partial derivative of $Q(\beta, \gamma)$ on $\gamma$ evaluated at $\beta = \hat{\beta}, \gamma = \gamma$. Since $\hat{\beta}$ does not have a close form, conditions are directly imposed on $Q'_c(\gamma)$ is not easily checked. To derive the feasible conditions, we define
\[
H(\beta, \gamma) = \frac{\partial Q(\beta, \gamma)}{\partial \gamma} - \frac{\partial Q(\beta, \gamma)}{\partial \beta} \left[ \frac{\partial F(\beta, \gamma)}{\partial \beta} \right]^{-1} \frac{\partial F(\beta, \gamma)}{\partial \gamma},
\]
which is a general form of $\frac{\partial Q_c(\gamma)}{\partial \gamma}$. Note that the dimension of $H(\beta, \gamma)$ is fixed. All matrix norms on $H(\beta, \gamma)$ are equivalent. The next condition bounds $\|Q_c(\gamma^*)\|$.

**Condition 6.** For $\beta \in B(\beta^*, \epsilon_n)$, it is required that $\|H^{-1}(\beta, \gamma^*)\| = O(h_{n5}/n^2)$, where $h_{n5}$ is a scalar factor.

Now we formally state the consistency result.

**Theorem 1.** Under Conditions 1–6, if equation (12) and the following hold:
\[
\left[ h_{n3} + \kappa_n b_{n1} \frac{M_n^2(h_{n1} + b_{n1} \kappa_n)}{m_n^3} \right] \sqrt{\frac{\log n}{n}} \times \max\{1, h_{n1} h_{n5}^2\} = o(1),
\]
then the moment estimator $\hat{\gamma}$ exists with probability approaching one and is consistent in the sense that
\[
\|\hat{\gamma} - \gamma^*\|_\infty = O_p \left( \left[ h_{n3} + \kappa_n b_{n2} \frac{M_n^2(h_{n1} + b_{n1})}{m_n^3} \right] \frac{h_{n5}(\log n)^{1/2}}{n^{1/2}} \right) = o_p(1),
\]
and
\[
\|\hat{\beta} - \beta^*\|_\infty = O_p \left( \frac{M_n^2(h_{n1} + b_{n1} \kappa_n)}{m_n^3} \sqrt{\frac{\log n}{n}} \right) = o_p(1).
\]

**4.2 Asymptotic representation**

Let $T_{ij}$ be a vector of length $n$ with $i$th and $j$th elements ones and other elements zeros and define
\[
s_{\beta,ij}(\beta, \gamma) = (a_{ij} - \mu_{ij}(\beta, \gamma))T_{ij}, \quad s_{\gamma,ij}(\beta, \gamma) = z_{ij}(a_{ij} - \mu_{ij}(\beta, \gamma)).
\]
Let
\[
V(\beta, \gamma) = \frac{\partial F(\beta, \gamma)}{\partial \beta}, \quad V_{\beta}(\beta, \gamma) = \frac{\partial Q(\beta, \gamma)}{\partial \beta^1},
\]
Then we define

\[ \tilde{s}_{\gamma ij}(\beta, \gamma) = s_{\gamma ij}(\beta, \gamma) - V_{\gamma\beta}(\beta, \gamma)[V(\beta, \gamma)]^{-1}s_{\beta ij}(\beta, \gamma). \]

When evaluating \( V(\beta, \gamma) \) at its true values, we omit the arguments \( \beta, \gamma \) in \( V \). Similarly, write \( V_{\gamma\beta} \) for \( V_{\gamma\beta}(\beta^*, \gamma^*) \) etc. That is, \( V = V(\beta^*, \gamma^*) \), \( V_{\gamma\beta} = V_{\gamma\beta}(\beta^*, \gamma^*) \), etc. Let \( N = n(n - 1) \) and

\[ \bar{H} = \lim_{n \to \infty} \frac{1}{N} H(\beta^*, \gamma^*). \]

where \( H(\beta, \gamma) \) is defined at (14). We assume that the above limiting exists. The asymptotic representation of \( \hat{\gamma} \) is stated below.

**Theorem 2.** Assume that the conditions in Theorem 1 holds. If

\[ \frac{b_{\gamma \beta} \kappa_n M_n^6 (b_{\gamma 1} + b_{\gamma 1} \kappa_n)^3 (\log n)^{3/2}}{n^{1/2} m_n^2} = o(1), \]

then we have

\[ \sqrt{N}(\hat{\gamma} - \gamma^*) = \bar{H}^{-1} B_* + \bar{H}^{-1} \frac{1}{\sqrt{N}} \sum_{j<i} \tilde{s}_{\gamma ij}(\beta^*, \gamma^*) + o_p(1), \]

where

\[ B_* = \lim_{n \to \infty} \frac{1}{2 \sqrt{N}} \sum_{i=1}^{n} \left( \sum_{j \neq i} \frac{z_{ij} \partial \mu_{ij}}{\partial \pi_{ij}} \right)^2. \]

**Remark 1.** The asymptotic expansion of \( \hat{\gamma} \) contains a bias term \( N^{-1/2} \bar{H} B_* \). If the parameter vector \( \beta \) and all homophily components \( z_i^T \gamma \)’s are bounded, then \( \| \bar{H} \| = O(1) \) and \( \| B_* \|_{inf} = O(1) \). It follows that \( \hat{\gamma} \) has a convergence rate at round \( n^{-1} \). Since \( \hat{\gamma} \) is not centered at the true parameter value, the confidence intervals and the p-values of hypothesis testing constructed from \( \hat{\gamma} \) cannot achieve the nominal level without bias-correction under the null: \( \gamma^* = 0 \). This is referred to as the so-called incidental parameter problem in econometric literature [Neyman and Scott (1948); Fernández-Vál and Weidner (2016); Dzemski (2017)]. The produced bias is due to the appearance of additional parameters. Here, we propose to use the analytical bias correction formula: \( \hat{\gamma}_{bc} = \hat{\gamma} - N^{1/2} H^{-1}(\hat{\beta}, \hat{\gamma}) \hat{B} \) where \( \hat{B} \) is the estimate of \( B_* \) by replacing \( \beta^* \) and \( \gamma^* \) in their expressions with their estimators \( \hat{\beta} \) and \( \hat{\gamma} \), respectively.

**Remark 2.** If \( N^{-1/2} \sum_{j<i} \tilde{s}_{\gamma ij}(\beta^*, \gamma^*) \) asymptotically follows a multivariate normal distribution, then \( \sqrt{N}(\hat{\gamma} - \gamma^*) \) converges in distribution to be normal.

**Theorem 3.** Let \( W = V^{-1} - S \) and \( B = \text{Cov}(d - E d) \). Assume that the conditions in
Theorem 1 holds. If
\[
\frac{M_n^2 (\varphi_{n1}^2 + \varphi_{n2}^2 \kappa_n^2) b_{n2} \log n}{m_n^3} = o(1),
\]
and
\[
\frac{M_n \kappa_n \|\hat{\gamma} - \gamma^*\|_1}{m_n^3} = o_p(n^{-1/2}), \quad \max_i (WBW^\top)_{ii} = o\left(\frac{1}{n}\right),
\]
then for any fixed \(i\),
\[
\hat{\beta}_i - \beta_i = v_i^{-1}(d_i - Ed_i) + o_p(n^{-1/2}),
\]
where
\[
\varphi_{n1} = \frac{M_n^2 (h_{n1} + b_{n1} \kappa_n)}{m_n^3}, \quad \varphi_{n2} = h_{n3} + \kappa_n b_{n2} h_{n5} \varphi_{n1}.
\]

Remark 3. We make a remark about the condition \(M_n^2 m_n^{-3} \kappa_n \|\hat{\gamma} - \gamma^*\|_1 = o(n^{-1/2})\). According to the asymptotic expansion of \(\hat{\gamma}\) in Theorem 2, if \(\frac{1}{\sqrt{N}} \sum_{j<i} \tilde{s}_{ij}(\beta^*, \gamma^*)\) converges in distribution to the normal distribution, then \(\|\hat{\gamma} - \gamma^*\|_\infty\) is in the magnitude of \(n^{-1}\) with probability approaching one. So this condition is mild and generally holds.

Remark 4. We discuss the condition \(\max_i (WBW^\top)_{ii} = o(n^{-1})\). If \(V = \text{Cov}(d - Ed)\), then \(WBW^\top = V^{-1} - S - S(I - VS)\), where \(I\) is the identify matrix of order \(n\). Similar to deriving equation (B.2) in Yan et al. (2016b), one can show that \(\|WVW^\top\| = O(n^{-2} M_n^2 m_n^{-3})\). In this case, \(\max_i (WBW^\top)_{ii} = o(n^{-1})\). If all random edges \(\{a_{ij}\}_{j<i}\) are independent and their distributions belong to the exponential family, then \(V = \text{Cov}(d - Ed)\).

Remark 5. If for any fixed \(k\), the vector \((d_1 - Ed_1, \ldots, d_k - Ed_k)\) is asymptotically multivariate normal distribution with mean 0 and covariance matrix \(\Sigma_{kk}\), then the vector
\[
(S_{kk}^{-1} \Sigma_{kk} S_{kk}^{-1})^{-1/2} (\hat{\beta}_1 - \beta_1, \ldots, \hat{\beta}_k - \beta_k)^\top
\]
converges in distribution to the standard normal distribution by Theorem 3, where \(S_{kk}\) is the the upper left \(k \times k\) submatrix of \(S\). In the case of edge independence that means \(\{a_{ij}\}_{j=1}^n\) is an independent random variable sequence for any fixed \(i\), \(v_i^{-1/2}(d_i - Ed_i)\) converges in distribution to the standard normality can be checked by various kinds of classical conditions such as Lyapunov’s condition [Billingsley (1995), page 362] and Lindeberg’s (1922) condition under which the central limit theorem holds. In the dependent case, it is complex to verify the central limit theorem for \(d_i\) and relevant examples can be find in Yan (2018).
5 Applications

In this section, we illustrate the theoretical result by an application to the logistic distribution $f(\cdot)$. Following Graham (2017), we assume that all dyads $(a_{ij}, a_{ji})$’s are independent. Under this assumption, the maximum likelihood equations are identical to the moment equations in (2). Thus, we relax the assumption that restricts the MLE to a compact set made by Graham (2017). The probit model with dyad dependent structures is given in an independent work [Yan (2018)] since verifying these conditions appears to complex.

The model is

$$
\mathbb{P}(a_{ij} = 1) = \frac{e^{z_{ij}^\top \gamma + \beta_i + \beta_j}}{1 + e^{z_{ij}^\top \gamma + \beta_i + \beta_j}}.
$$

The moment equations are

$$
d_i = \sum_{j \neq i} \frac{e^{z_{ij}^\top \gamma + \beta_i + \beta_j}}{1 + e^{z_{ij}^\top \gamma + \beta_i + \beta_j}} , \quad i = 1, \ldots, n,
$$

$$
\sum_{j < i} a_{ij} z_{ij} = \sum_{j < i} \frac{z_{ij} e^{z_{ij}^\top \gamma + \beta_i + \beta_j}}{1 + e^{z_{ij}^\top \gamma + \beta_i + \beta_j}}
$$

which is identical to the maximum likelihood equations. In the case of logistic regression, $\mu(x) = e^x / (1 + e^x)$. It can be shown that

$$
\mu'(x) = \frac{e^x}{(1 + e^x)^2}, \quad \mu''(x) = \frac{e^x(1 - e^x)}{(1 + e^x)^3}, \quad \mu'''(x) = \frac{e^x(1 - 4e^x + e^{2x})}{(1 + e^x)^4}.
$$

It is easily checked that

$$
|\mu'(x)| \leq \frac{1}{4}, \quad |\mu''(x)| \leq \frac{1}{4}, \quad |\mu'''(x)| \leq \frac{1}{4},
$$

where the last two inequalities are due to

$$
\mu''(x) \leq \frac{e^x}{(1 + e^x)^2} \times \frac{(1 - e^x)}{(1 + e^x)},
$$

and

$$
\mu'''(x) = \frac{e^x}{(1 + e^x)^2} \times \left[ \frac{1}{(1 + e^x)^3} - \frac{4e^x}{(1 + e^x)^2(1 + e^x)} + \frac{e^x}{(1 + e^x)^2(1 + e^x)} \right],
$$

respectively. So $b_{n1} = b_{n2} = b_{n3} = 1/4$ in inequality (4). By Hoeffding’s inequality, with the similar lines of arguments for proving Lemma 3 in Yan et al. (2016a), we have $h_{n1} = 1$. With the similar lines of arguments for condition 3 in Yan et al. (2016b), Condition 2 holds with $\lambda = 4n(-1)$. When $\beta \in B(\beta^*, \epsilon_{n1})$, $\gamma \in B(\gamma^*, \epsilon_{n2})$, $-\frac{\partial F(\beta, \gamma)}{\partial \beta} \in \mathcal{L}_n(m_n, M_n), \ldots$
where
\[ m_n = \frac{\epsilon^2\|\beta^\star\|_{\infty} + \|\gamma^\star\|_1 \kappa_n + 2\epsilon n_1 + \rho \epsilon n_2}{1 + \epsilon^2\|\beta^\star\|_{\infty} + \|\gamma^\star\|_1 \kappa_n + 2\epsilon n_1 + \rho \epsilon n_2}, \quad M_n = \frac{1}{4}. \]

This verifies Condition 3. Again, by Hoeffding’s inequality, we have
\[ \|Q(\beta^\star, \gamma^\star)\| = O_p(\kappa_n n (\log n)^{1/2}). \]

So \( h_n^3 = \kappa_n \) in Condition 4. It also can be shown that Condition 5 holds with \( h_n^4 = O(\kappa_n^2 M_n^6 n^2) \), whose proof is given in the supplementary material. Let \( \lambda_n \) be the smallest eigenvalue of \( \bar{H}(\beta^\star, \gamma^\star) \). Then Condition 6 holds with \( h_n^5 = \lambda_n \). So by Theorem 1, we have the following corollary.

**Corollary 1.** If
\[ \lambda_n^4 \kappa_n^4 e^{24\|\beta^\star\|_{\infty} + 12\epsilon n_1} \gamma^\star \|_\infty \sqrt{\frac{\log n}{n}} = o(1), \]

then \( \|\widehat{\gamma} - \gamma^\star\|_{\infty} = o_p(1) \) and \( \|\widehat{\beta} - \beta^\star\|_{\infty} = o_p(1) \).

Since \( a_{ij} \)’s (\( j < i \)) are independent, it is easy to show the central limit theorem for \( d_i \) and \( N^{-1/2} \sum_{j<i} \tilde{s}_{ij}(\beta, \gamma) \) as given in Su et al. (2018) and Graham (2017) respectively. So by Theorems 2 and 3, the central limit theorem holds for \( \widehat{\gamma} \) and \( \widehat{\beta} \). See also Su et al. (2018) and Graham (2017).

### 6 Summary and discussion

We have present the moment estimation for inferring the degree parameters and homophily parameter in model (1) that only specifies the marginal distribution. We establish the consistency of the moment estimator under several conditions and also derive its asymptotic representation. It is worth noting that the conditions imposed on \( m_n \) and \( M_n \) may not be best possible. In particular, the conditions guaranteeing the asymptotic normality seem stronger than those guaranteeing the consistency. We will investigate this in future studies and note that the asymptotic behavior of the MLE depends not only on \( m_n \) and \( M_n \), but also on the configuration of the parameters.

Throughout the paper, we assume that \( \max_{i,j} \|z_{ij}\|_{\infty} \leq \kappa_n \). Conditions imposed in the theorems imply that \( \kappa_n \) can be allowed to increase only with a slow rate. What can be said when some of \( \|z_{ij}\|_{\infty} \)’s are large? For example, some of the covariates information for edges may increases with a fast rate. If the proportion of large values of \( \|z_{ij}\|_{\infty} \)’s is bounded, then this will have little effect on the moment estimators when \( n \) is large, so that the consistency and asymptotic representation still hold. A more interesting case is when the proportion of large \( \|z_{ij}\|_{\infty} \)’s is not bounded. Is there any asymptotic properties
of the moment estimator? We plan to investigate this and other related situations in the future.

Acknowledgements

Yan is partially supported by by the National Natural Science Foundation of China (No. 11771171).

References

Billingsley, P. (1995). *Probability and measure. 3rd edition.* Wiley, New York.

Blitzstein, J. and Diaconis, P. (2011). A sequential importance sampling algorithm for generating random graphs with prescribed degrees. *Internet Mathematics, 6*(4):489–522.

Chatterjee, S., Diaconis, P., and Sly, A. (2011). Random graphs with a given degree sequence. *The Annals of Applied Probability,* pages 1400–1435.

Chung, F. and Lu, L. (2002). The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences, 99*(25):15879–15882.

Currranini, S., Jackson, M. O., and Pin, P. (2009). An economic model of friendship: homophily, minorities, and segregation. *Econometrica, 77*(4):1003–1045.

Delyon, B. (2009). Exponential inequalities for sums of weakly dependent variables. *Electronic Journal of Probability,* pages 752–779.

Dzemski, A. (2017). An empirical model of dyadic link formation in a network with unobserved heterogeneity. *Working Papers in Economics,* page No. 698.

Fernández-Vál, I. and Weidner, M. (2016). Individual and time effects in nonlinear panel models with large n, t. *Journal of Econometrics,* 192(1):291 – 312.

Fienberg, S. E. (2012). A brief history of statistical models for network analysis and open challenges. *Journal of Computational and Graphical Statistics,* 21(4):825–839.

Goldenberg, A., Zheng, A. X., Fienberg, S. E., and Airoldi, E. M. (2010). A survey of statistical network models. *Foundations and Trends in Machine Learning,* 2(2):129–233.

Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica,* 85(4):1033–1063.
Hillar, C. and Wibisono, A. (2013). Maximum entropy distributions on graphs. *arXiv preprint arXiv:1301.3321*.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30.

Hoff, P. D. (2005). Bilinear mixed-effects models for dyadic data. *Journal of the American Statistical Association*, 100(469):286–295.

Holland, P. W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, 76(373):33–50.

Ioannides, D. and Roussas, G. (1999). Exponential inequality for associated random variables. *Statistics & Probability Letters*, 42(4):423 – 431.

Jochmans, K. (2017). Semiparametric analysis of network formation. *Journal of Business & Economic Statistics*, page To appear.

Kolaczyk, E. D. (2009). *Statistical analysis of network data*. Springer-Verlag, New York.

Kossinets, G. and Watts, D. J. (2006). Empirical analysis of an evolving social network. *Science*, 311(5757):88–90.

Krivitsky, P. N., Handcock, M. S., Raftery, A. E., and Hoff, P. D. (2009). Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social networks*, 31(3):204213.

Lindeberg, J. W. (1922). Eine neue herleitung des exponentialgesetzes in der wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 15:211–225.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models, Second Edition*. Chapman and Hall.

McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444.

Neyman, J. and Scott, E. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, (16):1–32.

Park, J. and Newman, M. E. J. (2004). Statistical mechanics of networks. *Physical Review E*, 70(6):066117.

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

16
