Estimating a dynamic network formation model with degree heterogeneity: Application to international trade network

Yapeng Zheng* Xiao Wang† Bo Zhang‡
International Institute of Finance, School of Management
University of Science and Technology of China

June 29, 2023

Abstract

In order to estimate trade network dynamics among heterogeneous countries, we study a dynamic network model with unobserved heterogeneity. To circumvent the intractable problem in maximizing the dynamic logit type likelihood function with fixed effects, this model assumes a quadratic exponential distribution of connection probabilities across periods. A decomposition of quadratic exponential likelihood inspires a two-step estimation method: The first step requires a conditional maximum likelihood estimation and the problem is converted to estimating a static network with multiple connection statuses in the second step. The two-step estimators are consistent and asymptotically normal with the assumption of dense network and Monte Carlo simulation evidence confirms superior performance of estimators. Empirical work demonstrates that there is significant state dependence in global trade network and degree heterogeneity of countries is more dispersed with accelerated globalization process.

Keywords: β-model; Dynamic network model; International trade network; Quadratic exponential distribution.

*zypeng12@mail.ustc.edu.cn
†iriswx@ustc.edu.cn
‡wbchpmp@ustc.edu

1 Introduction

Global trade networks have gained increasing prominence given that the fast development of globalization (Graham & De Paula, 2020; Jackson, 2010). In the context of trade network, each node represents a distinct country and the links between nodes depict international trade relationships. There are two salient features of global trade networks—node heterogeneity and link state dependence. The first feature indicates that larger countries tend to possess a greater number of trade partners, while smaller countries may establish connections with only a limited subset of other countries. As for the second feature of link state dependence, it summarizes changes in the global trade network: While the process of globalization has accelerated since the 1980s and resulted in more countries becoming interconnected through trade links, uneven distribution of trade benefits and economic crises have occasionally incited resistance against globalization. Consequently, trade links between countries have the propensity to form, dissolve, and rebuild, causing global trade network to evolve over time. Nevertheless, to the best of our knowledge, there is no econometric model that can estimate the dynamic nature of trade network while accounting for node heterogeneity. Given increasing trade conflicts and the role of trade network in global economy, a dynamic network model is urgently needed.

To incorporate node-level heterogeneity and state dependence of trade network, we introduces a novel dynamic network model. This model builds upon the β-model framework that captures node heterogeneity in networks (Chatterjee et al., 2011; Graham, 2017; Yan & Xu, 2013; Yan et al., 2019). Specifically, we extend the static network model by assuming a quadratic exponential distribution to describe the dynamics of node connections, as proposed by Bartolucci and Nigro (2010). This distribution offers a flexible and suitable framework for modeling inter-temporal evolution of network connections. Notably, this model carries similar economic implications as a dynamic logit model while remaining within the exponential family, and thus enables us to derive sufficient statistics for time-invariant variables. Leveraging these sufficient statistics, we can decompose the likelihood of observed dynamic networks into two distinct components: the likelihood for the sufficient statistics and the conditional likelihood of the observed dynamic networks given these sufficient statistics. Moreover, the second part can be converted to the likelihood of a β-model with multiple statuses over periods (we call it a multiple β-model).

From this decomposition, we can identify the effects of dyad-level covariates, node-level fixed effects and state dependence via a two-step estimation method. In the first step, we estimate the effects of time-varying covariates and state dependence by conditional maximum likelihood (CML as in Andersen, 1970). After obtaining estimated parameters in the first step, we substitute them back to the likelihood function of sufficient statistics for fixed effects and time-invariant covariates. We adopt joint maximum likelihood (JML as in Graham, 2017) to estimate node-level heterogeneity and the effects of time-invariant covariates simultaneously. We also find that the multiple β-model has nearly identical properties with the β-model in the JML estimation. The two-step estimators are proven to be consistent and asymptotically normally distributed under the assumption of dense network (the ratio of realized over all possible links falls between 0 and 1), or a less stringent condition of nearly dense network. Monte Carlo simulations in supplementary material yield favorable results for both the JML estimation of the multiple β-model and the two-step estimation approach.

Based on three global trade network samples from the 1980s to the 2000s, we estimate state dependence and determinants of international trade in the dynamic network model with country-level heterogeneity. Our estimation results show strong state dependence in trade networks. Furthermore, the effects of covariates in the dynamic network model is qualitatively consistent with those in the baseline probit model. As a by-product, we obtain estimation of degree heterogeneity in three samples. With the accelerated globalization process, countries in the global trade network in the 2000s have a higher degree of heterogeneity, indicating that some countries may play as hubs with more trade links but others become peripheral relatively.

Our work contributes to four trends of literature. First, we contribute to the network perspective of international trade study. In the trend of literature on network formation, Allen (2014) and Chaney (2014) propose that information friction may affect the dynamics of international trade network and Chaney (2016) surveys the impact of social relationship on trade network formation. As Bernard and Moxnes (2018) point out that less attention has been paid to the trade network formation despite its importance, our model fills the gap by providing an econometric tool to investigate trade network heterogeneity and dynamics simultaneously. Another trend of literature views trade network as exogenous and focuses on spillover through trade network. Barrot and Sauvagnat (2016) and Bernard et al. (2019) analyze how exogenous shocks may transmit along seller-buyer networks. However, trade network is built endogenously and macroeconomic factors may also affect trade relationship. Our econometric model helps to control for the endogeneity problem in spillovers through trade network.

Second, our work contributes to the literature on the Erdős-Rényi type model (Erdős and Rényi, 1960; Gilbert, 1959) in analysing the probability of node connections and the literature on dynamic network models. We model dynamic link probability through quadratic exponential likelihood functions, which helps us develop a two-step estimation method. This quadratic exponential likelihood modelling method can be extended to other dynamic network models to simplify analysis. We also contribute to the literature on quadratic exponential distribution that can trace back to Cox (1958) and Cox (1972). While Bartolucci and Nigro (2010) and Bartolucci and Nigro (2012) employ the quadratic method to estimate a dynamic model for binary panel data, to our knowledge, we are the first to apply quadratic model to the dynamic network model. Furthermore, in the literature on dynamic network models, we can estimate network state dependence under the assumption of fixed periods, in contrast to Jiang et al. (2020) that estimates an autoregressive network model with the assumption that time periods increase to infinity. The most relevant literature to our work is Graham (2016), he focuses on identifying and estimating homophily and transitivity of the dynamic network formation model. His identification strategy is constructive and he estimates the structural parameter of the model by a type of conditional maximum likelihood estimation. Different with his method, our approach is more intuitive and can utilize all information in the likelihood function.

Third, our work contributes to the literature on the β-model by extending it to the dynamic network model. Chatterjee et al. (2011) proposes the β-model, provides estimating algorithms to degree heterogeneity and proves the consistency of maximum likelihood estimators (MLE). Following this pioneering work, Yan and Xu (2013) proves the central limit theorem of MLE; Graham (2017) introduces covariates into the β-model; Yan et al. (2019) extends the β-model to the directed network; Chen et al. (2021) studies the sparse β-model. We develop a multiple β-model to allow multiple connection statuses in history and apply it to the study of dynamic network.

Last but not the least, we contribute to the literature on the conditional maximum likelihood estimation and two-step estimation methods. Andersen (1970) proves the asymptotic properties of conditional maximum likelihood estimators. Both McFadden (1974) and Amemiya (1978) use a similar two-step estimation method with conditional likelihood estimation in the first step. We innovatively apply conditional maximum likelihood estimation and two-step estimation methods to the dynamic network model, broadening the application scope of these methods.

The remaining of this paper is organized as follows. In the next section, we propose our dynamic network model. We employ our method to estimate the global trade network in Section 3 and give some concluding remarks in Section 4. In the Supplementary Material, we illustrate performance of our method using Monte Carlo simulations.

2 Model

2.1 Baseline assumptions

Consider a large population of potentially connected agents with their link statuses evolving in period t, $t = 0, 1, \ldots, T$. Let $i = 1, \ldots, n$ index a random sample of agents from the population and the sample size n does not change over time. Each of the $N \overset{\text{def}}{=} \binom{n}{2} = n(n-1)/2$ pairs of sampled agents constitute a dyad.

For each dyad (i, j), let $A_{ij,t} = 1$ if i and j are connected in period t and zero otherwise. Both i and j will sign a contract and form a trade link if their potential profit is positive, which means that the network is undirected (i.e., $A_{ij,t} = A_{ji,t}$). As trade cannot be done by only one agent, self-ties are ruled out in the network (i.e., $A_{ii,t} = 0$). The $n \times n$ matrix A_t, with the (i, j)th element $A_{ij,t}$, is called as an adjacency matrix in period t. This matrix is binary and symmetric with zeros on its main diagonal. The adjacency matrix encodes the structure of links across all sampled agents. We will refer to a set of such links as, equivalently, a network or graph. We employ the bold case letters to denote the stack of matrices or scalars in different periods, i.e., $\mathbf{A} = (A_1, A_2, \ldots, A_T)'$ and $\mathbf{A}_{ij} = (A_{ij,1}, \ldots, A_{ij,T})'$. Furthermore, we implicitly assume that the network is dense in every period, namely the ratio of realized versus all possible links is between 0 and 1: $\sum_{i=1}^{n} \sum_{j \neq i} A_{ijt}/N \in (0,1)$ for $t = 0, 1, \ldots, T$.

An agent's *degree* is the number of links she has in each period t: $A_{i+,t} = \sum_{j \neq i} A_{ij,t}$ (where "+" denotes "leave-own-out" summation). The row sums of the adjacency matrix in each period, denoted by the $n \times 1$ vector $A_{+,t} = (A_{1+,t}, \ldots, A_{n+,t})'$, give the network's *degree sequence*. A dyad's *score* equals the total number of links it has from $t = 1$ to $t = T$: $A^s_{ij,+} = \sum_{t=1}^{T} A_{ij,t}$ (where"+" denotes summation over the periods). The $n \times n$ matrix A^s_+ with the ijth element $A^s_{ij,+}$, gives the network's *score matrix*. An agent's total score across potential links, or total degree across different periods, is $d_i \stackrel{\text{def}}{=} \sum_{t=1}^{T} A_{i+,t} = \sum_{j \neq i} A^s_{ij,+}$, the sequence of total scores is denoted as $\mathbf{d} = (d_1, d_2, \ldots, d_n)'$.

In each period, we also observe $\tilde{X}_{ij,t} = (W'_{ij}, X'_{ij,t})'$, a $(K_1 + K_2) \times 1$ dyad-level vector of link-dependent covariates, where $K_1 \times 1$ vector W_{ij} is time-invariant and $K_2 \times 1$ vector $X_{ij,t}$ is time-varying. All covariates are symmetric, i.e., $\tilde{X}_{ij,t} = \tilde{X}_{ji,t}$.[1] $\tilde{\mathbf{X}}_{ij} \stackrel{\text{def}}{=} \{\tilde{X}_{ij,t}; t = 1, \ldots, T\}$ represents the union of covariates of dyad (i, j) across periods. Similarly, $\tilde{\mathbf{X}}_t \stackrel{\text{def}}{=} \{\tilde{X}_{ij,t}; 1 \leq i \neq j \leq n\}$ is the union of covariates in period t across dyads and $\tilde{\mathbf{X}} \stackrel{\text{def}}{=} \{\tilde{X}_{ij,t}; 1 \leq i \neq j \leq n, t = 1 \ldots, T\}$ is the union of covariates across dyads and periods.

Next, we assume that the link between i and j is determined by a latent stochastic process $A^*_{ij,t}$:

$$\begin{aligned} A_{ij,t} =& \mathbf{1}\left(A^*_{ij,t} \geq 0\right) \\ A^*_{ij,t} =& \beta_i + \beta_j + W'_{ij}\lambda + X'_{ij,t}\gamma + A_{ij,t-1}\eta - \epsilon_{ij,t}, \ t = 1, 2, \ldots, T, \end{aligned} \tag{1}$$

where $\mathbf{1}(\cdot)$ denotes the indicator function. The latent process consists of five components. The first component is the unobserved agent-level attributes $\{\beta_i\}_{i=1}^{n}$ (*degree heterogeneity*).

[1] These covariates may be constructed by agent level attributes, see Graham (2017).

Degree heterogeneity vector is denoted as $\boldsymbol{\beta}_n = (\beta_1, \ldots, \beta_n)'$. The second time-invariant component $W'_{ij}\lambda$ and the third time-varying component $X'_{ij,t}\gamma$ are systematic components that vary with observed dyad attributes. The fourth component $A_{ij,t-1}\eta$ is state dependence and the last component $\epsilon_{ij,t}$ is idiosyncratic and symmetric, assumed independently and identically distributed (i.i.d.) across dyads and periods. Implicit in equation (1), $A^*_{ij,t}$ is transferable across directly linked agents.

Then we are ready to write the conditional probability for a link between i and j in period t and further the conditional likelihood function of the entire path $\mathbf{A}_{ij}$ over all periods. If we use superscript 0 to denote true values of parameters, the conditional probability for a link between i and j is

$$\begin{aligned}
&P(A_{ij,t} = 1 | \beta^0_i, \beta^0_j, \lambda^0, \gamma^0, \eta^0, \tilde{\mathbf{X}}_{ij}, A_{ij,t-1}) \\
=&\frac{\exp(\beta^0_i + \beta^0_j + W'_{ij}\lambda^0 + X'_{ij,t}\gamma^0 + A_{ij,t-1}\eta^0)}{1 + \exp(\beta^0_i + \beta^0_j + W'_{ij}\lambda^0 + X'_{ij,t}\gamma^0 + A_{ij,t-1}\eta^0)}, \ t = 1, 2, \ldots, T.
\end{aligned}$$

Given the initial status A_0, the likelihood function of entire path $\mathbf{A}_{ij}$ becomes

$$\begin{aligned}
&P(\mathbf{A}_{ij} | \beta^0_i, \beta^0_j, \lambda^0, \gamma^0, \eta^0, \tilde{\mathbf{X}}_{ij}, A_{ij,0}) \\
=&\frac{\exp\left(A^s_{ij,+}(\beta^0_i + \beta^0_j + W'_{ij}\lambda^0) + \sum_{t=1}^T A_{ij,t}X'_{ij,t}\gamma^0 + A_{ij*}\eta^0\right)}{\prod_{t=1}^T \left\{1 + \exp(\beta^0_i + \beta^0_j + W'_{ij}\lambda^0 + X'_{ij,t}\gamma^0 + A_{ij,t-1}\eta^0)\right\}},
\end{aligned} \tag{2}$$

where $A_{ij*} = \sum_{t=1}^T A_{ij,t-1}A_{ij,t}$. This is an extension of the autoregressive networks of Jiang et al. (2020). However, it is difficult to directly maximize (2) to obtain a MLE when $\eta \neq 0$. More intuitively, the conditional likelihood function (2) does not belong to the exponential family as its denominator is dependent on the observation $\mathbf{A}_{ij}$.

In order to bypass the intractable problem related to the state dependence in (2), we follow Bartolucci and Nigro (2012) to take first-order Taylor expansion of the conditional likelihood (2) at $(\boldsymbol{\beta}'_n, \lambda', \gamma', \eta)' = 0$. Bartolucci and Nigro (2012) proposes that the first-order Taylor expansion is a good approximation of the original likelihood function up to a correction term. Furthermore, this approximation is a quadratic exponential model which keeps all properties of the exponential family (Cox, 1972). We assume that the distribution of $\mathbf{A}_{ij}$ is

$$
\begin{aligned}
&P(\mathbf{A}_{ij}|\beta_i,\beta_j,\lambda,\gamma,\eta,\tilde{\mathbf{X}}_{ij},A_{ij,0})\\
=&\frac{\exp\left(A^s_{ij,+}(\beta_i+\beta_j+W'_{ij}\lambda)+\sum_t A_{ij,t}X'_{ij,t}\gamma+A_{ij,T}e^*(X_{ij,T})+A_{ij*}\eta\right)}{\sum_{\mathbf{z}}\exp\left(z_+(\beta_i+\beta_j+W'_{ij}\lambda)+\sum_t z_tX'_{ij,t}\gamma+z_Te^*(X_{ij,T})+z_{ij*}\eta\right)}\\
=&\frac{\exp\left(A^s_{ij,+}(\beta_i+\beta_j+W'_{ij}\lambda)\right)}{\mu_{ij}(\beta_i,\beta_j,\lambda,\boldsymbol{\theta})}\exp\left(\sum_t A_{ij,t}X'_{ij,t}\gamma+A_{ij,T}e^*(X_{ij,T})+A_{ij*}\eta\right), \qquad (3)
\end{aligned}
$$

with $\boldsymbol{\theta}=(\gamma',\eta)'$ and

$$
\mu_{ij}(\beta_i,\beta_j,\lambda,\boldsymbol{\theta})=\sum_{\mathbf{z}}\exp\left(z_+(\beta_i+\beta_j+W'_{ij}\lambda)+\sum_t z_tX'_{ij,t}\gamma+z_Te^*(X_{ij,T})+z_{ij*}\eta\right),
$$

where $z_{ij*}=A_{ij,0}z_1+\sum_{t>1}z_{t-1}z_t$ and the algebraic operator $\sum_{\mathbf{z}}$ indicates summation over all possible outcome vector $\mathbf{z}$.[2] And $e^*(X_{ij,T})=\psi+X'_{ij,T}\phi$, where ψ and ϕ are nuisance parameters (see detailed definition in Bartolucci and Nigro, 2010 and Bartolucci and Nigro, 2012).

From equation (3), we can obtain the distribution of score $A^s_{ij,+}$:

$$
\begin{aligned}
&P(A^s_{ij,+}|\beta_i,\beta_j,\lambda,\boldsymbol{\theta},\tilde{\mathbf{X}}_{ij},A_{ij,0})\\
=&\frac{\exp\left(A^s_{ij,+}(\beta_i+\beta_j+W'_{ij}\lambda)\right)}{\mu_{ij}(\beta_i,\beta_j,\lambda,\boldsymbol{\theta})}\sum_{\mathbf{z}(A^s_{ij,+})}\exp\left(\sum_t z_tX'_{ij,t}\gamma+z_Te^*(X_{ij,T})+z_{ij*}\eta\right), \qquad (4)
\end{aligned}
$$

where $\sum_{\mathbf{z}(A^s_{ij,+})}$ denotes the sum over all outcomes such that $\sum_{t=1}^T z_t=A^s_{ij,+}$. We are interested in the score distribution because it contains all information of parameters for time-invariant covariates, $\boldsymbol{\beta}_n$ and λ. Thus we can estimate $\boldsymbol{\beta}^0_n$ and λ^0 from distribution of $A^s_{ij,+}$.

[2]For example, when $T=2$, all outcomes of (z_1,z_2) are $(0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$ respectively. Typically, there are 2^T different forms for an outcome vector with length T. Similarly, $z_+=\sum_{t=1}^T z_t$ and $z_{ij*}=A_{ij,0}z_1+\sum_{t=1}^T z_{t-1}z_t$.

To see this more clearly, we re-write the distribution of $A^s_{ij,+}$ in (4) as

$$\begin{aligned} & P(A^s_{ij,+} = t|\beta_i, \beta_j, \lambda, \boldsymbol{\theta}, \tilde{\mathbf{X}}_{ij}, A_{ij,0}) \\ = & \frac{w_t \exp\left(t(\beta_i + \beta_j + W'_{ij}\lambda)\right)}{1 + \sum_{k=1}^{T} w_k \exp\left(k(\beta_i + \beta_j + W'_{ij}\lambda)\right)}, \ t = 0, 1, \ldots, T. \end{aligned} \tag{5}$$

Here $w_k = w(k; \boldsymbol{\theta}, \mathbf{X}_{ij})$ depends on time varying covariates $\mathbf{X}_{ij}$ and their coefficients $\boldsymbol{\theta}$, see its definition equation (S.1) in the Supplemental Material. We suppress the dependence in w_k for notation simplicity. One can verify the derivation details in the Supplemental Material. We can interpret equation (5) as an extension of the β-model in Chatterjee et al. (2011), where "β" refers to the degree heterogeneity of nodes. Each dyad in the β-model has two possible statuses: connected or not. In contrast, there are $T+1$ potential statuses for dyad (i, j) in the distribution of $A^s_{ij,+}$. To demonstrate the difference with the β-model, we define the *multiple β-model* as the formation of a static network with multiple statuses according to the rule (5). Now we convert the dynamic network model to a static network model with a sufficient statistic for its dynamic history. In the next subsection, we will explain how to estimate $\boldsymbol{\beta}_n$ and λ in the multiple β-model.

2.2 Estimation of the multiple β-Model

MLE for $\boldsymbol{\theta} = (\gamma', \eta)'$ is difficult to obtain because of the fixed effects in (3). As suggested by Bartolucci and Nigro (2010), a conditional maximum likelihood estimation for $\boldsymbol{\theta}$ is applicable under the quadratic exponential model (3). Therefore, we propose a two-step estimation method in this subsection.

We first give an inspiration of this method: By intuition, the sufficient statistic $A^s_{ij,+}$ summarizes the connecting probability of dyad (i, j) in every period. So $A^s_{ij,+}$ contains information of $(\boldsymbol{\beta}'_n, \lambda')'$, the coefficients of time invariant variables. Moreover, conditional on $A^s_{ij,+}$, the entire link history $\mathbf{A}_{ij}$ can contain additional information for network dynamics.

For example, link history $\mathbf{A}_{ij} = (0,0,1,1,1)'$ may imply a higher (positive) state dependence than $\mathbf{A}_{ij} = (1,0,1,0,1)'$, even through these two have identical $A^s_{ij,+}$. Based on this inspiration, we give a short description of our two-step estimation method. In the first step, we estimate coefficients of *time-varying* variables $\boldsymbol{\theta} = (\gamma', \eta)'$ by a conditional maximum likelihood estimation given $A^s_{ij,+}$; in the second step, after substituting estimated value of $\boldsymbol{\theta}$ into the likelihood function of $A^s_{ij,+}$, we employ a joint maximum likelihood approach to estimate degree heterogeneity vector $\boldsymbol{\beta}_n$ and the coefficient vector for *time-invariant* variables λ simultaneously. While the first step is just a direct application from Bartolucci and Nigro (2010), the second step requires us to derive theorems on the multiple β-model for the estimation. Thus we start our two-step estimation method by a discussion on the multiple β-model.

It is helpful to use a new notation M to represent the score matrix A^s_+ here because the estimation method in this sub-section is general, not just valid for the score matrix. All other notations are inherited. For dyad (i,j),

$$\begin{aligned} &P(M_{ij} = m|\beta_i, \beta_j, \lambda, \boldsymbol{\theta}, W_{ij}) \\ =&\frac{w_m \exp\left(m(\beta_i + \beta_j + W'_{ij}\lambda)\right)}{1 + \sum_{t=1}^{T} w_t \exp\left(t(\beta_i + \beta_j + W'_{ij}\lambda)\right)}, \ m = 0, 1, \ldots, T. \end{aligned} \tag{6}$$

We denote $P(M_{ij} = m|\beta_i, \beta_j, \lambda, \boldsymbol{\theta}, W_{ij})$ as $p_{m,ij}(\lambda, \beta_i, \beta_j)$ and define $p_{ij}(\lambda, \beta_i, \beta_j)$ as the vector of these items. We further write the probability under true values, $p_{m,ij}(\lambda^0, \beta^0_i, \beta^0_j)$, as $p_{m,ij}$ and the corresponding vector as p_{ij}.

Taking logarithm of equation (6) and summing over all dyads, we obtain joint log likelihood function for the score matrix M:

$$
\begin{aligned}
&l_n(\lambda, \boldsymbol{\beta}_n) = \sum_{i=1}^{n}\sum_{j>i} \log P(M_{ij}|\beta_i, \beta_j, \lambda, \boldsymbol{\theta}, W_{ij}) \\
&= \sum_{i=1}^{n}\sum_{j>i} \left[\log(w_{M_{ij}}) + M_{ij}(\beta_i+\beta_j) + M_{ij}W_{ij}'\lambda\right] - \sum_{i=1}^{n}\sum_{j>i} \log\left[1 + \sum_{k=1}^{T} w_k \exp\left(k(\beta_i+\beta_j+W_{ij}'\lambda)\right)\right] \\
&= \sum_{i=1}^{n}\beta_i d_i + \sum_{i=1}^{n}\sum_{j>i} M_{ij}W_{ij}'\lambda - \sum_{i=1}^{n}\sum_{j>i}\log\left[1+\sum_{k=1}^{T} w_k \exp\left(k(\beta_i+\beta_j+W_{ij}'\lambda)\right)\right] + C, \quad (7)
\end{aligned}
$$

where $C = \sum_{i=1}^{n}\sum_{j>i}\log(w_{M_{ij}})$ doesn't affect estimation and thus we will omit it below.

Recall that "JML" refers to joint maximum likelihood, the JML estimators $\hat{\lambda}_{\text{JML}}$ and $\hat{\boldsymbol{\beta}}_n$ are obtained by simultaneously maximizing (7). We can fix λ and find $\hat{\boldsymbol{\beta}}_n(\lambda)$ that maximizes $l_n(\lambda, \boldsymbol{\beta}_n)$ at first, i.e., $\hat{\boldsymbol{\beta}}_n(\lambda) = \arg\max_{\boldsymbol{\beta}_n \in \mathbb{R}^n} l_n(\lambda, \boldsymbol{\beta}_n)$. Then we obtain a concentrated (profile) log likelihood function about λ

$$
\begin{aligned}
&l_n^c(\lambda, \hat{\boldsymbol{\beta}}_n(\lambda)) \\
&= \sum_{i=1}^{n}\hat{\beta}_i(\lambda)d_i + \sum_{i=1}^{n}\sum_{j>i} M_{ij}W_{ij}'\lambda - \sum_{i=1}^{n}\sum_{j>i}\log\left[1+\sum_{k=1}^{T} w_k \exp(k(\hat{\beta}_i(\lambda)+\hat{\beta}_j(\lambda)+W_{ij}'\lambda)\right]
\end{aligned}
$$

where the superscript c denotes "concentrate". Note that $\hat{\lambda}_{\text{JML}}$ also maximizes $l_n^c(\lambda, \hat{\boldsymbol{\beta}}_n(\lambda))$.

By adapting Theorem 1.5 of Chatterjee et al. (2011), we can give a fixed point representation of $\hat{\boldsymbol{\beta}}_n(\lambda)$ when it exists. The first-order condition of equation (7) for β_i is

$$
d_i - \sum_{j\neq i} m_{ij}(\lambda, \boldsymbol{\beta}_n, \mathbf{W}_i) = 0, i = 1, \ldots, n
$$

where

$$
m_{ij}(\lambda, \boldsymbol{\beta}_n, \mathbf{W}_i) = \mathbb{E}M_{ij} = \frac{\sum_{k=1}^{T} k w_k \exp\left(k(\beta_i+\beta_j+W_{ij}'\lambda)\right)}{1+\sum_{k=1}^{T} w_k \exp\left(k(\beta_i+\beta_j+W_{ij}'\lambda)\right)},
$$

with $\mathbf{W}_i = (W_{i1}, \ldots, W_{i(i-1)}, W_{i(i+1)}, \ldots, W_{in})'$. Taking logarithm of the first-order condition yields for $i = 1, \ldots, n$,

$$
\begin{aligned}
\log(d_i) &= \log\left[\sum_{j\neq i} m_{ij}(\lambda, \boldsymbol{\beta}_n, \mathbf{W}_i)\right] \\
&= \log[\exp(T\beta_i) r_i(\lambda, \boldsymbol{\beta}_n, \mathbf{W}_i)] = T\beta_i + \log r_i(\lambda, \boldsymbol{\beta}_n, \mathbf{W}_i),
\end{aligned}
$$

with $r_i(\lambda, \boldsymbol{\beta}_n, \mathbf{W}_i) = \sum_{j \neq i} \exp(-T\beta_i) m_{ij}(\lambda, \boldsymbol{\beta}_n, \mathbf{W}_i)$. Rearranging terms, we know that $\hat{\boldsymbol{\beta}}_n$ is a solution of the fixed point equation:

$$\hat{\boldsymbol{\beta}}_n(\lambda) = \varphi(\hat{\boldsymbol{\beta}}_n(\lambda)), \tag{8}$$

where

$$\varphi(\boldsymbol{\beta}_n) = \frac{1}{T} \begin{pmatrix} \log d_1 - \log r_1(\lambda, \boldsymbol{\beta}_n, \mathbf{W}_1) \\ \dots \\ \log d_n - \log r_n(\lambda, \boldsymbol{\beta}_n, \mathbf{W}_n) \end{pmatrix}$$

Our adaptation has two key points. First, it adapts to the setting with covariates in the regression; second, different with Graham (2017), the normalization constant is T to ensure it behaves like a contraction mapping. As proved in the Supplemental Material, $\varphi(\varphi(\cdot))$ is a contraction mapping in a compact support of $\boldsymbol{\beta}_n$. Therefore the iteration process along equation (8) converges to a fixed point $\hat{\boldsymbol{\beta}}_n(\lambda)$ at an exponential rate. This fixed point representation helps us compute MLE of $\boldsymbol{\beta}_n$ efficiently as long as λ is given.

To study the large sample properties of JML estimators for the multiple β-model, we specify two assumptions on parameters below.

Assumption 1 (**Restrictions**).

(a) T is fixed and finite,

(b) for any dyad (i,j) and t, $0 < w^1 < w(t; \boldsymbol{\theta}, \mathbf{X}_{ij}) < w^2$ where w^1 and w^2 are positive constants.

Assumption 1 restricts the multiple β-model to a much smaller class in which all properties in the β-model still hold.[3] Some special cases of the multiple β-model are studied in Hillar and Wibisono (2018), where they called them entropy models on graph. However, these cases do not allow the weights to depend on covariates.

[3]See Yan and Xu (2013) and Graham (2017) for undirected case and Yan et al. (2019) for directed case

The second assumption is designed to identify the fixed effects and parameters of covariates jointly.

Assumption 2 (Joint fixed effects identification).

(a) $\lambda^0 \in \mathrm{int}(\mathbb{L})$, *with* $\mathbb{L}$ *a compact subset of* $\mathbb{R}^{K_1}$.

(b) The support of W_{ij} *is* $\mathbb{W}$, *a compact subset of* $\mathbb{R}^{K_1}$, $i = 1, \ldots, n; i < j \leq n$.

(c) The support of β_i^0 *is* $\mathbb{B}$, *a compact subset of* $\mathbb{R}$, $i = 1, \ldots, n$.

(d) $\mathbb{E}[l_n(\lambda, \boldsymbol{\beta}_n)|\mathbf{W}, \boldsymbol{\beta}_n^0]$ *is uniquely maximized at* $\lambda = \lambda^0$ *and* $\boldsymbol{\beta}_n = \boldsymbol{\beta}_n^0$ *for a large enough* n.

Part (a), (b) and (c) of Assumption 2 determine that for $m = 0, \ldots, T$, $p_{m,ij} \in (\kappa, 1-\kappa)$ uniformly for a positive constant $\kappa > 0$. This means that the limit network will not be sparse or a full graph. We need to point out that part (c) of this assumption might be relaxed to a condition that β_i^0 diverges with a slow rate such as $\sup_{1 \leq i \leq n} |\beta_i^0| = O(\log\log(n))$, see Yan and Xu (2013). Part (d) is a standard assumption in the literature (e.g., Section 6 of Newey and McFadden, 1994) that ensures the consistency of JML estimators. To identify n fixed effects of agents, Assumption 2 requires a network with $\Omega(n^2)$ links. Although rigorous, we need to admit that this may be a proper condition, see the discussion in next section.

Now we are ready to articulate four theorems on $\hat{\boldsymbol{\beta}}(\lambda)$ and $\hat{\lambda}_{\mathrm{JML}}$. The first theorem is an extension of Theorem 1.5 of Chatterjee et al. (2011), also see Graham (2017). For simplicity, we write $\boldsymbol{\beta}$ for $\boldsymbol{\beta}_n$ if we also need to specify superscripts of $\boldsymbol{\beta}_n$; $\boldsymbol{\beta}^k$ represents the kth iterative value when finding the fixed point $\hat{\boldsymbol{\beta}}_n(\lambda)$. The first inequality in Theorem 1 ensures that the joint MLE for $\boldsymbol{\beta}(\lambda)$ will exist and the second inequality ensures that $\boldsymbol{\beta}^k(\lambda)$ will converge to $\boldsymbol{\beta}_n(\lambda)$, the fixed point of equation (8).

Theorem 1. *Suppose the concentrated MLE $\hat{\boldsymbol{\beta}}(\lambda)$ lies in the interior of $\mathbb{B}^n$, then for some δ such that $\boldsymbol{\beta}^{k+1}(\lambda) = \varphi(\boldsymbol{\beta}^k(\lambda))$,*

$$\|\boldsymbol{\beta}^{k+2}(\lambda) - \boldsymbol{\beta}^{k+1}(\lambda)\|_\infty \leq \left(1 - \frac{2(n-2)}{n-1}\delta^2\right) \|\boldsymbol{\beta}^k(\lambda) - \boldsymbol{\beta}^{k-1}(\lambda)\|_\infty$$

and

$$\|\boldsymbol{\beta}^{k+1}(\lambda) - \hat{\boldsymbol{\beta}}(\lambda)\|_\infty \leq \left(1 - \frac{2(n-2)}{n-1}\delta^2\right) \|\boldsymbol{\beta}^{k-1}(\lambda) - \hat{\boldsymbol{\beta}}(\lambda)\|_\infty.$$

Under Assumptions 1 and 2, we have that JML estimation of the multiple β-model is consistent and asymptotically normal and summarize the results in Theorem 2, 3 and 4 below.

Theorem 2. *Under Assumptions 1 and 2*

$$\hat{\lambda}_{\text{JML}} \xrightarrow{p} \lambda^0.$$

The intuition for Theorem 2 is that we can decompose the log likelihood function as:

$$\begin{aligned} l_n(\lambda, \boldsymbol{\beta}_n) = & \sum_{i=1}^{n} \sum_{j>i} \sum_{k=1}^{T} (M_{k,ij} - p_{k,ij}) \ln\left(\frac{p_{k,ij}(\lambda, \beta_i, \beta_j)}{p_{0,ij}(\lambda, \beta_i, \beta_j)}\right) \\ & - \sum_{i=1}^{n} \sum_{j>i} D_{\text{KL}}\left(p_{ij} || p_{ij}(\lambda, \beta_i, \beta_j)\right) - \sum_{i=1}^{n} \sum_{j>i} \mathbf{S}(p_{ij}), \end{aligned} \tag{9}$$

where $M_{k,ij} = 1\{M_{ij} = k\}$, $k = 0, 1, \ldots, T$. Also recall that $p_{ij} = \left(p_{0,ij}(\lambda^0, \beta_i^0, \beta_j^0), \ldots, p_{T,ij}(\lambda^0, \beta_i^0, \beta_j^0)\right)'$ is $(T+1) \times 1$ distribution vector under true values. The first item in equation (9) measures the difference between sample M_{ij} and expectation p_{ij}. The function $D_{\text{KL}}(\cdot||\cdot)$ is the Kullback-Leibler divergence and measures how the distribution under $(\boldsymbol{\beta}_n', \lambda')'$ is different with the distribution under the true values $(\boldsymbol{\beta}_n^{0\prime}, \lambda^{0\prime})'$. And $\mathbf{S}(\cdot)$ is the entropy function which measures the amount of information in random variable M_{ij}. Notice that $(M_{k,ij} - p_{k,ij}) \ln\left(\frac{p_{k,ij}(\lambda, \beta_i, \beta_j)}{p_{0,ij}(\lambda, \beta_i, \beta_j)}\right)$ is a bounded random variable by applying Hoeffding's inequality (see, e.g., Theorem 2.8 of Boucheron et al., 2013), we can show that the first component of $l_n(\lambda, \boldsymbol{\beta}_n)$ is $o_p(n^2)$ in probability (so that it will be $o_p(1)$ after normalizing by n^2). Then the maximizer of equation (9) is approximately the minimizer of the sum of $\binom{n}{2}$ Kullback-Leibler measures of divergence of $p_{ij}(\lambda, \beta_i, \beta_j)$ from p_{ij}. Actually, Kullback-Leibler measure is minimized when divergence is 0, i.e., when $\hat{\beta}_i = \beta_i^0$ and $\hat{\lambda} = \lambda^0$. By part (d) of Assumption 2, this minimizer will be unique and consistent as $n \to \infty$.

The next theorem states that under λ^0, the maximum likelihood estimation of $\beta_i, i = 1, \ldots, n$ will be consistent, too. Actually, this theorem gives an upper bound of the estimation error of $\hat{\beta}_i(\lambda^0)$.

Theorem 3. *Under Assumption 1 and 2,*

$$P\left(\sup_{1\le i\le n} |\hat{\beta}_i(\lambda^0) - \beta_i^0| \le C\sqrt{\frac{\ln n}{n}}\right) = 1 - O(n^{-2}),$$

where C is a uniform constant.

Theorem 3 ensures that $\hat{\beta}_i$ will not have a bias larger than $O(\sqrt{\ln n/n})$ with probability $1 - O(n^{-2})$. This loose upper bound can be explained by an observation: the estimation of λ utilizes information contained in $N = \binom{n}{2}$ dyads while the estimation of β_i only uses information from $n-1$ agents that might be linked with agent i. To give an analytic form for the bias of $\hat{\lambda}_{\text{JML}}$, we need to compute the inverse of Hessian matrix for $\boldsymbol{\beta}_n$ ($H^{-1}_{n,\boldsymbol{\beta}\boldsymbol{\beta}}$, see the detailed definition in the Supplemental Material). However, it is difficult to directly compute the inverse because the dimension of $H_{n,\boldsymbol{\beta}\boldsymbol{\beta}}$ is $n \times n$, which will be large when the number of agents in the sample increases. Thanks to the special structure of $H_{n,\boldsymbol{\beta}\boldsymbol{\beta}}$, we follow Yan and Xu (2013) and Graham (2017) to find a good approximation of $H^{-1}_{n,\boldsymbol{\beta}\boldsymbol{\beta}}$ and define the approximation matrix as Q_n. Based on Q_n, we give an asymptotically linear representation of $\hat{\lambda}_{\text{JML}}$ which characterizes the bias term.

In order to introduce Theorem 4, we define

$$\mathcal{I}_n(\lambda) = -\binom{n}{2}^{-1} \frac{\partial^2 l_n^c(\lambda^0, \hat{\boldsymbol{\beta}}(\lambda^0))}{\partial \lambda \partial \lambda'} \tag{10}$$

and

$$B_0 = -\lim_{n\to\infty} \frac{1}{2\sqrt{N}} \sum_{i=1}^{n} \frac{\sum_{j\neq i} h_{ij} W_{ij}}{\sum_{j\neq i} v_{ij}} \tag{11}$$

with $v_{ij} = \mathrm{Var}_0(M_{ij})$ and $h_{ij} = \mathbb{E}_0(M_{ij} - m_{ij})^3$. The subscript 0 above indicates these expressions are under true parameters λ^0 and $\boldsymbol{\beta}_n^0$. We also define the limit of $\mathcal{I}_n(\lambda)$ as $\mathcal{I}_0(\lambda)$.

Theorem 4. *Under Assumptions 1 and 2, as $n \to \infty$, the $K_1 \times 1$ vector $\sqrt{N}(\hat{\lambda}_{\mathrm{JML}} - \lambda^0)$ is asymptotically multivariate normal distributed with mean $\mathcal{I}_0(\lambda)^{-1} B_0$ and covariance matrix $\mathcal{I}_0^{-1}(\lambda)$.*

The magnitude of bias term is $O(n^{-1})$, which may be omitted in our empirical application. But in some cases when n is not large enough, we should correct the estimation bias. Bias corrected estimator is computed directly by $\hat{\lambda}_{\mathrm{BC}} = \hat{\lambda}_{\mathrm{JML}} - \mathcal{I}_0(\hat{\lambda}_{\mathrm{JML}})^{-1} B_0/\sqrt{N}$ like the method in Yan et al. (2019).

2.3 Two-step estimation

We have discussed how to estimate fixed effects and coefficients of time invariant covariates in the multiple β-model with a given weight functions w_t. In this subsection, we apply a conditional maximum likelihood approach to estimate coefficients of time-varying covariates γ and state dependence η. Combining (3) and (5), we obtain the distribution of $\mathbf{A}_{ij}$ given $A_{ij,+}^s$:

$$\begin{aligned} &P(\mathbf{A}_{ij} | \beta_i, \beta_j, \lambda, \boldsymbol{\theta}, \mathbf{X}_{ij}, A_{ij,+}^s, A_{ij,0}) \\ =&\frac{\exp\left(\sum_t A_{ij,t} X_{ij,t}\gamma + A_{ij,T} e^*(X_{ij,T}) + A_{ij*}\eta\right)}{\sum_{\mathbf{z}(A_{ij,+}^s)} \exp\left(\sum_t z_t X_{ij,t}\gamma + z_T e^*(X_{ij,T}) + z_{ij*}\eta\right)}. \end{aligned} \tag{12}$$

As this likelihood function is independent on $\boldsymbol{\beta}_n$ and λ, we can denote it as $P(\mathbf{A}_{ij}|\mathbf{X}_{ij}, A^s_{ij,+}, A_{ij,0})$. For an observed sample $(\mathbf{X}_{ij}, \mathbf{A}_{ij}, A_{ij,0}), i, j = 1, \ldots, n, i \neq j$, the conditional likelihood has logarithm

$$l_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{j>i} 1\{0 < A^s_{ij,+} < T\} \log \left[P(\mathbf{A}_{ij}|\mathbf{X}_{ij}, A^s_{ij,+}, A_{ij,0}) \right]. \tag{13}$$

We remove items with $A^s_{ij,+} = 0$, T because those items do not contain information of η (those items all equal 0). The number of non-zero terms in equation (13) is defined as the true sample size. We can estimate $\boldsymbol{\theta}$ by directly maximizing the conditional log likelihood function (13).

To formally describe the two-step procedure, we decompose the total likelihood function of $\mathbf{A}$ as:

$$\begin{aligned}
&P(\mathbf{A}|\boldsymbol{\beta}_n, \lambda, \boldsymbol{\theta}, \tilde{\mathbf{X}}, A_0) \\
=&\prod_{i=1}^{n} \prod_{j>i} P(\mathbf{A}_{ij}|\beta_i, \beta_j, \lambda, \boldsymbol{\theta}, \mathbf{X}_{ij}, A^s_{ij,+}, A_{ij,0}) P(A^s_{ij,+}|\beta_i, \beta_j, \lambda, \boldsymbol{\theta}, W_{ij}, A_{ij,0}) \\
=&\prod_{i=1}^{n} \prod_{j>i} P(\mathbf{A}_{ij}|\boldsymbol{\theta}, \mathbf{X}_{ij,t}, A^s_{ij,+}, A_{ij,0}) P(A^s_{ij,+}|\beta_i, \beta_j, \lambda, \boldsymbol{\theta}, W_{ij}, A_{ij,0}),
\end{aligned} \tag{14}$$

after taking logarithm and suppressing the notation of given covariates, we obtain

$$L_n(\lambda, \boldsymbol{\beta}_n, \boldsymbol{\theta}|\mathbf{A}) \stackrel{\text{def}}{\equiv} \log \left[P(\mathbf{A}|\boldsymbol{\beta}_n, \boldsymbol{\theta}, \tilde{\mathbf{X}}, A_0) \right] = l_n(\boldsymbol{\theta}) + l_n(\lambda, \boldsymbol{\beta}_n, \boldsymbol{\theta}),$$

where $l_n(\boldsymbol{\theta})$ is defined by equation (13) and $l_n(\lambda, \boldsymbol{\beta}_n, \boldsymbol{\theta})$ comes from equation (7) with an additional $\boldsymbol{\theta}$ to indicate the dependence on $\boldsymbol{\theta}$. Thus the estimation can be done by a two-step maximization procedure as in the definition below.

Definition 1 (Two-step estimation procedure)**.** *We estimate* $\boldsymbol{\theta} = (\gamma', \eta)'$ *and* $(\boldsymbol{\beta}'_n, \lambda')'$ *by following two-step procedure:*

$$\hat{\boldsymbol{\theta}} := \arg\max_{\boldsymbol{\theta}} l_n(\boldsymbol{\theta}), \tag{15}$$

$$(\hat{\boldsymbol{\beta}}_n', \hat{\lambda}_{\text{TS}}')' := \arg\max_{\boldsymbol{\beta}_n,\lambda} l_n(\lambda, \boldsymbol{\beta}_n, \hat{\boldsymbol{\theta}}), \tag{16}$$

where the subscripts "TS" indicate that the estimators are from the two-step estimation.

It is clear that the second step will be done by estimating the multiple β-model with the estimated parameters $\hat{\boldsymbol{\theta}}$ from the first step. So that we can also write $\hat{\lambda}_{\text{TS}} := \hat{\lambda}_{\text{TS}}(\hat{\boldsymbol{\theta}})$ to demonstrate this dependence on $\hat{\boldsymbol{\theta}}$. This method with the first step as a conditional likelihood estimation resembles the two-step estimation method of a multivariate logit model in McFadden (1974) and Amemiya (1978).

We introduce two more assumptions before stating main results on the two-step estimation.

Assumption 3. *The number of non-zero items in (13) is $s(n)$,*

(a) $\lim_{n\to\infty} s(n) = \infty$,

(b) or a stronger condition: $s(n) = \Omega(n^2)$.

This assumption ensures that the "true" sample size used for estimating $\boldsymbol{\theta}$ is large enough to guarantee consistency and asymptotic normality. Similar with the case of estimation of λ is influenced by the error of $\boldsymbol{\beta}_n$, the estimation error of the first step will cause a finite sample bias in second-step estimation. And this bias is in proportion to $O(s(n)^{-1})$, see Section 4 of Hahn and Newey (2004). What's more, $s(n)$ also influences the asymptotic variance of the second step estimators because $s(n)$ is the number of dyads that have link changing history. We need to point out that if Assumption 2 is satisfied, $s(n) = \Omega(n^2)$ will be ensured with probability 1. However, if $s(n) = O(n)$ or even diverges at a slower rate (i.e., a very sparse network), we can ensure the asymptotic properties in the first step estimation only but not in the second step. So when Assumption 2 is satisfied, two-step estimation procedure is well behaved. When Assumption 2 fails but part (a) of Assumption

3 holds, we can only employ the first stage estimation, in other words, only identify the time-variant parameters and state dependence.

Assumption 4 (Two-step estimation identification).

(a) The k-th element of $\boldsymbol{\theta}$ satisfies $\theta_k \in \operatorname{int}(\Theta)$, with Θ a compact subset of $\mathbb{R}$, $k = 1, \ldots, K_2$.

(b) The support of $X_{ij,t}$ is $\mathbb{X}$, a compact subset of $\mathbb{R}^{K_2}$ for $t = 1, \ldots, T$ and all dyads (i, j).

(c) $\mathbb{E}[l_n(\boldsymbol{\theta})|\mathbf{X}, \boldsymbol{\theta}^0]$ is uniquely maximized at $\boldsymbol{\theta} = \boldsymbol{\theta}^0$ for a large enough n.

Part (a) and (b) of Assumption 4 ensure that w_t is positive and finite for all t. Part (c) of Assumption 4 is standard to identify $\boldsymbol{\theta}^0$. Under Assumption 1, 2, 3 and 4, we can construct the consistency and asymptotic normality in Theorems 5 and 6.

Theorem 5. *Under Assumptions 1, 2, part (b) of 3 and 4, as $n \to \infty$, the first step estimator $\hat{\boldsymbol{\theta}}$ and the second-step estimator $\hat{\lambda}$ are all consistent,*

$$\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}^0,$$

$$\hat{\lambda}_{\text{TS}} \xrightarrow{p} \lambda^0.$$

Theorem 6. *Under Assumptions 1, 2, part (b) of 3 and 4, with $n \to \infty$,*

1. *$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \xrightarrow{d} N(0, \mathbf{J}^{-1})$, with $\mathbf{J}$ will be defined in the Supplemental Material, see equation* (S.20)*.*

2. *$K_1 \times 1$ vector $\sqrt{N}(\hat{\lambda}_{\text{TS}} - \lambda^0)$ is asymptotically multivariate normal distributed with mean $\mathcal{I}_0(\lambda)^{-1}B_0$ and covariance matrix Σ, which will be defined in the Supplemental Material, see equation* (S.30)*.*

Recall Definition 1, the intuition of Theorem 6 is that $\hat{\lambda}_{\mathrm{TS}}$ is also the JML estimation of λ under $\hat{\boldsymbol{\theta}}$, so that we can denote $\hat{\lambda}_{\mathrm{TS}}$ as $\hat{\lambda}_{\mathrm{JML}}(\hat{\boldsymbol{\theta}})$. Then we can rewrite $\hat{\lambda}_{\mathrm{TS}} - \lambda^0$ as

$$\begin{aligned}\hat{\lambda}_{\mathrm{TS}} - \lambda^0 &= \hat{\lambda}_{\mathrm{TS}}(\hat{\boldsymbol{\theta}}) - \lambda^0 \\ &= \underbrace{[\hat{\lambda}_{\mathrm{JML}}(\hat{\boldsymbol{\theta}}) - \hat{\lambda}_{\mathrm{JML}}(\boldsymbol{\theta}^0)]}_{\text{First-step error}} + \underbrace{[\hat{\lambda}_{\mathrm{JML}}(\boldsymbol{\theta}^0) - \lambda^0]}_{\text{Second-step error}},\end{aligned}$$

where the first term measures the bias arisen from the estimation error of $\hat{\boldsymbol{\theta}}$ and the second term gives the error that results from the JML estimation method. This decomposition helps us recognize the source of estimation error: as Theorem 4 specifies the asymptotic bias of the second term, we just need to focus on the first term which has an analytically asymptotic expression by the Taylor expansion at $\boldsymbol{\theta}^0$, see the Supplementary Material.

In the Supplementary Material, Monte Carlo simulation results show that the two-step method behaves well. We examine whether our estimators are consistent and whether their coverage rate of true values admits the asymptotic normality. Our estimators are superior across all experiments.

3 Empirical application: international trade network

3.1 Data and empirical framework

Our real world examples are three global trade network samples in different periods. The first one is between 1980 and 1989 in 158 countries/areas extracted from Helpman et al. (2008); the next two are trade networks among the 1990s and the 2000s in 125 countries/areas extracted from Barbieri and Keshk (2016) and Barbieri et al. (2009).[4] We

[4]The sample from Helpman et al. (2008) cannot be easily extended by adding trade data in later years because some countries disintegrated and others established in the early 1990s. Thus we supplemented the data from Barbieri and Keshk (2016) and Barbieri et al. (2009) and 125 countries are the maximum of

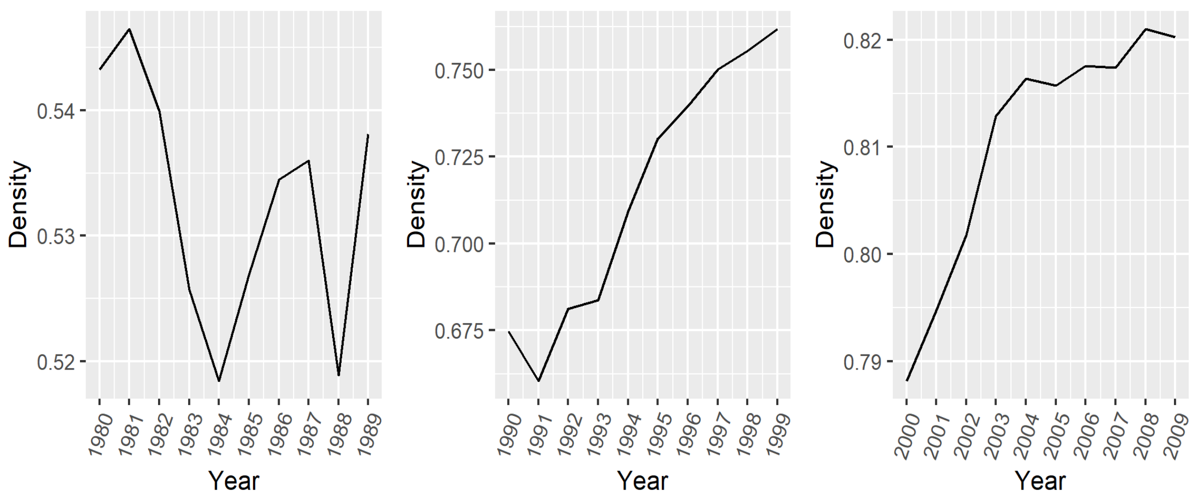


Figure 1: Network density series in three decades

construct these samples based on two reasons: First, we can estimate the sample in the 1980s by our two-step estimation method and directly compare the results with Helpman et al. (2008) to examine its performance. Second, as international trade network evolves with a dramatic change in last decades, estimating and comparing results in three samples help to exloit this change. Especially in the 1990s and the 2000s, as more countries joined the World Trade Organization (WTO), trade network became denser. We also estimate a sample between 2012 and 2021 in 131 countries/areas collected from UNcomtrade. However, this network is almost completely connected everywhere with a density of 0.98, which makes any estimation method meaningless (the result is presented in the Supplemental Material).[5]

Network in the first year of each sample is set as the initial status A_0. We implement our estimation method in three samples to study network dynamics and degree heterogeneity

intersection from two samples.

[5]The trade network at the sector or firm level between 2012 and 2021 is sparse and we plan to investigate the inference method for sparse network in a subsequent paper. More discussions are in the section of conclusion.

of different countries/areas. Specifically, we assign $A_{ij,t} = 1$ if country pair (i, j) in period t trade with each other in terms of either import or export given the nature of non-directional network in our estimation method.

Helpman et al. (2008) employs a micro-founded model to prove that if their profit from international trade is non-negative, two countries trade with each other; trade determinants are related to productivity and cost of two countries. Our dynamic network model in (1) exactly describes such a decision: the profit from international trade $A^*_{ij,t}$ between countries i and j depends on decree heterogeneity β_i and β_j, time-invariant and time-varying determinants W_{ij}, $X_{ij,t}$, and prior trade status $A_{ij,t-1}$. Main time-invariant and time-varying covariates are consistent with those in Helpman et al. (2008). Specifically, trade determinants include (i) geographic factors such as distance between two countries, whether countries share land border, are islands or landlocked; (ii) institutional factors such as whether countries have common legal system, language, religion, or colonial ties; (iii) economic factors such as whether two countries are in the same currency union or free trade agreement, or WTO members.[6]

3.2 Degree distribution and descriptive statistics

In order to examine whether the trade network is dense, we define the normalized degree of each country i in period t as

$$\tilde{d}_{i,t} \overset{\text{def}}{=} \frac{\sum_{j \neq i} A_{ij,t}}{n-1}$$

[6] We find that two dummy variables $Currency\ Union$ and FTA have almost no time variation for each country. Thus we convert $Currency\ Union$ and FTA to time invariant, following the rule that the dummy is one if the country pair is in the same currency union or FTA in at least one year through the time span.

and measure it in three samples. Because

$$\frac{\sum_{i=1}^{n} \tilde{d}_{i,t}}{n} = \frac{2\sum_{i=1}^{n}\sum_{j>i} A_{ij,t}}{n(n-1)},$$

the mean of $\tilde{d}_{i,t}$ across countries in period t is the density of trade network. In Figure 1, we plot time series of network density in three samples. While it keeps relatively stable in the 1980s, network density increases significantly in the 1990s and the 2000s. This salient change demonstrates that the global trade network is dynamic. We further report the minimum, three quantiles, and the maximum of normalized degree distribution for three samples in Tables 1, 2, and 3. These detailed statistics show that trade networks satisfy our dense network assumption and confirm the validity of our estimation method in these samples.

Moreover, we present descriptive statistics for trade data in Table 4. Geographic and institution determinants for trade are relative stable across three samples. Changes in *WTO none* and *WTO all* indicate that more countries becomes WTO members. Variation inflation factors of covariates in three periods are less than 2, which excludes the possibility of multicollinearity, see Table S.7 in Supplementary Materials.

Table 1: Normalized degree distribution for trade network (the 1980s)

	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1980s
Minimum	0.070	0.070	0.057	0.076	0.045	0.064	0.038	0.045	0.045	0.045	0.055
1st quantile	0.377	0.366	0.344	0.339	0.320	0.320	0.346	0.325	0.312	0.338	0.339
Median	0.490	0.497	0.490	0.468	0.455	0.462	0.478	0.468	0.452	0.462	**0.472**
3rd quantile	0.697	0.726	0.726	0.712	0.691	0.718	0.715	0.717	0.691	0.739	0.713
Maximum	1.000	0.994	0.994	1.000	0.994	1.000	1.000	1.000	1.000	1.000	0.998

Table 2: Normalized degree distribution for trade network (the 1990s)

	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	1990s
Minimum	0.121	0.129	0.177	0.177	0.210	0.194	0.234	0.250	0.250	0.258	0.200
1st quantile	0.484	0.468	0.484	0.484	0.524	0.556	0.573	0.573	0.573	0.589	0.531
Median	0.645	0.637	0.661	0.653	0.702	0.758	0.758	0.766	0.798	0.798	**0.718**
3rd quantile	0.927	0.887	0.903	0.895	0.911	0.935	0.944	0.960	0.952	0.960	0.927
Maximum	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 3: Normalized degree distribution for trade network (the 2000s)

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2000s
Minimum	0.298	0.306	0.290	0.315	0.347	0.331	0.315	0.306	0.323	0.290	0.312
1st quantile	0.637	0.629	0.645	0.677	0.669	0.669	0.669	0.669	0.669	0.669	0.660
Median	0.839	0.831	0.839	0.855	0.847	0.847	0.863	0.847	0.863	0.863	**0.849**
3rd quantile	0.960	0.976	0.984	0.984	0.984	0.984	0.984	0.976	0.976	0.984	0.979
Maximum	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

3.3 Estimation results

We estimate international trade networks in the 1980s, the 1990s and the 2000s by our two-step estimation method. To compare with static network models, we estimate a network snapshot in each sample by the JML estimation in Graham (2017) and the probit model as in Helpman et al. (2008). Below we compare estimation results in three samples by different methods and show how our dynamic network estimation helps to increase identification accuracy and unveil the characteristics of international trade network.

We first analyse results for the trade network in the 1980s. In Table 5, we find that there is a positive state dependence in international trade network, which demonstrates that trade decisions in last year will help countries keep trading in this year. Results on other control variables are qualitatively consistent with Helpman et al. (2008). Specifically, two countries have a higher probability of trading with each other if they are geographically closer, they

Table 4: Descriptive statistics for trade data

	1980s		1990s		2000s	
	mean	standard deviation	mean	standard deviation	mean	standard deviation
log(Distance)	4.177	0.781	4.156	0.780	4.115	0.779
Land border	0.017	0.131	0.016	0.127	0.019	0.135
Island	0.365	0.481	0.371	0.483	0.322	0.467
Landlock	0.271	0.444	0.261	0.439	0.268	0.443
Legal	0.369	0.483	0.379	0.485	0.374	0.484
Language	0.286	0.452	0.304	0.460	0.308	0.462
Colonial ties	0.010	0.098	0.012	0.107	0.012	0.105
Currency Union	0.011	0.103	0.012	0.111	0.013	0.112
Free trade agreement (FTA)	0.006	0.079	0.148	0.355	0.172	0.377
Religion	0.242	0.295	0.255	0.300	0.252	0.302
WTO none	0.205	0.404	0.060	0.237	0.016	0.127
WTO both	0.296	0.457	0.587	0.494	0.754	0.431

Note: *Land border* indicates whether two countries share a common physical boundary. *Legal*, *Language*, *Colonial ties*, *Currency union*, *FTA*, and *Religion* are defined similarly. *Island* is a binary variable that equals one if both countries are islands, and zero otherwise. *Landlock* is a dummy that equals one if both countries have no direct access to sea, and zero otherwise.

are not islands or land locked, they share a common language, they belong to the same currency union or regional free trade agreement, or they have common religion. Estimation results on WTO *none* and WTO *both* indicate that countries are more likely to trade when they are WTO members. Sharing common land border may reduce the probability of trade between countries, possibly due to border conflicts as analyzed in Helpman et al. (2008). The only difference from Helpman et al. (2008) is that our results show that common legal system may have no statistically significant impact on the probability of trade, although the estimation coefficient is positive.

Comparing estimation results between the dynamic and static network models, we can find that the magnitude of covariate coefficients in the dynamic model is smaller. This comparison implies that some estimated effects of time-invariant covariates in the static model may be attributed to state dependence. Meanwhile, standard errors estimated in dynamic network model are smaller than those in the static model, too. Smaller standard errors suggest higher estimation accuracy, which is intuitive because we include more observations from different periods in the dynamic network model. Thus we recommend researchers conduct inference based on the dynamic network model when observed networks from different periods are available at hand.

We turn our attention to estimation results for the trade network in the 1990s and the 2000s.[7] From Tables 6 and 7, we also find strong state dependence in network dynamics. We further find that the impacts of covariates in the dynamic network model is qualitatively consistent with those in the probit models.

Finally, we present distributions of degree heterogeneity in three samples. It is clear

[7] At this point, we exclude *Colonial ties* because the standard errors of its coefficient are very high in all regressions and do not fit the likelihood estimation method. Estimation results of other variables keep quantitatively consistent after the removal.

Table 5: Estimation results of global trade network: The 1980s

Sample range (model)	1980s (Dynamic β-model)		1986 (Static β-model)	1986 (probit)	1980s (probit)	
Variables	$A_{ij,t}$	$A_{ij,t}$	$A_{ij,t}$	$A_{ij,t}$	$A_{ij,t}$	$A_{ij,t}$
$A_{ij,t-1}$	1.381***	1.378***				
	(0.029)	(0.030)				
log(Distance)	-0.638***	-0.639***	-1.175***	-0.675***	-0.609***	-0.609***
	(0.033)	(0.044)	(0.054)	(0.036)	(0.011)	(0.011)
Land border	-0.627***	-0.628***	-1.095***	-0.602***	-0.557***	-0.578***
	(0.102)	(0.130)	(0.244)	(0.151)	(0.043)	(0.043)
Island	-0.293***	-0.293***	-0.546***	-0.317***	-0.282***	-0.282***
	(0.063)	(0.076)	(0.186)	(0.109)	(0.032)	(0.032)
Landlock	-0.214***	-0.215**	-0.283	-0.043	-0.188***	-0.188***
	(0.073)	(0.084)	(0.338)	(0.152)	(0.044)	(0.044)
Legal	0.031	0.032	0.051	0.095**	0.070***	0.069***
	(0.023)	(0.029)	(0.074)	(0.045)	(0.013)	(0.013)
Language	0.301***	0.303***	0.474***	0.299***	0.281***	0.280***
	(0.053)	(0.069)	(0.095)	(0.057)	(0.017)	(0.017)
Colonial ties	0.125	0.123	-0.455	-0.182	0.168	0.172
	(0.522)	(0.755)	(0.893)	(0.515)	(0.153)	(0.153)
Currency union	0.411***	0.414***	1.133***	0.474**	0.245***	0.242***
	(0.105)	(0.075)	(0.292)	(0.198)	(0.056)	(0.056)
FTA	2.398***	2.422***	3.458***	1.194**	1.515***	1.517***
	(0.346)	(0.353)	(0.915)	(0.524)	(0.177)	(0.177)
Religion	0.170***	0.174***	0.399***	0.280***	0.234***	0.232***
	(0.057)	(0.064)	(0.145)	(0.091)	(0.027)	(0.013)
WTO none		-0.518***				-0.194***
		(0.124)				(0.040)
WTO both		0.462***				0.257***
		(0.122)				(0.077)
Year FE	YES	YES			YES	YES
Observations	124030	124030	12403	12403	124030	124030

Notes: all reported coefficients are bias corrected and adjusted standard errors are reported in parentheses. *, **, and *** denote statistic significance at 10%, 5%, and 1% respectively.

Table 6: Estimation results of global trade network: The 1990s

Sample range (model)	1990s (Dynamic β-model)		1996 (Static β-model)	1996 (probit)	1990s (probit)	
Variables	$A_{ij,t}$	$A_{ij,t}$	$A_{ij,t}$	$A_{ij,t}$	$A_{ij,t}$	$A_{ij,t}$
$A_{ij,t-1}$	1.454***	1.448***				
	(0.043)	(0.043)				
log(Distance)	-0.898***	-0.902***	-2.261***	-0.787***	-0.737***	-0.738***
	(0.082)	(0.105)	(0.100)	(0.063)	(0.019)	(0.019)
Land border	-0.242	-0.214	-1.116**	-0.378	-0.314***	-0.314***
	(0.201)	(0.215)	(0.509)	(0.277)	(0.084)	(0.084)
Island	-0.356***	-0.369***	-0.582**	-0.159	-0.353***	-0.353***
	(0.111)	(0.126)	(0.290)	(0.177)	(0.056)	(0.056)
Landlock	-0.346***	-0.339***	-0.539	0.022	-0.241***	-0.240***
	(0.089)	(0.091)	(0.338)	(0.198)	(0.059)	(0.059)
Legal	0.180***	0.176***	0.298***	0.236***	0.226***	0.226***
	(0.028)	(0.029)	(0.105)	(0.063)	(0.020)	(0.020)
Language	0.232***	0.238**	0.465***	0.239***	0.140***	0.141***
	(0.081)	(0.103)	(0.144)	(0.087)	(0.027)	(0.027)
Currency union	0.989***	0.979***	1.255***	0.547**	0.589***	0.588***
	(0.186)	(0.223)	(0.424)	(0.267)	(0.084)	(0.084)
FTA	0.205**	0.205**	-0.291	0.045	-0.007	-0.007
	(0.089)	(0.098)	(0.259)	(0.155)	(0.044)	(0.044)
Religion	0.106	0.124	-0.075	0.389***	0.274***	0.274***
	(0.089)	(0.108)	(0.726)	(0.131)	(0.040)	(0.040)
WTO none		-0.378***				-0.075*
		(0.135)				(0.045)
WTO both		0.065				0.069**
		(0.075)				(0.035)
Year FE	YES	YES			YES	YES
Observations	77750	77750	7775	7775	77750	77750

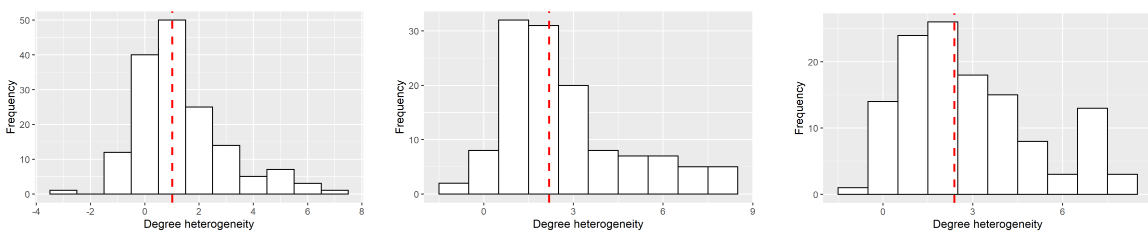
Notes: all reported coefficients are bias corrected and adjusted standard errors are reported in parentheses. *, **, and *** denote statistic significance at 10%, 5%, and 1% respectively.

Table 7: Estimation results of global trade network: The 2000s

Sample range (model)	2000s (Dynamic β-model)		2006 (Static β-model)	2006 (probit)	2000s (probit)	
Variables	$A_{ij,t}$	$A_{ij,t}$	$A_{ij,t}$	$A_{ij,t}$	$A_{ij,t}$	$A_{ij,t}$
$A_{ij,t-1}$	1.442***	1.438***				
	(0.054)	(0.054)				
log(Distance)	-0.991***	-1.017***	-2.876***	-0.829***	-0.746***	-0.746***
	(0.144)	(0.351)	(0.116)	(0.078)	(0.025)	(0.025)
Land border	-0.409*	-0.373	-2.402***	-0.475	-0.485***	-0.484***
	(0.221)	(0.294)	(0.587)	(0.323)	(0.104)	(0.105)
Island	-0.505***	-0.513**	-1.230***	-0.206	-0.313***	-0.313***
	(0.151)	(0.256)	(0.337)	(0.207)	(0.066)	(0.066)
Landlock	-0.100	-0.102	-0.489	0.142	-0.058	-0.058
	(0.093)	(0.093)	(0.330)	(0.196)	(0.063)	(0.063)
Legal	0.115***	0.111**	0.219**	0.200***	0.190***	0.190***
	(0.037)	(0.048)	(0.046)	(0.069)	(0.023)	(0.023)
Language	0.246**	0.256	0.346**	0.183*	0.155	0.155***
	(0.108)	(0.243)	(0.160)	(0.097)	(0.031)	(0.031)
Currency union	1.246***	1.211***	1.683***	0.876**	1.042***	1.041***
	(0.214)	(0.362)	(0.450)	(0.295)	(0.097)	(0.097)
FTA	0.275	0.261	-1.291***	-0.206	0.127**	0.127**
	(0.188)	(0.245)	(0.343)	(0.204)	(0.065)	(0.065)
Religion	0.390***	0.422	0.031	0.557***	0.564***	0.565***
	(0.146)	(0.297)	(0.233)	(0.144)	(0.046)	(0.046)
WTO none		-1.475**				-0.146
		(0.621)				(0.092)
WTO both		0.438**				0.140*
		(0.209)				(0.077)
Year FE	YES	YES			YES	YES
Observations	77500	77500	7750	7750	77500	77500

Notes: all reported coefficients are bias corrected and adjusted standard errors are reported in parentheses. *, **, and *** denote statistic significance at 10%, 5%, and 1% respectively.

that these distributions are not concentrated, which confirms our assumption that countries in network formation process are degree heterogeneous. Compare Figure 2 (a), (b), and (c), we observe that the medians of the 1990s and the 2000s are higher than the 1980s, indicating that countries are more possible to trade with others in later decades. We also find that the distribution of degree heterogeneity in the 2000s have a relatively fat right tail in comparison with the sample of the 1980s, as approximately 20 countries have estimated fixed effect higher than 7. This implies that the some countries can be "hubs" in the trade network and they make the trade network become denser.



(a) The 1980s (b) The 1990s (c) The 2000s

Figure 2: Degree heterogeneity distributions

Note: red dash lines are medians of estimated degree heterogeneity.

4 Concluding remarks

In this paper, we study how to model a dynamic network with first order autoregressive property and additive degree heterogeneity. To circumvent the intractable problem in directly maximizing the dynamic logit type likelihood function with fixed effects, we employ a quadratic exponential distribution and develop a two-step estimation method. Our approach has good theoretical properties and Monte Carlo simulation results in the Supplementary Material confirm estimation performance across a variety of designs.

We further apply this two-step method to estimate global trade networks in three decades: the 1980s, the 1990s, and the 2000s. The estimation results show the state dependence of trade between countries and increasing dynamics with the process of globalization. The effects of covariates in the dynamic network model is qualitatively comparable with those in the baseline probit model. We also find that degree heterogeneity of countries are more dispersed and some countries become hubs of the trade network in the 2000s.

Given the trade conflicts between countries and increasing likelihood of economic crisis in current times, it becomes more important to investigate trade network formation. This need exactly motivates us to develop our dynamic network model and we believe that our method can facilitate theorists to test their network formation theory. Another application area is in economic shock transmission and spillover through international trade network. This quickly growing trend of literature assumes trade network as given, even though countries select their trade partners and macro shocks may also affect trade links. This assumption of exogenous network may result in biased estimate because of the endogeneity problem in trade network. Our method can control for the endogeneity problem and thus help to measure the effect of shock transmission along trade network accurately. Moreover, our dynamic network model is not limited to trade network but can similarly be applied to network formation and shock transmission along network in other disciplines.

Supplementary Material

It includes derivation of equation (5), proofs of Theorems 1-6, Monte Carlo simulation studies and additional estimation results for the sample between 2012 and 2021 collected from UNcomtrade.

References

Allen, T. (2014). Information frictions in trade. *Econometrica*, *82*(6), 2041–2083.

Amemiya, T. (1978). On a two-step estimation of a multivariate logit model. *Journal of Econometrics*, *8*(1), 13–21.

Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society: Series B (Methodological)*, *32*(2), 283–301.

Barbieri, K., Keshk, O. M. G., & Pollins., B. (2009). Trading data: Evaluating our assumptions and coding rules. *Conflict Management and Peace Science.*, *26*(5), 471–491.

Barbieri, K., & Keshk, O. M. G. O. (2016). *Correlates of war project trade data set codebook, version 4.0. online: Https://correlatesofwar.org.*

Barrot, J.-N., & Sauvagnat, J. (2016). Input specificity and the propagation of idiosyncratic shocks in production networks. *The Quarterly Journal of Economics*, *131*(3), 1543–1592.

Bartolucci, F., & Nigro, V. (2010). A dynamic model for binary panel data with unobserved heterogeneity admitting a sqrt n-consistent conditional estimator. *Econometrica*, *78*(2), 719–733.

Bartolucci, F., & Nigro, V. (2012). Pseudo conditional maximum likelihood estimation of the dynamic logit model for binary panel data. *Journal of Econometrics*, *170*(1), 102–116.

Bernard, A. B., & Moxnes, A. (2018). Networks and trade. *Annual Review of Economics*, *10*(1), 65–85.

Bernard, A. B., Moxnes, A., & Saito, Y. U. (2019). Production networks, geography, and firm performance. *Journal of Political Economy*, *127*(2), 639–688.

Boucheron, S., Lugosi, G., & Massart, P. (2013). *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press.

Chaney, T. (2014). The network structure of international trade. *American Economic Review*, *104*(11), 3600–3634.

Chaney, T. (2016). Networks in international trade. *Oxford Handbook of the Economics of Networks*.

Chatterjee, S., Diaconis, P., & Sly, A. (2011). Random graphs with a given degree sequence. *Annals of Applied Probability*, *21*(4), 1400–1435.

Chen, M., Kato, K., & Leng, C. (2021). Analysis of networks via the sparse beta-model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*, *20*(2), 215–232.

Cox, D. R. (1972). The analysis of multivariate binary data. *Applied Statistics*, 113–120.

Erdős, P., & Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, *5*(1), 17–60.

Gilbert, E. N. (1959). Random graphs. *The Annals of Mathematical Statistics*, *30*(4), 1141–1144.

Graham, B. S. (2016). Homophily and transitivity in dynamic network formation. *Working Paper*.

Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica*, *85*(4), 1033–1063.

Graham, B. S., & De Paula, A. (2020). *The econometric analysis of network data*. Academic Press.

Hahn, J., & Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, *72*(4), 1295–1319.

Helpman, E., Melitz, M., & Rubinstein, Y. (2008). Estimating trade flows: Trading partners and trading volumes. *Quarterly Journal of Economics*, *123*(2), 441–487.

Hillar, C., & Wibisono, A. (2018). Maximum entropy distributions on graphs. *arXiv preprint arXiv:1301.3321*.

Jackson, M. O. (2010). *Social and economic networks*. Princeton university press.

Jiang, B., Li, J., & Yao, Q. (2020). Autoregressive networks. *arXiv preprint arXiv:2010.04492*.

McFadden, D. (1974). The measurement of urban travel demand. *Journal of Public Economics*, *3*(4), 303–328.

Newey, W., & McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics*, *4*, 2111–2245.

Yan, T., Jiang, B., Fienberg, S. E., & Leng, C. (2019). Statistical inference in a directed network model with covariates. *Journal of the American Statistical Association*, *114*(526), 857–868.

Yan, T., & Xu, J. (2013). A central limit theorem in the beta-model for undirected random graphs with a diverging number of vertices. *Biometrika*, *100*(2), 519–524.