CAMBRIDGE
UNIVERSITY PRESS

RESEARCH ARTICLE

A pairwise strategic network formation model with group heterogeneity: With an application to international travel

Tadao Hoshino*

School of Political Science and Economics, Waseda University, 1-6-1 Nishi-waseda, Shinjuku-ku, Tokyo 169-8050, Japan
Corresponding author. Email: thoshino@waseda.jp

Action Editor: Fernando Vega-Redondo

Abstract

This study considers a network formation model in which each dyad of agents strategically determines the link status. Our model allows the agents to have unobserved group heterogeneity in the propensity of link formation. For the model estimation, we propose a three-step maximum likelihood method, in which the latent group structure is estimated using the binary segmentation algorithm in the second step. As an empirical illustration, we focus on the network data of international visa-free travels. The results indicate the presence of significant strategic complementarity and a certain level of degree heterogeneity in the network formation behavior.

1. Introduction

Empirical modeling of network formation is an important research topic that has been studied for several decades. While most of these models have been developed in the mathematical statistics literature, as the importance of network structure in many economic activities has been increasingly recognized, there is currently a growing number of econometric studies that focus on network formation in conjunction with the significant advancement in the related econometric techniques.[1]

Econometric studies on network formation can be classified into two types: those that attempt to explicitly incorporate the interaction of individuals in the realizing network structure endogenously affecting the network formation behavior (e.g., Leung, 2015; Mele, 2017; Sheng, 2020) and those that do not account for such simultaneous interactions but emphasize modeling a flexible form of individual heterogeneity (e.g., Graham, 2017; Jochmans, 2018; Dzemski, 2019). For the former type, network formation is modeled as a game in which agents strategically form links to maximize their payoffs. Although this game-theoretic approach is (economic) theoretically well-underpinned, we often encounter serious analytical difficulties due to the presence of multiple equilibria. To circumvent these difficulties, we typically need to introduce some ad hoc behavioral assumptions into the network formation process, or we simply discontinue point-identifying the models and resort to partial identification. Compared with the former, the latter is more "descriptive" than "structural," but has great flexibility in the model specification. These models are relatively easy to implement and, thus, are appealing to empirical researchers. However,

---

* For recent developments regarding econometric approaches for analyzing network formation, we refer readers to, for example, Chandrasekhar (2016) and de Paula (2013).

they are not suitable for analyzing the interactions of agents in network formation, which should be an essential factor in economic and social network data.

These two types of econometric models have their own advantages. Hence, it is an ingenuous idea to construct a new model by combining them. However, to my best knowledge, there are only a few papers that address this way of model extension (e.g., Graham, 2016; Graham & Pelican, 2020; Pelican & Graham, 2020). Graham (2016) considers the "dynamic" (rather than the instantaneous) interdependencies in network formation. The latter two papers consider a quite general framework that incorporates both a general form of strategic interaction and unobserved degree heterogeneity into a single model. However, they mainly focus on testing the presence of interactions, but not on the estimation of the models.

In this paper, we propose a new "pairwise" network formation model that is empirically tractable while retaining the nice properties of the above-mentioned approaches. More specifically, we assume that each link connection is determined by the strategic interaction solely between the corresponding dyad of agents, without affecting or being affected by other dyads, rather than regarding the realized network as a consequence of a large n-player game. Although ignoring such network externalities limits the range of applications of our model, it would still cover a fairly large number of interesting empirical situations, and, most importantly, the multiple equilibria problem can be greatly mitigated. In addition, we allow each agent's payoff to depend on two unobserved preference parameters that represent his/her outgoing and incoming propensity, which we refer to as the sender and the receiver effect, respectively.

For estimating our pairwise network formation model, assuming that the model error terms follow some parametric distribution, we propose using the maximum likelihood (ML) method. Although multiple equilibria can occur even in our binary game situation, we can address this issue in the same manner as in Bresnahan & Reiss (1990) and Berry (1992). The agent-specific preference heterogeneity parameters are treated as the fixed-effect parameters to be estimated. Note that if we include these heterogeneity parameters directly into the model, as the dimension of the parameters increases proportionally to the sample size, our ML estimator suffers from the *incidental parameter* problem. As will be confirmed in the numerical simulations reported later, the incidental parameter bias can be severe. Thus, to avoid the incidental parameter problem, as a second novel part of this study, we focus on a situation where the individuals can be classified into several unknown groups and their specific effects are homogeneous within each group. If the number of the latent groups is fixed, we can expect that the ML estimator becomes asymptotically unbiased at the parametric rate, and hence, the standard inference procedure can be applied.

Note that if one is interested only in the "common" parameters rather than in the latent group heterogeneity, or believes that the agents are intrinsically heterogeneous and no such group structure exists, employing a conventional fixed-effect bias correction method such as those in Graham (2017) and Yan et al. (2019) would be reasonable to address the incidental parameter problem. However, there may be some empirical fields (e.g., international relations, marketing) where identifying some group structure itself has important policy implications. Another practical benefit of grouping the observations is that, even when they are not exactly grouped into small groups in reality, it can play the role of regularization to avoid overfitting to the data.

In the literature on statistical network data analysis, uncovering such latent group structures in networks has been intensively studied (e.g., Newman & Girvan, 2004; Bickel & Chen, 2009; Fortunato, 2010; Karrer & Newman, 2011; Rohe et al., 2011; Abbe, 2017). In panel data analysis also, identification of unobserved grouped heterogeneity is one of the most active research areas (e.g., Bonhomme & Manresa, 2015; Ke et al., 2016; Su et al., 2016; Wang et al., 2018). Although the application of these grouping methods to econometric network models has been relatively limited thus far, it is a promising approach, as discussed in Bonhomme (2020). In this study, among several alternative methods, we adopt the binary segmentation (BS) method (see, e.g., Bai, 1997; Ke et al., 2016; Lian et al., 2021; Wang & Su, 2021). Compared to the other grouping methods, the BS method has several favorable properties including fast computation speed and robustness.

The whole estimation procedure is divided into three steps. The first step is to obtain the ML estimator without considering the latent group structures. Although this estimator suffers from the incidental parameter bias, it is still possible to produce consistent estimates for the heterogeneity parameters. In the second step, we apply the BS method with respect to the estimated sender effect parameters and the receiver effect parameters separately to identify each agent's group memberships. In the third step, we re-estimate the model using the ML method given the estimated group structure. Under certain regularity conditions, we show that the proposed estimator is asymptotically unbiased and normal at the parametric rate. Furthermore, the estimator is asymptotically equivalent to the "oracle" estimator that is obtained based on the (unknown) true group memberships.

To illustrate our model empirically, we investigate the formation of international visa-free travel networks, where the dependent variable of interest is defined as follows: $g_{i,j} = 1$ if country i allows the citizens in country j to visit i without visas and $g_{i,j} = 0$ if not. We apply our model framework to the network of 57 countries selected mainly from Asia, the Middle East, the former USSR, and Oceania. As expected, we find the presence of a certain level of degree heterogeneity in both the sender and the receiver effects. Interestingly, there seems to be a negative correlation between the sender effects and the receiver effects; in other words, there is a tendency that a country's sender effect increases as its receiver effect decreases. Our estimation result also suggests that there is a significant strategic complementarity in the network formation behavior. Another interesting finding is that the countries are *homophilous*—tend to connect with similar others—in terms of the political system.

Organization of the paper: The remainder of the paper is organized as follows. In Section 2, we formally introduce the model investigated in this study. In this section, we demonstrate that our pairwise model exhibits multiple equilibria and discuss the conditions under which the model can be point-identified. Section 3 provides a detailed explanation about our three-step ML estimator. We also investigate the asymptotic properties of the proposed estimator in this section. In Section 4, we present a set of Monte Carlo experiments to evaluate the finite sample performance of the proposed estimator. Section 5 presents our empirical analysis, and, finally, Section 6 concludes. All the technical details are relegated to Appendix.

Notation: For a natural number n, I_n denotes an $n \times n$ identity matrix. $\mathbf{1}\{\cdot\}$ denotes the indicator function, which is one if its argument is true and zero otherwise. For a matrix A, we use $||A||$ to denote its Frobenius norm: $||A|| = \sqrt{\mathrm{tr}\{AA^\top\}}$, where $\mathrm{tr}\{\cdot\}$ is a trace of a matrix. When A is a square matrix, we use $\lambda_{\min}(A)$ to denote its smallest eigenvalue. For a vector $\mathbf{a} = (a_1, \ldots, a_k)^\top$, $||\mathbf{a}||_\infty$ denotes its maximum norm: $||\mathbf{a}||_\infty = \max_{1 \le i \le k} |a_i|$. For a general set $\mathcal{X}$, we use $\mathcal{X}^{\mathrm{int}}$ to denote its interior. In addition, $|\mathcal{X}|$ denotes the cardinality of $\mathcal{X}$. c (possibly with subscript) denotes a generic positive constant whose exact value may vary per case.

2. Model setup and identification

2.1 Pairwise strategic network formation model

Suppose that we have a sample of n agents that form social networks whose connections are represented by an $n \times n$ adjacency matrix $G_n = (g_{i,j})_{1 \le i,j \le n}$. These agents can be individuals, firms, municipalities, or nations depending on the context. The network is directed; that is, regardless of the value of $g_{j,i}$, we observe $g_{i,j} = 1$ if agent i links to j and $g_{i,j} = 0$ otherwise. There are no self-loops; that is, the diagonal elements of G_n are all zero. Throughout the paper, we assume that the status of $(g_{i,j}, g_{j,i})$ is determined solely by the pair of agents (i, j), without considering the status of other network links. Specifically, for each pair (i, j), suppose that i's marginal payoff of forming a link with j given $g_{j,i} = q$ is written as

$$u_{i,j}(q) = Z_{i,j}^\top \beta_0 + \alpha_0 q + A_{0,i} + B_{0,j} - \epsilon_{i,j}, \ \text{ for } i \ne j.$$

Here, $Z_{i,j} \in \mathbb{R}^{d_z}$ is a vector of observed covariates, $A_{0,i} \in \mathbb{R}$ is agent i's individual specific effect as a "sender," $B_{0,j} \in \mathbb{R}$ is j's individual specific effect as a "receiver," $\epsilon_{i,j} \in \mathbb{R}$ is an unobservable payoff component, and $\beta_0 \in \mathbb{R}^{d_z}$ and $\alpha_0 \in \mathbb{R}$ are unknown coefficient vector and the interaction effect parameter, respectively. The covariates $Z_{i,j}$ and $Z_{j,i}$ may contain common elements; however, in the later discussion, we require that they must have some agent-specific elements that can vary across the partners. The individual specific effects $A_{0,i}$ and $B_{0,j}$, which we call the sender and the receiver effect, respectively, can be interpreted as the level of i's willingness to create connections with others and the popularity of j, respectively, that generate degree heterogeneity across the agents. We treat $\{(A_{0,i}, B_{0,i})\}$ as fixed-effect parameters to be estimated.

We assume that the agents have complete information; that is, the realizations of $(Z_{i,j}, Z_{j,i})$ and $(\epsilon_{i,j}, \epsilon_{j,i})$ are common knowledge to both i and j. Then, if we assume that the observed network G_n is formed by a collection of Nash equilibrium actions, we obtain the following econometric model:

$$
\begin{aligned}
g_{i,j} &= \mathbf{1}\left\{Z_{i,j}^\top \beta_0 + \alpha_0 g_{j,i} + A_{0,i} + B_{0,j} \geq \epsilon_{i,j}\right\} \\
g_{j,i} &= \mathbf{1}\left\{Z_{j,i}^\top \beta_0 + \alpha_0 g_{i,j} + A_{0,j} + B_{0,i} \geq \epsilon_{i,j}\right\}, \quad \text{for } i \neq j
\end{aligned}
\tag{1}
$$

The following are the two examples to which the above framework can be potentially applied.

Example 2.1 (Online social networking). The analysis of online social networking behavior is an active research topic in network science. For some social networking sites, users can easily establish links to others (become a *follower*) without mutual consent. In addition, whether a person becomes a follower of someone is often irrelevant to who else they are following. This would be a situation where our framework reasonably fits.

Example 2.2 (International visa-free network). In the research on international migration and tourism, investigating the determinants and impacts of visa policies is one of the central interests (e.g., Neiman & Swagel, 2009; Neumayer, 2010; McKay & Tekleselassie, 2018). As bilateral visa policies are naturally observed as a consequence of strategic (economic and/or political) interactions between the two countries, our model would be an appropriate analytical tool here. In our empirical study presented in Section 5, by setting $g_{i,j} = 1$ if country i allows visa-free entry for the citizens of country j and $g_{i,j} = 0$ if not, we will show that the magnitude of the bilateral interaction in visa policies is significant.

In these examples, we can naturally imagine that the strategic interaction effect α_0 is positive (i.e., strategic complements). We assume that strategic complementarity would be reasonable for most empirical situations of network formation games. Then, throughout the paper, we impose this assumption: $\alpha_0 > 0$. Under strategic complementarity, each pair's Nash equilibrium action can be summarized in Figure 1. As shown in the figure, the space of $(\epsilon_{i,j}, \epsilon_{j,i})$ cannot be partitioned into non-overlapping regions associated with the four alternative realizations of $(g_{i,j}, g_{j,i})$. That is, both $(g_{i,j}, g_{j,i}) = (1, 1)$ and $(g_{i,j}, g_{j,i}) = (0, 0)$ can occur in the shaded area in the figure, and the link status is not uniquely determined in this area (i.e., multiple equilibria). This non-uniqueness of model-consistent decisions is called *incompleteness* and has been extensively studied in the literature on simultaneous equation models for discrete outcomes (e.g., Tamer, 2003; Lewbel, 2007; Ciliberto & Tamer, 2009; Chesher & Rosen, 2020).

There are several approaches to handle this incompleteness issue in the literature.[2] Among them, this study adopts the traditional approach developed by Bresnahan & Reiss (1990) and Berry (1992) that focuses only on the unique equilibrium outcomes. That is, we consider estimating the model using only the information about "one-way links" in the network. Specifically, we develop an ML estimator based on the probabilities of the three regions partitioned by the bold solid lines in Figure 1.
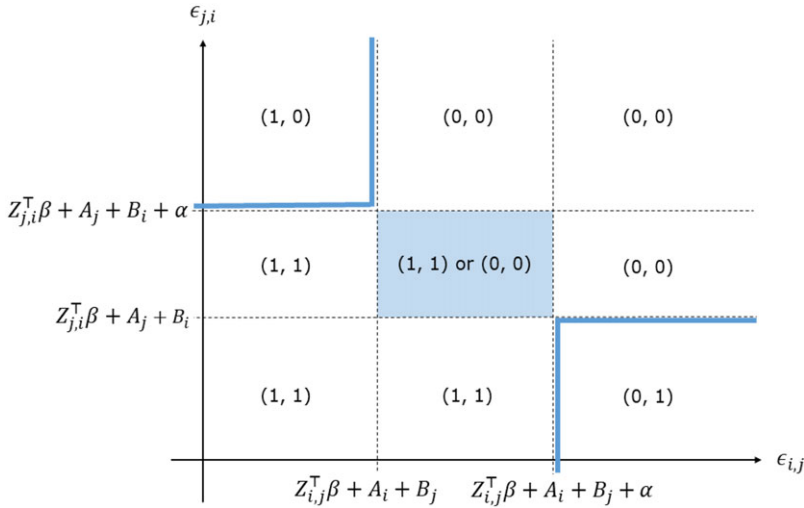
Figure 1. Pure strategy of Nash equilibrium.

In the following, we assume that $\{\epsilon_{i,j}\}$ are identically distributed with a known cumulative distribution function (CDF) F. Further, we assume that the pairs $\{(\epsilon_{i,j}, \epsilon_{j,i})\}$ are independent and identically distributed (i.i.d.) across pairs, and their joint distribution is represented by $H(\cdot, \cdot; \rho_0)$ such that $\Pr(\epsilon_{i,j} \leq a_1, \epsilon_{j,i} \leq a_2) = H(a_1, a_2; \rho_0)$, where $\rho_0 \in \mathbb{R}$ is a parameter controlling the correlation between $\epsilon_{i,j}$ and $\epsilon_{j,i}$. Define $y_{i,j}^{(1,0)} \equiv \mathbf{1}\{(g_{i,j}, g_{j,i}) = (1,0)\}$ and $y_{i,j}^{(0,1)} \equiv \mathbf{1}\{(g_{i,j}, g_{j,i}) = (0,1)\}$. Let $\chi_{n,a}$ be the a-th column of I_n, $\mathbf{A}_0 = (A_{0,1}, \ldots, A_{0,n})^\top$, and $\mathbf{B}_0 = (B_{0,1}, \ldots, B_{0,n})^\top$, so that we can write $Z_{i,j}^\top \beta_0 + A_{0,i} + B_{0,j} = W_{i,j}^\top \Pi_0$, where $W_{i,j} = (Z_{i,j}^\top, \chi_{n,i}^\top, \chi_{n,j}^\top)^\top$ and $\Pi_0 = (\beta_0^\top, \mathbf{A}_0^\top, \mathbf{B}_0^\top)^\top$. In addition, we denote $\theta_0 = (\beta_0^\top, \alpha_0, \rho_0)^\top$ and $\gamma_0 = (\mathbf{A}_0^\top, \mathbf{B}_0^\top)^\top$. Then, the probabilities of $\{y_{i,j}^{(1,0)} = 1\}$ and $\{y_{i,j}^{(0,1)} = 1\}$ are respectively given as follows:

$$P_{i,j}^{(1,0)}(\theta_0, \gamma_0) \equiv F(W_{i,j}^\top \Pi_0) - H(W_{i,j}^\top \Pi_0, W_{j,i}^\top \Pi_0 + \alpha_0; \rho_0),$$
$$P_{i,j}^{(0,1)}(\theta_0, \gamma_0) \equiv F(W_{j,i}^\top \Pi_0) - H(W_{i,j}^\top \Pi_0 + \alpha_0, W_{j,i}^\top \Pi_0; \rho_0)$$

Here, note that the equalities $y_{i,j}^{(1,0)} = y_{j,i}^{(0,1)}$ and $P_{i,j}^{(1,0)}(\theta, \gamma) = P_{j,i}^{(0,1)}(\theta, \gamma)$ hold. Thus, the likelihood function can be concentrated with respect to $(y_{i,j}^{(1,0)}, P_{i,j}^{(1,0)}(\theta, \gamma))$; hereinafter, we omit the superscripts and denote $y_{i,j} = y_{i,j}^{(1,0)}$ and $P_{i,j}(\theta, \gamma) = P_{i,j}^{(1,0)}(\theta, \gamma)$ when there is no confusion. Then, the log-likelihood function can be written as

$$\mathcal{L}_n(\theta, \gamma) = \frac{2}{N} \sum_{i=1}^{n} \sum_{j>i} \left[y_{i,j} \ln P_{i,j}(\theta, \gamma) + y_{j,i} \ln P_{j,i}(\theta, \gamma) + (1 - y_{i,j} - y_{j,i}) \ln (1 - P_{i,j}(\theta, \gamma) - P_{j,i}(\theta, \gamma)) \right]$$

$$= \frac{1}{N} \sum_{i=1}^{n} \sum_{j \neq i} \left[y_{i,j} \ln P_{i,j}(\theta, \gamma) + y_{j,i} \ln P_{j,i}(\theta, \gamma) + (1 - y_{i,j} - y_{j,i}) \ln (1 - P_{i,j}(\theta, \gamma) - P_{j,i}(\theta, \gamma)) \right] \qquad (2)$$

$$= \frac{1}{N} \sum_{i=1}^{n} \sum_{j \neq i} \left[2y_{i,j} \ln P_{i,j}(\theta, \gamma) + (1 - 2y_{i,j}) \ln (1 - P_{i,j}(\theta, \gamma) - P_{j,i}(\theta, \gamma)) \right]$$

where $N \equiv n(n-1)$. As above, we can consider three equivalent representations for the log-likelihood function and switch between them according to analytical convenience.

Remark 2.1 (Asymptotic framework). We investigate the identification and asymptotic properties of our estimator as n, the network size, increases, which means that we consider a sequence of networks $\{G_n\}$ with n tending to infinity. However, when each agent represents a country (as in our empirical analysis), the size of the network is fixed. In such situations, in the causal inference literature, researchers often adopt a finite-population design-based framework (cf. Abadie et al., 2020). Particularly, Athey et al. (2018) considered the hypothesis testing of the causal effects under network interference using a finite-population network model. However, to the best of our knowledge, finite-population inference in the context of network formation models is still a theoretically unexplored topic.

2.2 Discrete heterogeneity

This study considers situations where the agents are grouped into several sub-samples, and the individual fixed effects are heterogeneous across these groups but are homogeneous within the groups in the following manner:

$$A_{0,i} = \sum_{k=1}^{K^A} a_{0,k} \cdot \mathbf{1}\{i \in \mathcal{C}^A_{0,k}\}, \quad B_{0,i} = \sum_{k=1}^{K^B} b_{0,k} \cdot \mathbf{1}\{i \in \mathcal{C}^B_{0,k}\} \tag{3}$$

That is, the agents can be classified into K^A groups $\mathcal{C}^A_0 \equiv \{\mathcal{C}^A_{0,1}, \ldots, \mathcal{C}^A_{0,K^A}\}$ in terms of the sender effects, where K^A is the total number of groups, which form a partition of $\{1, \ldots, n\}$ into K^A subsets. Similarly, in terms of the receiver effects, the agents can be grouped as $\mathcal{C}^B_0 \equiv \{\mathcal{C}^B_{0,1}, \ldots, \mathcal{C}^B_{0,K^B}\}$. The group where each individual belongs to remains unknown to us. Meanwhile, it is often assumed in the literature that the number of groups is known to researchers (e.g., Bonhomme & Manresa, 2015; Okui & Wang, 2021). Then, following these studies, we treat K^A and K^B as known values and assume that $K^A, K^B \geq 2$.[3] We discuss how to choose K^A and K^B in practice in Section 5. Note that transforming $(\mathbf{A}_0, \mathbf{B}_0)$ to $(\mathbf{A}_0 + c, \mathbf{B}_0 - c)$ for any constant c does not change the model (1). Thus, without loss of generality, we assume that $a_{0,1} = 0$ for normalization.

Under this setup, the full ML estimator solves

$$\max_{(\theta, a_2, \ldots, a_{K^A}, b_1, \ldots, b_{K^B}, \mathcal{C}^A, \mathcal{C}^B)} \mathcal{L}_n(\theta, \mathbf{A}, \mathbf{B}) \tag{4}$$

where $\mathcal{C}^A \equiv \{\mathcal{C}^A_1, \ldots, \mathcal{C}^A_{K^A}\}$, $\mathcal{C}^B \equiv \{\mathcal{C}^B_1, \ldots, \mathcal{C}^B_{K^B}\}$, $A_i = \sum_{k=1}^{K^A} a_k \cdot \mathbf{1}\{i \in \mathcal{C}^A_k\}$, and $B_i = \sum_{k=1}^{K^B} b_k \cdot \mathbf{1}\{i \in \mathcal{C}^B_k\}$. The maximization problem in (4) is clearly a combinatorial (NP-hard) optimization problem. In the context of panel data models, several authors have proposed iterative algorithms to obtain a (local) solution efficiently to the problems similar to (4) (e.g., Bonhomme & Manresa, 2015; Liu et al., 2020). However, the iterative algorithm is still computationally demanding. More importantly, it cannot be directly applied to our network model where each agent's heterogeneity parameters affect not only the value of his/her own likelihood function but also that of the others.

Hence, in this paper, we propose to break down the maximization problem in (4) into three steps. The first step is to estimate $\gamma_0 = (\mathbf{A}_0^\top, \mathbf{B}_0^\top)^\top$ using the full ML estimator based on the log-likelihood function in (2) without explicitly considering the group structure. Given the consistent estimates of these parameters, the second step is to estimate the group memberships $\mathcal{C}^A_0$ and $\mathcal{C}^B_0$ using the BS algorithm. The final step is to solve (4) with the group structure replaced by the estimated $\mathcal{C}^A_0$ and $\mathcal{C}^B_0$.

2.3 Identification

Before presenting the estimation procedure in detail, we discuss identification conditions for the parameters in model (1). It is important to note that, even when individual heterogeneity parameters have only finite variations groupwisely, each individual's parameters must be point-identified separately to estimate the group structure consistently. A reason for this is that our three-step estimator requires preliminary consistent estimates of $\mathbf{A}_0$ and $\mathbf{B}_0$ in the estimation of the group structure.[4] Here, again, we need some location normalization to identify $\mathbf{A}_0$ and $\mathbf{B}_0$. Similarly as above, we set $A_{0,1} = 0$ without loss of generality. To facilitate the discussion, we also introduce several simplifying assumptions, some of which are mentioned previously.

Assumption 2.1. (i) The payoff disturbances $\{\epsilon_{i,j}\}$ are identically distributed on the whole $\mathbb{R}$ with a known strictly increasing marginal CDF $F(\cdot)$. (ii) The pairs $\{(\epsilon_{i,j}, \epsilon_{j,i})\}$ are i.i.d. across dyads with joint CDF $H(\cdot, \cdot; \rho_0)$, and $H(\cdot, \cdot; \rho)$ is strictly increasing in each argument for all $\rho \in \mathcal{R}$. (iii) $F(\cdot)$ is three times continuously differentiable, and $H(\cdot, \cdot; \cdot)$ is three times continuously differentiable with respect to all arguments.

Let $\Theta \equiv \mathcal{B} \times \mathcal{A} \times \mathcal{R}$, $\mathbb{A}_n \equiv \{0\} \times \mathbb{A}^{n-1}$, $\mathbb{B}_n \equiv \mathbb{B}^n$, and $\mathbb{C}_n \equiv \mathbb{A}_n \times \mathbb{B}_n$, where $\mathcal{B} \subset \mathbb{R}^{d_z}$, $\mathcal{A} \subset \mathbb{R}_{++}$, $\mathcal{R} \subset \mathbb{R}$, $\mathbb{A} \subset \mathbb{R}$, and $\mathbb{B} \subset \mathbb{R}$ are parameter spaces for β, α, ρ, A_i's, and B_i's, respectively.

Assumption 2.2. (i) $\theta_0 \in \Theta^{\text{int}}$, where Θ is compact. (ii) For all i, $A_{0,i} \in \mathbb{A}^{\text{int,}}$ and $B_{0,i} \in \mathbb{B}^{\text{int}}$, where $\mathbb{A}$ and $\mathbb{B}$ are compact.

Assumption 2.3. The covariates $\{Z_{i,j}\}$ are uniformly bounded.

In Assumption 2.1(i), we assume that the marginal CDF of the error term is known. This assumption is typically adopted in the estimation of complete information games. As shown by Khan & Nekipelov (2018), when the marginal CDFs of $(\epsilon_{i,j}, \epsilon_{j,i})$ are unknown, it is generally impossible to estimate the interaction effect α_0 at the parametric rate. Assumption 2.1(ii) requires that the error terms are independent across dyads. Note that the parameters $(A_{0,i}, B_{0,i})$ have the role of accommodating all unobserved payoff components specific to i. Therefore, assuming the independence within the remainders $\{\epsilon_{i,1}, \ldots, \epsilon_{i,n}\}$ should not be too restrictive. The other requirements in Assumption 2.1 are standard in that they are satisfied in most of commonly used parametric models. In Assumption 2.2(ii), we assume that the fixed-effect parameters $\{(A_{0,i}, B_{0,i})\}$ are bounded. Although imposing boundedness on the degree heterogeneity parameters is commonly accepted in the literature on network formation models, some studies consider a more general framework where $||\mathbf{A}_0||_\infty$ and $||\mathbf{B}_0||_\infty$ can grow slowly (e.g., Yan et al., 2016, 2019). The parameter space $\mathcal{R}$ for the correlation parameter ρ depends on the choice of the functional form of H, which is typically $\mathcal{R} = [-1, 1]$. Assumption 2.3 should not be restrictive in practice. Hereinafter, we fix the values of $\{Z_{i,j}\}$; that is, we interpret the following analysis as being conditional on the realization of $\{Z_{i,j}\}$. Thus, any randomness in the model is considered to be due to the randomness of $\{\epsilon_{i,j}\}$.

An important implication from Assumptions 2.1–2.3 is that the one-way link probabilities $\{P_{i,j}(\theta, \gamma)\}$ are uniformly bounded away from 0 and 1 for all possible parameter values. In other words, we are assuming that our networks are dense such that the number of one-way links per agent will be about proportional to the number of sampled agents. The plausibility of this assumption depends on the context of application. For example, international trade networks and the visa-free travel networks given in Example 2.2 and Section 5 may be regarded as dense networks.[5]

Assumption 2.4. For any $(\beta_1, \alpha_1, \gamma_1), (\beta_2, \alpha_2, \gamma_2) \in \mathcal{B} \times \mathcal{A} \times \mathbb{C}_n$ such that $(\beta_1, \alpha_1, \gamma_1) \neq (\beta_2, \alpha_2, \gamma_2)$, either or both (a) and (b) hold:

$$a. \quad \liminf_{n \to \infty} \frac{1}{N} \sum_{i=1}^{n} \sum_{j \neq i} \mathbf{1} \left\{ W_{i,j}^\top (\Pi_1 - \Pi_2) > 0, \ W_{j,i}^\top (\Pi_1 - \Pi_2) + \alpha_1 - \alpha_2 < 0 \right\} > 0$$

$$b. \quad \liminf_{n \to \infty} \frac{1}{N} \sum_{i=1}^{n} \sum_{j \neq i} \mathbf{1} \left\{ W_{i,j}^\top (\Pi_1 - \Pi_2) < 0, \ W_{j,i}^\top (\Pi_1 - \Pi_2) + \alpha_1 - \alpha_2 > 0 \right\} > 0$$

where $\Pi_1 = (\beta_1^\top, \gamma_1^\top)^\top$, and $\Pi_2 = (\beta_2^\top, \gamma_2^\top)^\top$.

Assumption 2.4 is our main identification condition, which basically requires the following two conditions. The first condition is a standard full-rank condition for $\{W_{i,j}\}$ and $\{W_{j,i}\}$. The second condition is that at least either $Z_{i,j}$ or $Z_{j,i}$ should contain agent-specific covariates that have large enough supports and also have variations across all potential partners. If no such variables exist, since the signs of $Z_{i,j}^\top (\beta_1 - \beta_2)$ and $Z_{j,i}^\top (\beta_1 - \beta_2)$ cannot differ for some parameter values for all (i,j)'s, Assumption 2.4 does not hold. It should be noted that this assumption is not inconsistent with Assumption 2.3 under the compactness of the parameter space (i.e., Assumption 2.2). While the existence of player-specific continuous variables with unbounded supports is typically required in the identification of non/semiparametric game models—the so-called *identification-at-infinity* argument (see, e.g., Tamer, 2003; Kline, 2015), we can develop our identification result under less stringent conditions owing to the full-parametric model specification.

Theorem 2.1 (Identification). Suppose that Assumptions 2.1(i)–(ii) and 2.2–2.4 hold. (i) Then, if ρ_0 is known, $(\beta_0, \alpha_0, \gamma_0)$ is point-identified on $\mathcal{B} \times \mathcal{A} \times \mathbb{C}_n$. (ii) If ρ_0 is unknown, but it is a unique maximizer of $\mathcal{L}_n^*(\rho)$, then (θ_0, γ_0) is point-identified on $\Theta \times \mathbb{C}_n$, where $\mathcal{L}_n^*(\rho) \equiv \mathbb{E}\mathcal{L}_n((\widetilde{\beta}_0(\rho)^\top, \widetilde{\alpha}_0(\rho), \rho)^\top, \widetilde{\gamma}_0(\rho))$, and $(\widetilde{\beta}_0(\rho), \widetilde{\alpha}_0(\rho), \widetilde{\gamma}_0(\rho)) \equiv \operatorname{argmax}_{(\beta, \alpha, \gamma) \in \mathcal{B} \times \mathcal{A} \times \mathbb{C}_n} \mathbb{E}\mathcal{L}_n((\beta^\top, \alpha, \rho)^\top, \gamma)$.

These identification results are similar to those in Theorem 2 of Aradillas-Lopez & Rosen (2019). In the literature on network formation models, it is often assumed that $\epsilon_{i,j}$ and $\epsilon_{j,i}$ are independent (e.g., Hoff et al., 2002; Jochmans, 2018; Yan et al., 2019). If they are independent, since $\rho_0 = 0$ is known, condition (i) is satisfied. Condition (ii) clearly depends on the choice of H function and is difficult to verify in general; however, this is directly empirically testable. A more primitive sufficient condition for this is that $\mathcal{L}_n^*(\rho)$ is strictly concave, which is satisfied when $\partial^2 \mathcal{L}_n^*(\rho)/(\partial \rho)^2$ is strictly negative under Assumption 2.1(iii). We provide an explicit form of $\partial^2 \mathcal{L}_n^*(\rho)/(\partial \rho)^2$ in Appendix C.1. For another identification condition for ρ_0, see, for example, Theorem B.2 of Hoshino & Yanagi (2021), that is based on the monotonicity of the likelihood function with respect to ρ.

Remark 2.2 (Other identification strategies). There are several other routes for identification of our model than the one in Theorem 2.1. For example, as our model is fully parametric, a classical parametric identification approach may be used based on the properties of the information matrix (e.g., Rothenberg, 1971; Bjorn & Vuong, 1984), although it is not easy to verify in practice. If one admits the existence of a player-specific continuous variable that has a positive density on the whole $\mathbb{R}$, then the model can be easily identified by the identification-at-infinity approach in the same manner as in Tamer (2003). Since assuming the existence of such unbounded variables is restrictive in practice, several authors have proposed other approaches based on some shape restrictions on the distribution of unobservables (e.g., Kline, 2016). Investigating whether their approaches can be applied to our model is an interesting topic, but it is left for a future work.

3. Three-step ML estimation

3.1 First-step ML estimator

The first step of the ML estimation aims to obtain consistent estimates of $\gamma_0 = (\mathbf{A}_0^\top, \mathbf{B}_0^\top)^\top$. Let

$$(\widehat{\theta}_n, \widehat{\gamma}_n) = \underset{(\theta, \gamma) \in \Theta \times \mathbb{C}_n}{\mathrm{argmax}} \mathcal{L}_n(\theta, \gamma) \tag{5}$$

where $\widehat{\theta}_n = (\widehat{\beta}_n^\top, \widehat{\alpha}_n, \widehat{\rho}_n)^\top$, and $\widehat{\gamma}_n = (\widehat{\mathbf{A}}_n^\top, \widehat{\mathbf{B}}_n^\top)^\top$. Below, we present the asymptotic properties of the initial full ML estimator in (5) with the main focus on $\widehat{\gamma}_n$. Instead of introducing particular identification conditions, for generality, we directly assume that the true parameter (θ_0, γ_0) is a unique maximizer of $\mathbb{E}\mathcal{L}_n(\theta, \gamma)$.

Assumption 3.1. $\mathbb{E}\mathcal{L}_n(\theta, \gamma)$ is uniquely maximized at $(\theta_0, \gamma_0) \in \Theta \times \mathbb{C}_n$ for all sufficiently large n.

We first establish several consistency results in the next theorem.

Theorem 3.1. Suppose that Assumptions 2.1–2.3 and 3.1 hold. Then, we have (i) $\widehat{\theta}_n \xrightarrow{P} \theta_0$, (ii) $\frac{1}{n} \sum_{i=1}^n |\widehat{A}_{n,i} - A_{0,i}| \xrightarrow{P} 0$, (iii) $\frac{1}{n} \sum_{i=1}^n |\widehat{B}_{n,i} - B_{0,i}| \xrightarrow{P} 0$, and (iv) $||\widehat{\gamma}_n - \gamma_0||_\infty \xrightarrow{P} 0$.

Next, we derive the convergence rate of $\widehat{\gamma}_n$. To this end, it is convenient to re-define $\widehat{\theta}_n$ and θ_0 as

$$\widehat{\theta}_n = \underset{\theta \in \Theta}{\mathrm{argmax}} \, \mathcal{L}_n(\theta, \widetilde{\gamma}_n(\theta)) \text{ and } \theta_0 = \underset{\theta \in \Theta}{\mathrm{argmax}} \, \mathbb{E}\mathcal{L}_n(\theta, \widetilde{\gamma}_0(\theta))$$

respectively, where $\widetilde{\gamma}_n(\theta) \equiv \mathrm{argmax}_{\gamma \in \mathbb{C}_n} \mathcal{L}_n(\theta, \gamma)$ and $\widetilde{\gamma}_0(\theta) \equiv \mathrm{argmax}_{\gamma \in \mathbb{C}_n} \mathbb{E}\mathcal{L}_n(\theta, \gamma)$ for any given $\theta \in \Theta$. Further, we define $\gamma_{-1} = (A_2, \ldots, A_n, B_1, \ldots, B_n)^\top$,

$$\underset{(2n-1)\times(2n-1)}{\mathcal{H}_{n,\gamma\gamma}(\theta, \gamma)} \equiv \frac{\partial^2 \mathcal{L}_n(\theta, \gamma)}{\partial \gamma_{-1} \partial \gamma_{-1}^\top}, \qquad \underset{(d_z+2)\times(d_z+2)}{\mathcal{H}_{n,\theta\theta}(\theta, \gamma)} \equiv \frac{\partial^2 \mathcal{L}_n(\theta, \gamma)}{\partial \theta \partial \theta^\top}, \qquad \underset{(2n-1)\times(d_z+2)}{\mathcal{H}_{n,\gamma\theta}(\theta, \gamma)} \equiv \frac{\partial^2 \mathcal{L}_n(\theta, \gamma)}{\partial \gamma_{-1} \partial \theta^\top}$$

and $\mathcal{H}_{n,\theta\gamma}(\theta, \gamma) \equiv \mathcal{H}_{n,\gamma\theta}(\theta, \gamma)^\top$. The exact form of $\mathcal{H}_{n,\gamma\gamma}(\theta, \gamma)$ is given in (A.2) in Appendix A. Finally, let

$$\underset{(d_z+2)\times(d_z+2)}{\mathcal{I}_{n,\theta\theta}(\theta, \gamma)} \equiv \mathcal{H}_{n,\theta\theta}(\theta, \gamma) - \mathcal{H}_{n,\theta\gamma}(\theta, \gamma) \left[\mathcal{H}_{n,\gamma\gamma}(\theta, \gamma)\right]^{-1} \mathcal{H}_{n,\gamma\theta}(\theta, \gamma)$$

This $\mathcal{I}_{n,\theta\theta}(\theta, \gamma)$ serves as the Hessian matrix for the concentrated ML estimator $\widehat{\theta}_n$ (cf. Amemiya, 1985).

Assumption 3.2. $\widetilde{\gamma}_0(\theta)$ uniquely exists uniformly on $\{\Theta : ||\theta - \theta_0|| \leq \varepsilon\}$ for an arbitrary small $\varepsilon > 0$.

Assumption 3.3. For an arbitrary small $\varepsilon > 0$, there exist $c_\gamma, c_\theta > 0$ that may depend on ε such that (i) $\lambda_{\min}\left(-n \cdot \mathcal{H}_{n,\gamma\gamma}(\theta, \gamma)\right) > c_\gamma$ and (ii) $\lambda_{\min}\left(-\mathcal{I}_{n,\theta\theta}(\theta, \gamma)\right) > c_\theta$ with probability approaching one (w.p.a.1) uniformly on $\{\Theta \times \mathbb{C}_n : ||\theta - \theta_0|| \leq \varepsilon, \, ||\gamma - \gamma_0||_\infty \leq \varepsilon\}$.

Assumptions 3.2 and 3.3 should be fairly reasonable in practice. Then, under these additional assumptions, we can derive the ℓ_1-norm and max-norm convergence rates for $\widehat{\gamma}_n$, as shown in the next theorem.

Theorem 3.2. Suppose that Assumptions 2.1–2.3 and 3.1–3.3 hold. Then, we have (i) $\frac{1}{n} \sum_{i=1}^{n} |\widehat{A}_{n,i} - A_{0,i}| = O_P(n^{-1/2})$, (ii) $\frac{1}{n} \sum_{i=1}^{n} |\widehat{B}_{n,i} - B_{0,i}| = O_P(n^{-1/2})$, and (iii) $||\widehat{\gamma}_n - \gamma_0||_\infty = O_P(\sqrt{\ln n/n})$.

The max-norm convergence rate obtained in Theorem 3.2 is consistent with the result of Theorem 3 in Graham (2017) and that of Theorem 3.1 in Yan et al. (2019).

3.2 Binary segmentation algorithm

Given the consistent estimates of $\mathbf{A}_0$ and $\mathbf{B}_0$, we use the BS algorithm to estimate the group structure. We first sort $\widehat{\mathbf{A}}_n$ and $\widehat{\mathbf{B}}_n$ in ascending order and write the order statistics as $\widehat{A}_{n,(1)} \leq \widehat{A}_{n,(2)} \leq \cdots \leq \widehat{A}_{n,(n)}$, and $\widehat{B}_{n,(1)} \leq \widehat{B}_{n,(2)} \leq \cdots \leq \widehat{B}_{n,(n)}$. In the following, we mainly describe the estimation of the group membership for the sender effects, $\mathcal{C}_0^A$. The exactly same procedure described below can be used to estimate $\mathcal{C}_0^B$.

A concept behind the BS algorithm is quite simple. If $\{A_{0,i}\}$ are heterogeneous across K^A latent groups but are homogeneous within the groups, there should exist $K^A - 1$ "break points" in the sorted $\{A_{0,i}\}$. Since $\widehat{\mathbf{A}}_n$ is uniformly consistent for $\mathbf{A}_0$, these break points appear also in the following sequence: $\widehat{A}_{n,(1)}, \ldots, \widehat{A}_{n,(n)}$ w.p.a.1. For $1 \leq i < j \leq n$, we define $\widehat{\Delta}^A(i,j)$ as the sum of squared variations over $\{\widehat{A}_{n,(i)}, \ldots, \widehat{A}_{n,(j)}\}$:

$$\widehat{\Delta}^A(i,j) \equiv \sum_{l=i}^{j} (\widehat{A}_{n,(l)} - \bar{A}_{n,i,j})^2, \quad \text{where } \bar{A}_{n,i,j} \equiv \frac{1}{j-i+1} \sum_{l=i}^{j} \widehat{A}_{n,(l)}$$

Further, we define

$$\widehat{S}_{i,j}^A(\kappa) \equiv \begin{cases} \frac{1}{j-i+1} \left(\widehat{\Delta}^A(i,\kappa) + \widehat{\Delta}^A(\kappa+1,j) \right) & \text{if } j < \kappa \\ \frac{1}{j-i+1} \widehat{\Delta}^A(i,j) & \text{if } j = \kappa \end{cases}$$

That is, $\widehat{S}_{i,j}^A(\kappa)$ provides the total variance of $\{\widehat{A}_{n,(i)}, \ldots, \widehat{A}_{n,(j)}\}$ when a break point is placed at κ. Assuming that $K^A \geq 2$, the BS algorithm proceeds as follows:

Step 1 ($K^A = 2$): We find the first break point, say $\widehat{t}_1$, by

$$\widehat{t}_1 = \underset{1 \leq \kappa < n}{\operatorname{argmin}} \widehat{S}_{1,n}^A(\kappa)$$

Then, we can partition $\{\widehat{A}_{n,(i)}\}$ into the following two subsets: $\{\widehat{A}_{n,(i)}\} = \{\widehat{A}_{n,(i)}\}_{i=1}^{\widehat{t}_1} \bigcup \{\widehat{A}_{n,(i)}\}_{i=\widehat{t}_1+1}^{n}$. If $K^A = 2$, assuming that $a_{0,1} < a_{0,2}$ without loss of generality, we obtain $\widehat{\mathcal{C}}_{n,1}^A \equiv \{i : \widehat{A}_{n,(1)} \leq \widehat{A}_{n,i} \leq \widehat{A}_{n,(\widehat{t}_1)}\}$ and $\widehat{\mathcal{C}}_{n,2}^A \equiv \{i : \widehat{A}_{n,(\widehat{t}_1+1)} \leq \widehat{A}_{n,i} \leq \widehat{A}_{n,(n)}\}$ as the estimators of $\mathcal{C}_{0,1}^A$ and $\mathcal{C}_{0,2}^A$, respectively, and the algorithm stops. If $K^A > 2$, we proceed to the next step.

Step 2 ($K^A = 3$): Now, if $K^A = 3$, there exists one more break point either in $\{\widehat{A}_{n,(i)}\}_{i=1}^{\widehat{t}_1}$ or in $\{\widehat{A}_{n,(i)}\}_{i=\widehat{t}_1+1}^{n}$. Either one of the two converges to a sequence of identical constants as the sample size increases. Therefore, if $\widehat{S}_{1,\widehat{t}_1}^A(\widehat{t}_1) > \widehat{S}_{\widehat{t}_1+1,n}^A(n)$ for example, the break point is likely to lie in the former subset. Thus, the second break point, say $\widehat{t}_2$, can be found by

$$\widehat{t}_2 = \begin{cases} \operatorname{argmin}_{1 \leq \kappa < \widehat{t}_1} \widehat{S}_{1,\widehat{t}_1}^A(\kappa) & \text{if } \widehat{S}_{1,\widehat{t}_1}^A(\widehat{t}_1) > \widehat{S}_{\widehat{t}_1+1,n}^A(n) \\ \operatorname{argmin}_{\widehat{t}_1+1 \leq \kappa < n} \widehat{S}_{\widehat{t}_1+1,n}^A(\kappa) & \text{if } \widehat{S}_{1,\widehat{t}_1}^A(\widehat{t}_1) \leq \widehat{S}_{\widehat{t}_1+1,n}^A(n) \end{cases}$$

Then, we add $\widehat{t}_2$ to the set of break points, and (with a little abuse of notation) we sort and re-label the points to ensure that $\widehat{t}_1 < \widehat{t}_2$. The resulting partition is $\{\widehat{A}_{n,(i)}\} = \{\widehat{A}_{n,(i)}\}_{i=1}^{\widehat{t}_1} \bigcup \{\widehat{A}_{n,(i)}\}_{i=\widehat{t}_1+1}^{\widehat{t}_2} \bigcup \{\widehat{A}_{n,(i)}\}_{i=\widehat{t}_2+1}^{n}$. Setting $a_{0,1} < a_{0,2} < a_{0,3}$ without loss of generality, we can define $\widehat{\mathcal{C}}_{n,1}^{A} \equiv \{i : \widehat{A}_{n,(1)} \le \widehat{A}_{n,i} \le \widehat{A}_{n,(\widehat{t}_1)}\}$, $\widehat{\mathcal{C}}_{n,2}^{A} \equiv \{i : \widehat{A}_{n,(\widehat{t}_1+1)} \le \widehat{A}_{n,i} \le \widehat{A}_{n,(\widehat{t}_2)}\}$, and $\widehat{\mathcal{C}}_{n,3}^{A} \equiv \{i : \widehat{A}_{n,(\widehat{t}_2+1)} \le \widehat{A}_{n,i} \le \widehat{A}_{n,(n)}\}$ as the estimators of $\mathcal{C}_{0,1}^{A}$, $\mathcal{C}_{0,2}^{A}$, and $\mathcal{C}_{0,3}^{A}$, respectively.

Step 3 ($K^A \ge 4$): When $K^A = 4$, if $\max\{\widehat{S}_{1,\widehat{t}_1}^{A}(\widehat{t}_1), \widehat{S}_{\widehat{t}_1+1,\widehat{t}_2}^{A}(\widehat{t}_2), \widehat{S}_{\widehat{t}_2+1,n}^{A}(n)\} = \widehat{S}_{1,\widehat{t}_1}^{A}(\widehat{t}_1)$ for example, the third break point should exist in $\{\widehat{A}_{n,(i)}\}_{i=1}^{\widehat{t}_1}$. Then, following the same procedure as above, we can obtain $\widehat{t}_3 = \operatorname{argmin}_{1 \le \kappa < \widehat{t}_1} \widehat{S}_{1,\widehat{t}_1}^{A}(\kappa)$. We repeat these steps until K^A groups are detected with $K^A - 1$ break points. Finally, letting $\widehat{t}_0 = 0$ and $\widehat{t}_{K^A} = n$ so that $\widehat{t}_0 < \widehat{t}_1 < \cdots < \widehat{t}_{K^A-1} < \widehat{t}_{K^A}$, each k-th group can be estimated by $\widehat{\mathcal{C}}_{n,k}^{A} \equiv \{i : \widehat{A}_{n,(\widehat{t}_{k-1}+1)} \le \widehat{A}_{n,i} \le \widehat{A}_{n,(\widehat{t}_k)}\}$. Then, $\widehat{\mathcal{C}}_n^A \equiv \{\widehat{\mathcal{C}}_{n,1}^{A}, \ldots, \widehat{\mathcal{C}}_{n,K^A}^{A}\}$ is our estimator for $\mathcal{C}_0^A$.

In the same manner as above, we define $\widehat{\mathcal{C}}_n^B \equiv \{\widehat{\mathcal{C}}_{n,1}^{B}, \ldots, \widehat{\mathcal{C}}_{n,K^B}^{B}\}$ for the estimation of $\mathcal{C}_0^B$. Note that the identification of group membership can be achieved only up to "label swapping." Thus, without loss of generality, we can label the groups according to the values of the group effects such that $a_{0,1} < a_{0,2} < \cdots < a_{0,K^A}$; that is, we define the k-th group as the group with the k-th smallest sender effect. Similarly, we set $b_{0,1} < b_{0,2} < \cdots < b_{0,K^B}$. To investigate the asymptotic properties of the BS algorithm, we introduce the following assumption.

Assumption 3.4. (i) There exist $c_A, c_B > 0$ such that $\min_{k \ne k'} |a_{0,k} - a_{0,k'}| > c_A$ and $\min_{k \ne k'} |b_{0,k} - b_{0,k'}| > c_B$. (ii) $|\mathcal{C}_{0,k}^{A}|/n \to \tau_k^A \in (0,1)$ for all $k = 1, \ldots, K^A$ and $|\mathcal{C}_{0,k}^{B}|/n \to \tau_k^B \in (0,1)$ for all $k = 1, \ldots, K^B$.

Assumption 3.4 is parallel to Assumption A2 in Wang & Su (2021). Similar assumptions are commonly used in the literature of panel data models with latent group structure. The following theorem provides the consistency result of the BS algorithm:

Theorem 3.3. Suppose that Assumptions 2.1–2.3 and 3.1–3.4 hold. Then, we have $\Pr(\widehat{\mathcal{C}}_n^A = \mathcal{C}_0^A) \to 1$ and $\Pr(\widehat{\mathcal{C}}_n^B = \mathcal{C}_0^B) \to 1$.

Although the proof of Theorem 3.3 is almost analogous to Ke et al. (2016) and Wang & Su (2021), for completeness, we provide it in Appendix B.

Remark 3.1 (Fine-tuning). Bai (1997) shows that the BS method tends to over/underestimate the location of break points depending on the share of each group and the gaps between the values of the group effects. To account for this problem, he proposes the following repartitioning method. Let $\{\widehat{t}_0, \ldots, \widehat{t}_{K^A}\}$ be the estimated break points obtained by the standard BS method. Then, we replace the initial estimate $\widehat{t}_k$ with $\widehat{t}_k^{\text{repart}} \equiv \operatorname{argmin}_{\widehat{t}_{k-1}+1 \le \kappa < \widehat{t}_{k+1}} \widehat{S}_{\widehat{t}_{k-1}+1,\widehat{t}_{k+1}}^{A}(\kappa)$ for all $k = 1, \ldots, K^A - 1$. Letting $\widehat{\mathcal{C}}_n^{A,\text{repart}}$ be the resulting repartitioned estimator of $\mathcal{C}_0^A$, given the result of Theorem 3.3, it is straightforward to see that $\widehat{\mathcal{C}}_n^{A,\text{repart}}$ is also consistent for $\mathcal{C}_0^A$. The above repartitioning procedure can be implemented recursively until convergence. The numerical simulations in Section 4 indicate that using this repartitioning method improves the probability of correctly predicting the group memberships.

Remark 3.2 (Other grouping techniques). Other than the BS algorithm, we can consider using alternative grouping methods such as the k-means and the C-Lasso (classifier-Lasso) developed by Su et al. (2016). As discussed in Wang & Su (2021), the main advantage of the BS algorithm over

these alternatives is its computational and interpretational simplicity. The k-means algorithm is NP-hard, and C-Lasso requires solving a large nonlinear optimization problem, whereas our BS method does not involve any computational optimization.[6] Additionally, the C-Lasso may end up leaving some unclassified agents. Therefore, we need to classify such observations afterward based on some criterion. Another possibility is to use finite-mixture modeling, which is more common and traditional in statistical literature. However, the mixture-based approach is more suitable for cross-sectional data. It only estimates the "share" of each group, requiring an additional classification step to estimate the group membership of each agent, as in C-Lasso. Additionally, when (K^A, K^B) is large, the mixture approach may be less practical because the marginal probability of each dyad is represented as a mixture of $(K^A K^B)^2$ conditional probabilities. As mentioned in the introduction, if the group structure is not of interest, the standard fixed-effect bias correction method should be also considered. For more details about these approaches, including some simulation results, see Appendices C.3 and C.4.

3.3 Post-grouping ML estimator

In the final step, we solve (4) approximately by using $(\widehat{\mathcal{C}}_n^A, \widehat{\mathcal{C}}_n^B)$ in the place of $(\mathcal{C}_0^A, \mathcal{C}_0^B)$. Recalling that a_1 is pinned at $a_1 = 0$, let $\delta \equiv (\theta^\top, a_2, \ldots, a_{K^A}, b_1, \ldots, b_{K^B})^\top$, and $\mathbb{D} \equiv \Theta \times \mathbb{A}^{K^A-1} \times \mathbb{B}^{K^B}$ for the parameter space of δ. We denote δ_0 as the true value of δ. Then, our final ML estimator for δ_0 is defined as

$$\widehat{\delta}_n = \underset{\delta \in \mathbb{D}}{\mathrm{argmax}} \; \widehat{\mathcal{L}}_n(\delta)$$

where $\widehat{\mathcal{L}}_n(\delta) \equiv \mathcal{L}_n\big(\theta, \big\{\sum_{k=1}^{K^A} a_k \cdot \mathbf{1}\{i \in \widehat{\mathcal{C}}_{n,k}^A\}\big\}, \big\{\sum_{k=1}^{K^B} b_k \cdot \mathbf{1}\{i \in \widehat{\mathcal{C}}_{n,k}^B\}\big\}\big)$. Similarly, we define

$$\widehat{\delta}_n^{\mathrm{oracle}} = \underset{\delta \in \mathbb{D}}{\mathrm{argmax}} \; \mathcal{L}_n(\delta)$$

where $\mathcal{L}_n(\delta) \equiv \mathcal{L}_n\big(\theta, \big\{\sum_{k=1}^{K^A} a_k \cdot \mathbf{1}\{i \in \mathcal{C}_{0,k}^A\}\big\}, \big\{\sum_{k=1}^{K^B} b_k \cdot \mathbf{1}\{i \in \mathcal{C}_{0,k}^B\}\big\}\big)$; that is, $\widehat{\delta}_n^{\mathrm{oracle}}$ is the "oracle" estimator that is computed based on the true $\mathcal{C}_0^A$ and $\mathcal{C}_0^B$. Since $\widehat{\delta}_n^{\mathrm{oracle}}$ is a standard parametric ML estimator, the estimator follows a normal distribution asymptotically with its asymptotic covariance matrix given by the inverse Fisher Information matrix. Meanwhile, we have shown in Theorem 3.3 that the estimated group memberships $(\widehat{\mathcal{C}}_n^A, \widehat{\mathcal{C}}_n^B)$ are equal to $(\mathcal{C}_0^A, \mathcal{C}_0^B)$ w.p.a.1. Therefore, the final ML estimator $\widehat{\delta}_n$ has asymptotically the same statistical performance as the oracle estimator $\widehat{\delta}_n^{\mathrm{oracle}}$. We formally state this result in the next theorem.

Theorem 3.4. Suppose that Assumptions 2.1–2.3 and 3.1–3.4 hold. In addition, we assume that $\mathcal{I}_{\delta\delta} \equiv -\lim_{n \to \infty} \mathbb{E}\big[\partial^2 \mathcal{L}_n(\delta_0)/(\partial\delta\partial\delta^\top)\big]$ exists and is positive definite. Then, $\widehat{\delta}_n$ and $\widehat{\delta}_n^{\mathrm{oracle}}$ have the same asymptotic distribution: $\sqrt{\frac{N}{2}}(\widehat{\delta}_n^{\mathrm{oracle}} - \delta_0) \xrightarrow{d} N(\mathbf{0}_{d_z+K^A+K^B+1}, \mathcal{I}_{\delta\delta}^{-1})$.

Recall that the asymptotic equivalence between $\widehat{\delta}_n$ and $\widehat{\delta}_n^{\mathrm{oracle}}$ relies on the dense network structure where each agent's specific effects can be point-identified. When the networks are not dense, $\widehat{\delta}_n$ is generally inconsistent, while $\widehat{\delta}_n^{\mathrm{oracle}}$ may be still consistent (potentially with a slower convergence rate). Finally, note that the above discussions hold true if the repartitioned estimator $(\widehat{\mathcal{C}}_n^{A,\mathrm{repart}}, \widehat{\mathcal{C}}_n^{B,\mathrm{repart}})$ is used instead of $(\widehat{\mathcal{C}}_n^A, \widehat{\mathcal{C}}_n^B)$.

4. Monte Carlo experiments

In this section, we examine the finite sample performance of the three-step ML estimator. We consider the following three data-generating processes (DGPs) for Monte Carlo experiments:

$$u_{i,j}(q) = Z_{i,j,1}\beta_{0,1} + Z_{i,j,2}\beta_{0,2} + \alpha_0 q + A_{0,i} + B_{0,j} - \epsilon_{i,j}, \text{ for } i \neq j$$

where

$$\text{DGP 1}: A_{0,i} = \sum_{k=1}^{K^A} a_{0,k} \cdot \mathbf{1}\{i \in \mathcal{C}_{0,k}^A\} \text{ and } B_{0,j} = \sum_{k=1}^{K^B} b_{0,k} \cdot \mathbf{1}\{j \in \mathcal{C}_{0,k}^B\} \text{ with } K^A = K^B = 2$$

$$\text{DGP 2}: A_{0,i} = \sum_{k=1}^{K^A} a_{0,k} \cdot \mathbf{1}\{i \in \mathcal{C}_{0,k}^A\} \text{ and } B_{0,j} = \sum_{k=1}^{K^B} b_{0,k} \cdot \mathbf{1}\{j \in \mathcal{C}_{0,k}^B\} \text{ with } K^A = K^B = 3$$

$$\text{DGP 3}: A_{0,i} \overset{i.i.d.}{\sim} \text{Uniform}[-1.4, 1.4] \ (A_{0,1} = 0) \text{ and } B_{0,j} \overset{i.i.d.}{\sim} \text{Uniform}[-1.4, 1.4]$$

DGP 3 is examined to investigate the performance of the proposed estimator under model mis-specification. For all GDPs, $Z_{i,j,1} = |X_i - X_j|$ with $X_i \overset{i.i.d.}{\sim} N(0, 0.6^2)$, $Z_{i,j,2} \overset{i.i.d.}{\sim} N(0, 1)$, $(\epsilon_{i,j}, \epsilon_{j,i})$ is i.i.d. across dyads as the standard bivariate normal with correlation $\rho_0 = 0.6$, and $(\alpha_0, \beta_{0,1}, \beta_{0,2}) = (0.6, -1, 1.5)$. The group heterogeneity parameters are set as follows: for DGP 1, $(a_{0,1}, a_{0,2}) = (0, r)$ and $(b_{0,1}, b_{0,2}) = -0.7 + (0, r)$, and for DGP 2 $(a_{0,1}, a_{0,2}, a_{0,3}) = (0, r, 2r)$, and $(b_{0,1}, b_{0,2}, b_{0,3}) = -0.7 + (-r, 0, r)$, where $r \in \{0.4, 0.8\}$. The smaller r becomes, the more difficult it is to identify the group structure. The group memberships are determined randomly while maintaining the equal size of each group. Exceptionally, for observation 1, $1 \in \mathcal{C}_{0,1}^A$ is fixed throughout the experiments. For each model setup, we consider two sample sizes: $n \in \{54, 72\}$. The number of Monte Carlo repetitions is set to 500 for each experiment.

Regarding DGPs 1 and 2, we compare the performances of the following six types of estimators: (1) the initial ML estimator (i.e., (5)), (2) the three-step ML estimator based on the BS method with correctly chosen (K^A, K^B), (3) the BS-based estimator with $K^A = K^B = 4$, (4) the three-step ML estimator where the k-means method is employed in the second step with correctly chosen (K^A, K^B), (5) the k-means-based estimator with $K^A = K^B = 4$, and (6) the oracle estimator based on the true group membership. For DGP 3, we compare estimators (1), (3), and (5) and also evaluate the BS- and k-means-based estimators with $K^A = K^B = 2$. For the BS-based estimator, we try both estimators with two iterations of fine-tuning and without repartitions. Note that when there are only two groups, the repartitioned BS method and the standard BS method yield the same grouping result.

To save space, we report the simulation results for DGPs 1, 2, and 3 in Tables C.1, C.2, and C.3 in Appendix C.2, in which we present the bias and root mean square error (RMSE) for estimating the common parameters $(\alpha_0, \beta_{0,1}, \beta_{0,2}, \rho_0)$ for the estimators listed above. Further, the results of the estimation of the group-specific effects in DGPs 1 and 2 are summarized in Table C.4. The main findings from these tables are as follows. We first focus on the results for DGPs 1 and 2. First, we can observe that the initial ML estimator is largely biased for the estimation of $(\beta_{0,1}, \beta_{0,2})$ owing to the incidental parameter problem. Although the three-step estimators also have some biases in the estimation of $(\beta_{0,1}, \beta_{0,2})$ when $r = 0.4$ and $n = 54$, the biases can be mitigated by increasing either r or n. Thus, these biases are probably due to the frequent misclassification of group memberships under small r and n. Probably for some DGP-specific reasons, the estimation of (α_0, ρ_0) is not biased for all estimators. Second, comparing the grouping methods, the results for DGP 2 show that although not always, repartitioning seems to be beneficial for the estimation of (α_0, ρ_0) and the group fixed-effect parameters in RMSE. Overall, the performance of the k-means-based estimator is very similar to that of the repartitioned BS estimator. Third, for both BS- and k-means-based estimators, misusing oversized group numbers tends to result in negative impacts in terms of RMSE. The impacts are relatively small in DGP 2 than DGP 1, as expected. Finally, although we can observe certain gaps between the oracle estimator and the three-step estimators in RMSE, the gaps can be reduced by increasing n, which is consistent with our theory.

Table 1. Distribution of $\{(g_{i,j}, g_{j,i}):1 \le i < j \le n\}$

	$g_{j,i} = 0$	$g_{j,i} = 1$
$g_{i,j} = 0$	495	343
$g_{i,j} = 1$	349	409

We move on to the results for DGP 3. First, when comparing the initial ML estimator and the three-step estimators, we can find that the incidental parameter bias is a more serious issue than the misspecification bias for the estimation of the slope parameters $(\beta_{0,1}, \beta_{0,2})$. Consequently, the three-step estimators tend to achieve smaller RMSE for $(\beta_{0,1}, \beta_{0,2})$ than the initial ML estimator despite the model misspecification. Next, comparing the two-group and the four-group estimators, the latter are more accurate in terms of bias (except for the estimation of α_0 with $n = 54$), which should be an understandable result. Which estimator has a smaller RMSE depends on the magnitude of bias improvement for each parameter. Finally, particularly in the estimation of the strategic effect α_0, the three-step estimators tend to have relatively large biases and RMSE compared with the initial ML estimator.

Table C.5 summarizes the simulation results for estimating the group memberships. Here, we compare the BS-based estimators with and without repartitions and the k-means-based estimator in terms of the ratio of correct group classification. First, the results indicate that the repartitioned BS method clearly outperforms the standard BS method without partitions. The performance of the k-means method is similar to that of the repartitioned BS method. As expected, as r decreases, predicting group membership correctly becomes significantly more difficult. If the gaps between the values of the group effects are sufficiently large and the sample size is not small, the estimators can attain more than 80% of correct classification even in DGP 2.[7] No clear difference can be observed between the estimation of $\mathcal{C}_0^A$ and that of $\mathcal{C}_0^B$.

5. Empirical application to international visa-free travels

As an empirical application of our model and method, we analyze the network of international visa-free travels. The dependent variable of interest is $G_n = (g_{i,j})_{1 \le i,j \le n}$, where $g_{i,j} = 1$ if country i allows the citizens in country j to visit i without visas, and $g_{i,j} = 0$ otherwise. Since the bilateral relationship about visa-free policy is expected to be complementary, this would fit into our model framework.[8]

In this empirical study, we consider 57 countries selected mainly from Asia, the Middle East, the former USSR, and Oceania.[9] The information about the visa policy of each country is taken from *Henly and Partners: Passport Index 2020* (https://www.henleypassportindex.com/passport).[10] The total number of dyads in this network is $57(57 - 1)/2 = 1,596$. From Table 1, which summarizes the distribution of the link connections, we can observe that the number of country pairs with one-way links is smaller than that with mutual links or no links. This would suggest the presence of complementarity in the network formation process. According to the above-mentioned passport index, Japan, Singapore, and South Korea are the top three countries among the 57 countries in terms of the number of all countries with visa-free access. For our restricted sample network, South Korea has the largest in-degree $\sum_i g_{i,\text{South Korea}} = 49$. For the out-degree, Nepal has the largest value $\sum_j g_{\text{Nepal},j} = 55$; that is, Nepal allows 55 countries (out of 57) to visit Nepal only with on-arrival visas. More detailed information can be found in Table C.8 in Appendix C.5.

The network for all the 57 countries is quite complicated and difficult to grasp the entire picture. As one illustration of our data, Figure 2 presents the sub-network obtained by restricting the vertices to the Eastern and Southeastern Asian countries. The left panel in the figure shows the whole shape of this sub-network. (Note that the direction of the arrows in the figure is "not" the
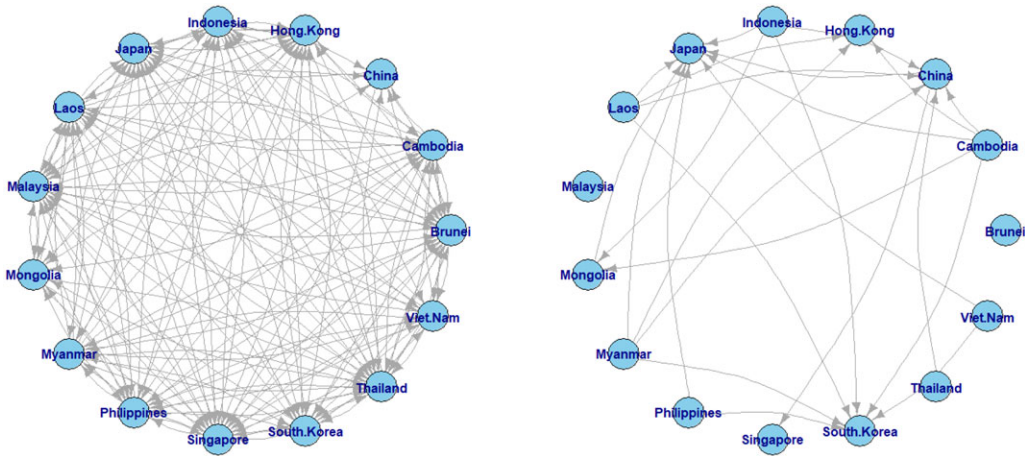
Figure 2. Eastern and Southeastern Asian sub-network. (Left panel: the whole sub-network, right panel: only one-way links.)

direction of visa-free access, but it represents that the target country is allowed to visit the country at the arrow's origin without visas.) The right panel shows the network created by leaving only one-way links from the left one; in other words, this is $(g_{i,j} \cdot (1 - g_{j,i}))_{i,j \in \text{Brunei},\dots,\text{Viet Nam}}$. From this figure, we can expect the existence of a certain level of degree heterogeneity. More specifically, Cambodia, for example, has five outgoing one-way links in this sub-network, suggesting that this country would have a larger sender effect. In contrast, countries such as Japan and South Korea would exhibit a larger receiver effect.

For estimating the network formation model, we consider five covariates; for their definitions, see Table 2. The summary statistics of the covariates are provided in Table C.9. With these variables, we consider the following payoff function:

$$u_{i,j}(q) = (\ln gdp\_pc_i)(\ln gdp\_pc_j)\beta_{0,1} + |free_i - free_j|\beta_{0,2} + \mathbf{1}\{region_i = region_j\}\beta_{0,3}$$
$$+ \ln(export_{ij} + 1)\beta_{0,4} + \ln(import_{ij} + 1)\beta_{0,5} + \alpha_0 q + A_{0,i} + B_{0,j} - \epsilon_{i,j}, \text{ for } i \neq j$$

where we assume that $(\epsilon_{i,j}, \epsilon_{j,i})$ have the standard bivariate normal distribution. To estimate our network formation model, we first need to determine the number of groups K^A and K^B. Then, following Ke et al. (2016) and Wang & Su (2021), the optimal (K^A, K^B) is selected as the minimizer of the BIC criterion: $-2\widehat{\mathcal{L}}_n(\widehat{\delta}_n) + (6 + K^A + K^B)\ln(1596)$. Then, as a result of searching over the models with $(K^A, K^B) \in \{2, \dots, 7\}^2$, we find that the model with $(K^A, K^B) = (7, 6)$ achieves the smallest BIC (see Table C.10), and this is the model reported here. We estimate not only our proposed model, which we call the grouped heterogeneity model, but also a model without strategic interaction and group heterogeneity as a benchmark. Although the benchmark model is a "complete" model that does not exhibit the multiple equilibria, we use the same log-likelihood function as in (2) for the comparison purpose.[11] For the estimation of the group memberships, we employ the repartitioning method with two iterations.

The estimation results are summarized in Table 3. First of all, as expected, our proposed model suggests that there is a significant strategic complementarity in the network formation behavior. We can also find a certain level of degree heterogeneity in terms of both the sender and the receiver effects. Comparing the grouped heterogeneity model and the benchmark model, the log-likelihood value for the former is apparently significantly larger than that for the latter. This large difference in the degree of model fitting also demonstrates the significance of strategic effect and unobserved heterogeneity (note however that these models are not nested). For specific parameter estimates, we can observe several non-negligible differences between the two

Table 2. Definitions of variables

Variables	Definitions
gdp_pc_i	GDP per capita in 2018 (1,000 USD)
$free_i$	Freedom rating (1 = Most Free, 7 = Least Free)*
$region_i$	Categorical variable: East Asia, Southeast Asia, Central Asia, Europe, Middle East, or Oceania.
$export_{ij}$	Total export value from country i to j in 2018 (million USD)[†]
$import_{ij}$	Total import value from country i to j in 2018 (million USD)[†]

Sources: *_Freedom in the World 2018_, Freedom House (`https://freedomhouse.org/`); [†]IMF DATA (`https://data.imf.org/`).

Table 3. Estimation results

	Grouped heterogeneity model		Benchmark model	
	Estimate	*t*-value	Estimate	*t*-value
Intercept			−0.208	−1.196
Strategic effect: α	0.482	3.842		
$(\ln gdp\_pc_i)(\ln gdp\_pc_i)$: β_1	0.063	4.234	0.047	2.366
$\|free_i - free_j\|$: β_2	−0.229	−7.044	−0.075	−1.692
$\mathbf{1}\{region_i = region_j\}$: β_3	1.402	9.496	0.691	4.400
$\ln(export_{ij} + 1)$: β_4	0.046	1.829	−0.073	−3.604
$\ln(import_{ij} + 1)$: β_5	0.056	2.365	0.137	6.535
Error correlation: ρ	−0.080	−0.450	0.088	1.656
Sender effects: A				
Group 1: a_1	0	–		
Group 2: a_2	1.060	4.552		
Group 3: a_3	2.051	10.199		
Group 4: a_4	2.846	13.572		
Group 5: a_5	3.813	16.302		
Group 6: a_6	4.790	16.642		
Group 7: a_7	6.105	15.387		
Receiver effects: B				
Group 1: b_1	−5.670	−14.388		
Group 2: b_2	−4.590	−16.159		
Group 3: b_3	−3.902	−14.559		
Group 4: b_4	−3.446	−14.240		
Group 5: b_5	−2.561	−10.952		
Group 6: b_6	−1.558	−6.722		
Log-likelihood	−810.798		−1, 531.496	
# Observations	1,596		1,596	

models. For example, the effect of the export amount is predicted to be positive in the grouped heterogeneity model, whereas the benchmark model predicts a significantly negative impact. The error correlation parameter is not significantly different from zero in our model but is weakly positively significant in the benchmark model. This result would be understandable since the

Table 4. Estimated group memberships

	Sender effect: *A*	Receiver effect: *B*
Group 1	Australia, China, Iraq, Nauru, Oman, Russia	Iraq, Pakistan
Group 2	Bhutan, India, Japan, New Zealand, UAE	Bangladesh, Iran, Jordan, Lebanon, Nepal, Sri Lanka, Yemen
Group 3	Bahrain, Cyprus, Estonia, Georgia, Kiribati, Kuwait, Latvia, Lithuania, Mongolia, Myanmar, Papua New Guinea, Qatar, Saudi Arabia, South Korea, Tonga, Turkey, Ukraine, Viet Nam, Yemen	Armenia, Cambodia, Laos, Myanmar, Qatar, Russia, Viet Nam
Group 4	Azerbaijan, Belarus, Brunei, Fiji, Hong Kong, Israel, Kazakhstan, Kyrgyzstan, Moldova, Pakistan, Singapore, Thailand, Uzbekistan	Belarus, Bhutan, China, Fiji, Georgia, India, Indonesia, Kazakhstan, Kiribati, Kyrgyzstan, Moldova, Mongolia, Nauru, Papua New Guinea, Philippines, Tajikistan, Thailand, Tonga, Ukraine, Uzbekistan
Group 5	Armenia, Bangladesh, Jordan, Lebanon, Malaysia, Philippines, Tajikistan, Vanuatu	Azerbaijan, Bahrain, Kuwait, Latvia, Lithuania, Oman, Saudi Arabia, Turkey, UAE, Vanuatu
Group 6	Indonesia, Iran, Laos	Australia, Brunei, Cyprus, Estonia, Hong Kong, Israel, Japan, Malaysia, New Zealand, Singapore, South Korea
Group 7	Cambodia, Nepal, Sri Lanka	

benchmark model can account for the interdependence of the links only through the error correlation. Additionally, there are several interesting findings. For both models, if countries i and j are located in the same region, they become more likely to allow visa-free access, as expected. Not only in terms of geographical proximity, but we can also observe significant homophily in terms of the political system.

For the estimation results of country-specific effects, we report the estimated group memberships in Table 4. As expected from the above discussion, countries such as Cambodia are indeed classified into the highest group (i.e., Group 7) in terms of the sender effect. The other two countries that have Group-7 sender effect are Nepal and Sri Lanka. For the receiver effect, these two countries are classified as Group 2, and Cambodia is in Group 3. Overall, interestingly, there seems to be a weak negative correlation between the sender effects and the receiver effects. As expected from the above discussion, Japan and South Korea indeed belong to the group with the highest receiver effect (i.e., Group 6). The magnitudes of the receiver effects seem to roughly correlate with the size of the countries' economies (with some exceptions, such as China, India, and Russia).

6. Conclusion

This paper proposed a network formation model with pairwise strategic interaction and grouped degree heterogeneity. Assuming some parametric form for the error distribution, we proved that the model parameters can be identified under the availability of agent-specific covariates that have large supports and also have variations across all potential partners. For estimating the model, we proposed the three-step ML procedure: in the first step, the model is estimated without considering the group structure; subsequently, we estimate the group memberships using the BS algorithm given the estimates for the heterogeneity parameters obtained in the first step; and, finally, based on the estimated group memberships, we re-estimate the model. Under certain regularity conditions, we showed that the proposed estimator is asymptotically unbiased and distributed as normal at the parametric rate. An empirical application to international visa-free travel networks indicates the usefulness of the proposed model.

Several limitations and extensions are as follows. First, our approach can be used only in pairwise network formation games with no network externalities to/from the rest of the links, and

this limits the empirical applicability. Therefore, it would be worthwhile to extend our results to network formation models with general network externalities involving more than two agents. Second, our approach requires that the degree heterogeneity parameters have discrete support, although, in reality, it is possible that they are continuous. To address this issue, it is of interest to modify our model in a similar manner to Bonhomme et al. (2017) and investigate the three-step ML estimator in which K^A and K^B grow slowly to infinity. Third, as our model is a dyadic binary game model, where a pairwise network formation model is its special case, we can consider its ordered-response game version as a natural extension. For example, we might be interested in analyzing bilateral military relations: non-alliance, quasi-alliance, or alliance. We expect that such extension can be relatively easily achieved by adopting the ML estimator discussed in Aradillas-Lopez & Rosen (2019). We leave these topics for future research.

Notes

1 For recent developments regarding econometric approaches for analyzing network formation, we refer readers to, for example, Chandrasekhar (2016) and de Paula (2020).

2 Refer to de Paula (2013) for a comprehensive survey on this topic.

3 This assumption should be debatable. Rather than directly specifying the number of groups, in the literature on the BS algorithm for example, Ke et al. (2016) and Lian et al. (2021) propose introducing an additional threshold parameter to detect the group structure. However, since the number of groups can vary only discretely with the threshold value, introducing the threshold parameter is essentially equivalent to selecting the number of groups. For the use of the BS algorithm, Wang & Su (2021) formally show that Bayesian Information Criterion (BIC)-type criterion can consistently select the correct number of groups as the sample size increases.

4 Note that we can always estimate "pseudo-true" group memberships based on the maximum likelihood principle even when some elements of $\mathbf{A}_0$ and $\mathbf{B}_0$ are not point-identified. However, they may not necessarily coincide with the true group memberships in general. A more formal investigation on this issue is left as a future work. For a related discussion, see Bonhomme & Manresa (2015).

5 Graham (2017) and Jochmans (2018) developed conditional likelihood methods that can be used to estimate the homophily parameters (i.e., β_0 in our context), even when the networks are sparse. However, their approach cannot be directly applied to our case because of the interdependence between $g_{i,j}$ and $g_{j,i}$.

6 Having said that, the recent statistical softwares are often equipped with fast heuristic algorithms for solving the k-means problem. In fact, as shown in Section 4, a version of our three-step estimator that employs the k-means classification in the second stage works equivalently well to the BS algorithm with repartitions. This is not surprising because the BS method and the k-means are based on a very similar grouping principle.

7 One might view that the results reported in Table C.5 are not particularly good for the BS algorithm. The main reason for this would be that our model is a bivariate binary response model, whereas most of the previous studies using the BS algorithm focused on models with a continuous outcome.

8 As pointed out by a referee, the validity of characterizing our data as a set of equilibrium actions in a simultaneous-move game, rather than a dynamic game, should be debatable. Indeed, in the game econometrics literature, there has been some discussion on the interpretation of applying a static model to data where the actions actually take place at different points in time. For example, some researchers justify the use of a static game by regarding the observed actions as the realization of a "long-run" equilibrium (e.g., Ciliberto & Tamer, 2009; Aradillas-López & Gandhi, 2016). If this assumption is not realistic, a dynamic network formation model such as that in Graham (2016) would be more suitable.

9 The list of countries used in this empirical study is as follows: Armenia, Australia, Azerbaijan, Bahrain, Bangladesh, Belarus, Bhutan, Brunei, Cambodia, China, Cyprus, Estonia, Fiji, Georgia, Hong Kong, India, Indonesia, Iran, Iraq, Israel, Japan, Jordan, Kazakhstan, Kiribati, Kuwait, Kyrgyzstan, Laos, Latvia, Lebanon, Lithuania, Malaysia, Moldova, Mongolia, Myanmar, Nauru, Nepal, New Zealand, Oman, Pakistan, Papua New Guinea, Philippines, Qatar, Russia, Saudi Arabia, Singapore, South Korea, Sri Lanka, Tajikistan, Thailand, Tonga, Turkey, UAE, Ukraine, Uzbekistan, Vanuatu, Viet Nam, and Yemen. These countries are selected based on geographical proximity and ease of data collection.

10 Based on their definition, we categorize electronic travel authorization (eTA) and on-arrival visa as visa-free access.

11 We also estimated another benchmark model with the strategic interaction but no group heterogeneity. Consequently, the estimated interaction effect was almost zero and insignificant. However, once the group effects are introduced into the model,

the interaction effect becomes positively significant even when K^A and K^B are small (e.g., when both are two, $\widehat{\alpha}_n = 0.519$). This result might be interpreted as that since the strategic interaction parameter α_0 is a type of "location-shift" parameter, if a larger part of the variation in the payoff remains unaccounted for, it becomes more difficult to identify it from the payoff disturbances. This should provide another motivation to account for unobserved degree heterogeneity. These results are not included here to save space, but are available on request.

References

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2020). Sampling-based versus design-based uncertainty in regression analysis. *Econometrica*, *88*(1), 265–296.

Abbe, E. (2017). Community detection and stochastic block models: Recent developments. *The Journal of Machine Learning Research*, *18*(1), 6446–6531.

Amemiya, T. (1985). *Advanced econometrics*. Cambridge, Massachusetts: Harvard University Press.

Aradillas-López, A., & Gandhi, A. (2016). Estimation of games with ordered actions: An application to chain-store entry. *Quantitative Economics*, *7*(3), 727–780.

Aradillas-Lopez, A., & Rosen, A. M. (2019). Inference in ordered response games with complete information, *Working paper*.

Athey, S., Eckles, D., & Imbens, G. W. (2018). Exact p-values for network interference. *Journal of the American Statistical Association*, *113*(521), 230–240.

Bai, J. (1997). Estimating multiple breaks one at a time. *Econometric Theory*, *13*, 315–352.

Berry, S. T. (1992). Estimation of a model of entry in the airline industry. *Econometrica*, *60*(4), 889–917.

Bickel, P. J., & Chen, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proceedings of the National Academy of Sciences*, *106*(50), 21068–21073.

Bjorn, P. A., & Vuong, Q. H. (1984). Simultaneous equations models for dummy endogenous variables: A game theoretic formulation with an application to labor force participation, *Working paper*.

Bonhomme, S. (2020). Econometric analysis of bipartite networks. In B. S. Graham, & Á. de Paula (Eds.), *The econometric analysis of network data* (pp. 83–121). Elsevier.

Bonhomme, S., Lamadon, T., & Manresa, E. (2017). Discretizing unobserved heterogeneity, *University of Chicago, Working paper*.

Bonhomme, S., & Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, *83*(3), 1147–1184.

Bresnahan, T. F., & Reiss, P. C. (1990). Entry in monopoly market. *The Review of Economic Studies*, *57*(4), 531–553.

Chandrasekhar, A. G. (2016). Econometrics of network formation. In Y. Bramoullé, A. Galeotti, & B. Rogers (Eds.), *The Oxford handbook of the economics of networks* (pp. 303–357). Oxford University Press.

Chesher, A., & Rosen, A. M. (2020). Structural modeling of simultaneous discrete choice, *cemmap Working paper*, CWP9/20.

Ciliberto, F., & Tamer, E. (2009). Market structure and multiple equilibria in airline markets. *Econometrica*, *77*(6), 1791–1828.

de Paula, Á. (2013). Econometric analysis of games with multiple equilibria. *Annual Review of Economics*, *5*, 107–131.

de Paula, Á. (2020). Econometric models of network formation. *Annual Review of Economics*, *12*, 775–799.

Dzemski, A. (2019). An empirical model of dyadic link formation in a network with unobserved heterogeneity. *Review of Economics and Statistics*, *101*(5), 763–776.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, *486*(3–5), 75–174.

Graham, B. S. (2016). Homophily and transitivity in dynamic network formation, *NBER Working Paper*, (22186).

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

Graham, B. S., & Pelican, A. (2020). Testing for externalities in network formation using simulation. In B. S. Graham, & Á. de Paula (Eds.), *The econometric analysis of network data* (pp. 63–82). Elsevier.

Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical association*, *97*(460), 1090–1098.

Hoshino, T., & Yanagi, T. (2021). Treatment effect models with strategic interaction in treatment decisions, arXiv, (1810.08350).

Jochmans, K. (2018). Semiparametric analysis of network formation. *Journal of Business & Economic Statistics*, *36*(4), 705–713.

Karrer, B., & Newman, M. E. (2011). Stochastic blockmodels and community structure in networks. *Physical Review E*, *83*(1), 016107.

Ke, Y., Li, J., Zhang, W. (2016). Structure identification in panel data analysis. *The Annals of Statistics*, *44*(3), 1193–1233.

Khan, S., & Nekipelov, D. (2018). Information structure and statistical information in discrete response models. *Quantitative Economics*, *9*(2), 995–1017.

Kline, B. (2015). Identification of complete information games. *Journal of Econometrics*, *189*(1), 117–131.

Kline, B. (2016). The empirical content of games with bounded regressors. *Quantitative Economics*, *7*(1), 37–81.

Leung, M. P. (2015). Two-step estimation of network-formation models with incomplete information. *Journal of Econometrics*, *188*(1), 182–195.

Lewbel, A. (2007). Coherency and completeness of structural models containing a dummy endogenous variable. *International Economic Review*, *48*(4), 1379–1392.

Lian, H., Qiao, X., & Zhang, W. (2021). Homogeneity pursuit in single index models based panel data analysis. *Journal of Business & Economic Statistics*, *39*, 386–401.

Liu, R., Shang, Z., Zhang, Y., & Zhou, Q. (2020). Identification and estimation in panel models with overspecified number of groups. *Journal of Econometrics*, *215*(2), 574–590.

McKay, A., & Tekleselassie, T. G. (2018). Tall paper walls: The political economy of visas and cross-border travel. *The World Economy*, *41*(11), 2914–2933.

Mele, A. (2017). A structural model of dense network formation. *Econometrica*, *85*(3), 825–850.

Neiman, B., & Swagel, P. (2009). The impact of post-9/11 visa policies on travel to the united states. *Journal of International Economics*, *78*(1), 86–99.

Neumayer, E. (2010). Visa restrictions and bilateral travel. *The Professional Geographer*, *62*(2), 171–181.

Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, *69*(2), 026113.

Okui, R., & Wang, W. (2021). Heterogeneous structural breaks in panel data models. *Journal of Econometrics*, *220*, 447–473.

Pelican, A., & Graham, B. S. (2020). An optimal test for strategic interaction in social and economic network formation between heterogeneous agents, *NBER Working Paper*, 27793.

Rohe, K., Chatterjee, S., Yu, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, *39*(4), 1878–1915.

Rothenberg, T. J. (1971). Identification in parametric models. *Econometrica*, *39*, 577–591.

Sheng, S. (2020). A structural econometric analysis of network formation games through subnetworks. *Econometrica*, **88**(5), 1829–1858.

Su, L., Shi, Z., & Phillips, P. C. (2016). Identifying latent structures in panel data. *Econometrica*, **84**(6), 2215–2264.

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *The Review of Economic Studies*, *70*(1), 147–165.

Wang, W., Phillips, P. C., & Su, L. (2018). Homogeneity pursuit in panel data models: Theory and application. *Journal of Applied Econometrics*, *33*(6), 797–815.

Wang, W., & Su, L. (2021). Identifying latent group structures in nonlinear panels. *Journal of Econometrics*, *220*, 272–295.

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., Leng, C., & Zhu, J. (2016). Asymptotics in directed exponential random graph models with an increasing bi-degree sequence. *The Annals of Statistics*, *44*(1), 31–57.