SUPPLEMENT TO "A STRUCTURAL MODEL OF DENSE NETWORK FORMATION: APPENDICES C–E"

(*Econometrica*, Vol. 85, No. 3, May 2017, 825–850)

BY ANGELO MELE

APPENDIX C: UNOBSERVED HETEROGENEITY

IT IS POSSIBLE TO INCORPORATE unobserved heterogeneity or random coefficients in the model. However, this would significantly increase the computational cost of estimation. The simplest way to introduce unobserved heterogeneity is to model the preference shock ε_{ij} as incorporating individual random effects. The decision of the player to form a link is modified as follows:

$$U_i(g_{ij}=1, g_{-ij}, X) + \eta_i + \eta_j + \nu_{ij1} \geq U_i(g_{ij}=0, g_{-ij}, X) + \eta_i + \nu_{ij0}, \tag{39}$$

where ν_{ij} is an i.i.d. shock with logistic distribution and the vector $\eta = \{\eta_1, \ldots, \eta_n\}$ is drawn at time 0 from a known distribution $W(\eta)$. In this formulation, we assume that the players observe the random effect η but the econometrician does not. Notice that the random effect of player i cancels out, while the choice of linking j is conditional on the random effect of player j (which is present only when the link is formed).

Conditioning on the realization of the vector $\eta \in Y$, the potential function is modified as follows:

$$\mathcal{Q}(g, X, \theta; \eta) = Q(g, X, \theta) + \sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}\eta_j. \tag{40}$$

To compute the unconditional likelihood, we need to integrate out the unobserved vector η to obtain

$$\pi(g, X, \theta) = \int_Y \frac{\exp\big[\mathcal{Q}(g, X, \theta; \eta)\big]}{\sum_{\omega \in \mathcal{G}} \exp\big[\mathcal{Q}(\omega, X, \theta; \eta)\big]}\, dW(\eta). \tag{41}$$

The integral above can be computed using Monte Carlo techniques, as it is standard in the empirical industrial organization literature or labor economics. However, the model does not allow standard Monte Carlo, because of the normalizing constant.

A more feasible strategy is to use data augmentation and Markov chain Monte Carlo methods as in the discrete choice literature (Rossi, McCulloch, and Allenby (1996), Athey and Imbens (2007)). Conditioning on the realization of the unobserved component η, we can use the exchange algorithm to sample from the posterior distribution of θ. Conditioning on the proposed θ, we can use a Metropolis–Hastings step to sample the unobserved component η.

Given an initial (θ, η) at simulation s, we propose a new θ' and use the exchange algorithm to accept or reject the proposal. Given the new value of θ_{s+1}, we propose a new vector of unobserved components η' and accept using a Metropolis–Hastings step. The

 DOI: 10.3982/ECTA10400

probability of η, conditioning on (θ, g, X), is

$$\Pr(\eta|g, X, \theta) = \frac{W(\eta)\pi(g, X, \theta; \eta)}{\pi(g, X, \theta)}. \tag{42}$$

The Metropolis–Hastings step proceeds by proposing a new η' from a distribution $q_\eta(\eta'|\eta)$, which is accepted with probability

$$\alpha_\eta(\eta, \eta', g, \theta_s) = \left\{1, \frac{W(\eta')\pi(g, X, \theta; \eta')q_\eta(\eta|\eta')}{W(\eta)\pi(g, X, \theta; \eta)q_\eta(\eta'|\eta)}\right\}. \tag{43}$$

Similar ideas apply to random coefficients. However, as discussed in Graham (2016), when we observe only one network in the data, it is not possible to separately identify the linking externalities and the unobserved heterogeneity.

The main cost of these extensions is the increased computational burden, which may be substantial.

APPENDIX D: LARGE NETWORKS ANALYSIS AND CONVERGENCE

In this paper, we developed a network formation game model, which results in an equilibrium network similar to a directed ERGM. The probability of observing network g is given by (notice that $g_{ij} = 1$ does not imply $g_{ji} = 1$, because it is a directed network)

$$\pi_n(g) = \frac{\exp\left[\sum_{i=1}^n\sum_{j=1}^n g_{ij}u_{ij} + \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n g_{ij}g_{ji}m_{ij} + \sum_{i=1}^n\sum_{j=1}^n\sum_{k\neq i,j}^n g_{ij}g_{jk}v_{ik}\right]}{c(\mathcal{G}_n)},$$

where the functions $u_{ij} = u(X_i, X_j, \theta_u)$, $m_{ij} = m(X_i, X_j, \theta_m)$, and $v_{ik} = v(X_i, X_k, \theta_v)$ are function of vectors of covariates X_i's and parameters $\theta = (\theta_u, \theta_m, \theta_v)$. To simplify, we will assume that all these functions are constants, so that we do not consider the covariates. Hence, the probability of observing network g with parameters α, β, γ

$$\pi_n(g; \alpha, \beta, \gamma) = \frac{\exp\left[\alpha\sum_{i=1}^n\sum_{j=1}^n g_{ij} + \frac{\beta}{2}\sum_{i=1}^n\sum_{j=1}^n g_{ij}g_{ji} + \gamma^o\sum_{i=1}^n\sum_{j=1}^n\sum_{k\neq i}^n g_{ij}g_{jk}\right]}{c(\alpha, \beta, \gamma, \mathcal{G}_n)}.$$

To apply the analysis of Chatterjee and Diaconis (2013), we re-scale the terms as

$$\pi_n(g; \alpha, \beta, \gamma) = \frac{\exp\left\{n^2\left[\alpha\frac{\sum_{i=1}^n\sum_{j=1}^n g_{ij}}{n^2} + \frac{\beta}{2}\frac{\sum_{i=1}^n\sum_{j=1}^n g_{ij}g_{ji}}{n^2} + \gamma\frac{\sum_{i=1}^n\sum_{j=1}^n\sum_{k\neq i}^n g_{ij}g_{jk}}{n^3}\right]\right\}}{c(\alpha, \beta, \gamma, \mathcal{G}_n)}. \tag{44}$$

Notice that γ needs to be re-scaled (i.e., divided by n) when we run the simulations using the usual ERGM form, that is, $\gamma^o = \frac{\gamma}{n}$ for simulations using the `ergm` package in the software R.

In the formula above, the term $\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}}{n^2}$ is the directed edge density of the network, the term $\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}g_{ji}}{n^2}$ is the reciprocity density, while $\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k\neq i}^{n} g_{ij}g_{jk}}{n^3}$ is the density of directed two-paths (in our model, the latter is intepreted as popularity or indirect links effect).

In this appendix, we provide the technical results about the graph limits, large deviations, and mean-field approximations of the model. In the exposition for graph limits and large deviations, we report some results for undirected networks from Chatterjee and Varadhan (2011) and Chatterjee and Diaconis (2013), for completeness.

D.1. *A Crash Course on Graph Limits*

Most of this brief digression follows the overview in Chatterjee and Diaconis (2013), focusing on directed graphs. For a more detailed introduction to graph limits, see Lovasz (2012), Borgs et al. (2008), and Lovasz and Szegedy (2007). Most of the theory is developed for dense graphs, but there are several results for sparse graphs. The model presented here generates a dense graph; therefore, we present only the relevant theory.

Consider a sequence of simple directed graphs G_n, where the number of nodes n tends to infinity. Let $|\hom(H, G)|$ denote the number of homomorphisms of simple directed graph H into G. An homomorphism is an arc-preserving map from the set of vertices $V(H)$ of H to the set of vertices $V(G)$ of G.[49] For the graph limits we are interested in the *homomorphism densities* of the form

$$t(H, G) = \frac{\left|\hom(H, G)\right|}{\left|V(G)\right|^{|V(H)|}}.$$

Intuitively, $t(H, G)$ is the probability that a random mapping $V(H) \to V(G)$ is a homomorphism. We are interested in the behavior of $t(H, G_n)$ when $n \to \infty$. In particular, we want to characterize the limit object $t(H)$, for any simple graph H. The work of Lovasz (see Lovasz (2012) for an extensive overview) provides the limit object for this problem. Let $h \in \mathcal{W}$ be a function in the space $\mathcal{W}$ of all measurable functions $h : [0, 1]^2 \to [0, 1]$. This slightly differs from the original paper of Chatterjee and Diaconis (2013) because we are considering directed graphs, and therefore we do not require the function h to be symmetric. For comparison with the original formulation, let $\mathcal{W}_o$ denote the set of all measurable functions $h : [0, 1]^2 \to [0, 1]$ such that $h(x, y) = h(y, x)$.

The existence of such limit objects and the characterization for directed graphs is contained in Boeckner (2013) and extends the usual formulation for undirected graphs. If H is a simple directed graph with k vertices (i.e., $V(H) = \{1, 2, \ldots, k\}$), the limit object for $t(H, G_n)$ is

$$t(H, h) = \int_{[0,1]^k} \prod_{(i,j)\in E(H)} h(x_i, x_j)\, dx_1 \cdots dx_k,$$

[49] An important difference between homomorphisms for undirected graphs and directed graphs is that in the latter class of models, the existence of homomorphisms is not guaranteed. See Lovasz (2012) for some additional details.

where $E(H)$ is the set of directed edges of H. For example, if we are interested in homomorphisms of a directed edge, the homomorphism density is

$$t(H,G)=\frac{\big|\hom(H,G)\big|}{\big|V(G)\big|^{|V(H)|}}=\frac{\sum_i\sum_j g_{ij}}{n^2}$$

and the limit object is

$$t(H,h)=\int_{[0,1]^k}\prod_{(i,j)\in E(H)}h(x_i,x_j)\,dx_1\cdots dx_k=\int_0^1\int_0^1 h(x,y)\,dx\,dy.$$

If we are interested in the indirect links as in our model, we have

$$t(H,G)=\frac{\big|\hom(H,G)\big|}{\big|V(G)\big|^{|V(H)|}}=\frac{\sum_i\sum_j\sum_k g_{ij}g_{jk}}{n^3}$$

with limit object

$$t(H,h)=\int_{[0,1]^k}\prod_{(i,j)\in E(H)}h(x_i,x_j)\,dx_1\cdots dx_k=\int_0^1\int_0^1\int_0^1 h(x,y)h(y,z)\,dx\,dy\,dz.$$

A sequence of graphs $\{G_n\}_{n\geq 1}$ converges to h if, for every simple directed graph H,

$$\lim_{n\to\infty}t(H,G_n)=t(H,h).$$

The intuitive interpretation of this theory is simple: when n becomes large, we re-scale the vertices to a continuum interval $[0,1]$; and $h(x,y)$ is the probability that there is a directed edge from x to y. The limit object $h\in\mathcal{W}$ is called *graphon*. For any finite graph G with vertex set $\{1,\dots,n\}$, we can always define the graph limit representation f^G as

$$f^G(x,y)=\begin{cases}1 & \text{if } \big(\lceil nx\rceil,\lceil ny\rceil\big) \text{ is a directed edge of } G,\\ 0 & \text{otherwise}\end{cases}$$

where the symbol $\lceil a\rceil$ indicates the ceiling of a, that is, the smallest integer greater than or equal to a.

To study convergence in the space $\mathcal{W}$ of the functions h, we need to define a metric. We use the *cut distance*

$$d_\square(f,g)\equiv\sup_{S,T\subseteq[0,1]}\left|\int_{S\times T}\big[f(x,y)-g(x,y)\big]\,dx\,dy\right|,$$

where f and g are functions in $\mathcal{W}$. However, there is some nontrivial complication in the topology induced by the cut metric. To solve this complication, the usual approach is to work with a suitably defined quotient space $\widetilde{\mathcal{W}}$. We introduce an equivalence relation

in $\mathcal{W}$: $f \sim g$ if $f(x,y) = g_\sigma(x,y) = g(\sigma x, \sigma y)$ for some measure-preserving bijection $\sigma : [0,1] \to [0,1]$. We will use $\widetilde{h}$ to denote the equivalence class of h in $(\mathcal{W}, d_\square)$. Since $d_\square$ is invariant under σ, we can define a distance on the quotient space $\widetilde{\mathcal{W}}$ as

$$\delta_\square(\widetilde{f}, \widetilde{g}) \equiv \inf_\sigma d_\square(f, g_\sigma) = \inf_\sigma d_\square(f_\sigma, g) = \inf_{\sigma_1,\sigma_2} d_\square(f_{\sigma_1}, g_{\sigma_2}).$$

This makes $(\widetilde{\mathcal{W}}, \delta_\square)$ a metric space with several nice properties: it is compact and the homomorphism densities $t(H, h)$ are continuous functions on it. We associate f^G to any finite graph G and we have $\widetilde{G} = \tau f^G = \widetilde{f^G} \in \widetilde{\mathcal{W}}$, where τ is a mapping, $\tau : f \to \widetilde{f}$. For completeness, we prove the compactness of the metric space, which is crucial for some of the following proofs.

LEMMA 5: *The metric space* $(\widetilde{\mathcal{W}}, \delta_\square)$ *is compact*.

PROOF: The proof follows similar steps as in Theorem 5.1 of Lovasz and Szegedy (2007). For every function $h \in \mathcal{W}$ and a partition $\mathcal{P} = \{P_1, P_2, \ldots, P_k\}$ of $[0,1]$ into measurable sets, we define $h_\mathcal{P} : [0,1]^2 \to [0,1]$ to be the step-function obtained from h by replacing its value at $(x,y) \in P_i \times P_j$ by the average of h over $P_i \times P_j$.

Let $h_1, h_2, \ldots$ be a sequence of functions in $\mathcal{W}$. We need to construct a subsequence that has limit in $\widetilde{\mathcal{W}}$. According to Lemmas 3.1.20 and 3.1.21 in Boeckner (2013), we can create a partition $\mathcal{P}_{n,k} = \{P_{1,n,k}, \ldots, P_{m_k,n,k}\}$ of $[0,1]$ for every n and k. This partition corresponds to a step-function $h_{n,k} = h_{\mathcal{P}_{n,k}} \in \mathcal{W}$, such that:

1. $\delta_\square(h_n, h_{n,k}) \leq 1/k$,
2. $|\mathcal{P}_{n,k}| = m_k$ (where m_k only depends on k),
3. the partition $\mathcal{P}_{n,k+1}$ refines the partition $\mathcal{P}_{n,k}$ for every k.

Notice that since $\delta_\square(h_n, h_{n,k}) \leq 1/k$, we can rearrange the range of $h_{n,k}$ so that all the steps of the function are intervals. Select a subsequence of h_n such that the length of the ith interval $P_{i,n,1}$ of $h_{n,1}$ converges for every i as $n \to \infty$; and the value $h_{n,1}$ on $P_{i,n,1} \times P_{j,n,1}$ also converges for every i and j as $n \to \infty$. Hence, the sequence $h_{n,1}$ converges to a limit almost everywhere. Let us call the limit U_1: notice that U_1 is also a step-function with m_1 steps (that are themselves intervals). We can repeat this procedure for $k = 2, 3, \ldots$. We obtain subsequences for which $h_{n,k} \to U_k$ almost everywhere, and U_k is a step-function with m_k steps.

We know that for every $k < l$, the partition $\mathcal{P}_{n,l}$ is a refinement of partition $\mathcal{P}_{n,k}$. As a consequence, the partition into the steps of $h_{n,l}$ is a refinement of the partition into the steps of $h_{n,k}$. Clearly, the same relation must hold for U_l and U_k, that is, the partition into the steps of U_l is a refinement of the partition into the steps of U_k. By construction of $h_\mathcal{P}$, the function $h_{n,k}$ can be obtained from $h_{n,l}$ by averaging its value over each step. As a consequence, the same holds for U_l and U_k.

It is shown in the proof of Lemma 3.1.21 in Boeckner (2013) that if we pick a random point (X, Y) uniformly over $[0,1]^2$, the sequence $U_1(X,Y), U_2(X,Y), \ldots$ is martingale, and each element of the sequence is bounded. Using the Martingale Convergence Theorem, we can show that the sequence $U_1(X,Y), U_2(X,Y), \ldots$ converges almost everywhere. We define this limit U.

The rest of the proof is the same as in Theorem 5.1 of Lovasz and Szegedy (2007). Fix an $\varepsilon > 0$. Then there exists a $k > 3/\varepsilon$, which we denote as K, such that $\|U - U_k\|_1 < \varepsilon/3$. Fix $k = K$; then there is an N such that, for all $n \geq N$, we have $\|U_k - h_{n,k}\|_1 < \varepsilon/3$. Then

we finally have

$$\begin{aligned}\delta_{\square}(U, h_n) &\leq \delta_{\square}(U, U_k) + \delta_{\square}(U_k, h_{n,k}) + \delta_{\square}(h_{n,k}, h_n) \\ &\leq \|U - U_k\|_1 + \|U_k - h_{n,k}\|_1 + \delta_{\square}(h_{n,k}, h_n) \\ &\leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.\end{aligned}$$

As a consequence, $h_n \to U$ in the metric space $(\widetilde{\mathcal{W}}, \delta_{\square})$. *Q.E.D.*

D.2. *A Crash Course on Large Deviations for Random Graphs*

D.2.1. *Undirected Graphs (Original Chatterjee and Varadhan (2011) Formulation)*

Chatterjee and Varadhan (2011) developed a large deviation principle for the undirected Erdős–Rényi graph. Let $G(n, p)$ indicate the random *undirected* graph with n vertices where each link is formed independently with probability p. Define a function $I_p : [0, 1] \to \mathbb{R}$:

$$I_p(u) \equiv \frac{1}{2} u \log \frac{u}{p} + \frac{1}{2}(1 - u) \log \frac{1 - u}{1 - p} \tag{45}$$

whose domain is easily extended to $\mathcal{W}_o$ as

$$\begin{aligned}I_p(h) &= \int_0^1 \int_0^1 I_p\big(h(x, y)\big)\, dx\, dy \\ &= \frac{1}{2} \int_0^1 \int_0^1 \left[h(x, y) \log \frac{h(x, y)}{p} + \big(1 - h(x, y)\big) \log \frac{1 - h(x, y)}{p} \right] dx\, dy.\end{aligned} \tag{46}$$

Analogously, we can define I_p on $\widetilde{\mathcal{W}}_o$ as $I_p(\widetilde{h}) \equiv I_p(h)$. The graph $G(n, p)$ induces a probability distribution $P_{n,p}$ on $\mathcal{W}_o$, because we can use the map $G \to f^G$; and it induces a probability distribution $\widetilde{P}_{n,p}$ on $\widetilde{\mathcal{W}}_o$ according to the map $G \to f^G \to \widetilde{f}^G = \widetilde{G}$. Chatterjee and Varadhan (2011) stated a large deviation principle for the Erdős–Rényi random graph in both spaces $(\mathcal{W}_o, d_{\square})$ and $(\widetilde{\mathcal{W}}_o, \delta_{\square})$.

We report the main result of Chatterjee and Varadhan (2011) for completeness.

THEOREM 7—Large Deviation Principle for Erdős–Rényi Graph, Chatterjee and Varadhan (2011): *For each fixed $p \in (0, 1)$, the sequence $\widetilde{P}_{n,p}$ obeys a large deviation principle in the space $(\widetilde{\mathcal{W}}_o, \delta_{\square})$ with rate function $I_p(h)$ defined in* (46). *For any closed set $\widetilde{F} \subseteq \widetilde{\mathcal{W}}$,*

$$\limsup_{n \to \infty} \frac{1}{n^2} \log \widetilde{P}_{n,p}(\widetilde{F}) \leq - \inf_{\widetilde{h} \in \widetilde{F}} I_p(\widetilde{h}),$$

and for any open set $\widetilde{U} \subseteq \widetilde{\mathcal{W}}$,

$$\liminf_{n \to \infty} \frac{1}{n^2} \log \widetilde{P}_{n,p}(\widetilde{U}) \geq - \inf_{\widetilde{h} \in \widetilde{U}} I_p(\widetilde{h}).$$

D.2.2. *Directed Graphs*

First, we consider the extension of Theorem 7 to *directed* Erdős–Rényi graphs. Let $G_d(n, p)$ indicate the random *directed* graph with n vertices where each arc is formed independently with probability p. Define a function $\mathcal{I}_p : [0, 1] \to \mathbb{R}$

$$\mathcal{I}_p(u) \equiv u \log \frac{u}{p} + (1-u) \log \frac{1-u}{1-p} : \tag{47}$$

whose domain is easily extended to $\mathcal{W}$ as

$$\begin{aligned} \mathcal{I}_p(h) &= \int_0^1 \int_0^1 I_p\big(h(x,y)\big)\, dx\, dy \\ &= \int_0^1 \int_0^1 \left[h(x,y) \log \frac{h(x,y)}{p} + \big(1 - h(x,y)\big) \log \frac{1-h(x,y)}{p} \right] dx\, dy. \end{aligned} \tag{48}$$

Analogously, we can define $\mathcal{I}_p$ on $\widetilde{\mathcal{W}}$ as $\mathcal{I}_p(\widetilde{h}) \equiv \mathcal{I}_p(h)$. Chatterjee and Varadhan (2011) (see their Lemma 2.1) proved that this function is lower semicontinuous on $\widetilde{\mathcal{W}}$ under the metric $\delta_\square$.

The graph $G_d(n, p)$ induces a probability distribution $\mathcal{P}_{n,p}$ on $\mathcal{W}$, because we can use the map $G \to f^G$; and it induces a probability distribution $\widetilde{\mathcal{P}}_{n,p}$ on $\widetilde{\mathcal{W}}$ according to the map $G \to f^G \to \widetilde{f}^G = \widetilde{G}$. The large deviation principle for this case is presented in the following theorem.

THEOREM 8—Large Deviation Principle for Directed Erdős–Rényi Graph: *For each fixed $p \in (0, 1)$, the sequence $\widetilde{\mathcal{P}}_{n,p}$ obeys a large deviation principle in the space $(\widetilde{\mathcal{W}}, \delta_\square)$ with rate function $\mathcal{I}_p(h)$ defined in* (48). *For any closed set $\widetilde{F} \subseteq \widetilde{\mathcal{W}}$,*

$$\limsup_{n\to\infty} \frac{1}{n^2} \log \widetilde{\mathcal{P}}_{n,p}(\widetilde{F}) \leq -\inf_{\widetilde{h}\in\widetilde{F}} \mathcal{I}_p(\widetilde{h}),$$

and for any open set $\widetilde{U} \subseteq \widetilde{\mathcal{W}}$,

$$\liminf_{n\to\infty} \frac{1}{n^2} \log \widetilde{\mathcal{P}}_{n,p}(\widetilde{U}) \geq -\inf_{\widetilde{h}\in\widetilde{U}} \mathcal{I}_p(\widetilde{h}).$$

PROOF: The proof follows the same steps as in the original theorem for undirected graphs in Chatterjee and Varadhan (2011), but substituting the new rate function in (48). For the upper bound, we define $p_{i,j}$ as in the original paper, but we do not require symmetry. We use slightly different regularity conditions, as provided in Boeckner (2013), because of the directed nature of the graph. In particular, we use Lemmas 3.1.14, 3.1.20, and 3.1.21 in Boeckner (2013). With these small changes, Lemma 2.4, 2.5, and 2.6 in Chatterjee and Varadhan (2011) hold. The proof follows the same steps as in the undirected case. For the lower bound, the proof is identical, without the requirement of symmetry. *Q.E.D.*

D.3. *Undirected ERGM (Chatterjee and Diaconis (2013))*

Let $T : \widetilde{\mathcal{W}}_o \to \mathbb{R}$ be a bounded continuous function in space $(\widetilde{\mathcal{W}}_o, \delta_\square)$. For a given n, the probability function for the graphs is given by

$$\pi_n(G) = \exp\{n^2[T(\widetilde{G}) - \psi_n]\},$$

where $\widetilde{G}$ is defined on $\widetilde{\mathcal{W}}_o$ according to the map $G \to f^G \to \widetilde{f}^G = \widetilde{G}$, and ψ_n is a constant defined as

$$\psi_n = \frac{1}{n^2} \log \sum_{G \in \mathcal{G}_n} \exp\{n^2[T(\widetilde{G})]\}. \tag{49}$$

The re-scaling by n^2 is necessary to guarantee that the limits for $n \to \infty$ converge to some nontrivial quantity. We are interested in finding the value of ψ_n as $n \to \infty$. We define a rate function

$$I(u) \equiv \frac{1}{2} u \log u + \frac{1}{2}(1-u)\log(1-u) \tag{50}$$

which we extend to $\widetilde{\mathcal{W}_o}$ as

$$\begin{aligned} I(\widetilde{h}) &\equiv \frac{1}{2} \int_0^1 \int_0^1 I\big(h(x,y)\big)\, dx\, dy, \\ I(\widetilde{h}) &\equiv \frac{1}{2} \int_0^1 \int_0^1 I\big(h(x,y)\big)\, dx\, dy \\ &= \frac{1}{2} \int_0^1 \int_0^1 \big[h(x,y)\log h(x,y) + \big(1-h(x,y)\big)\log\big(1-h(x,y)\big)\big]\, dx\, dy. \end{aligned} \tag{51}$$

THEOREM 9—Theorem 3.1 for ERGM in Chatterjee and Diaconis (2013): *If* $T : \widetilde{\mathcal{W}}_o \to \mathbb{R}$ *is a bounded continuous function and* ψ_n *and* I *are defined as in* (49) *and* (51)*, respectively, then*

$$\psi \equiv \lim_{n\to\infty} \psi_n = \sup_{\widetilde{h} \in \widetilde{\mathcal{W}}_o} \{T(\widetilde{h}) - I(\widetilde{h})\}.$$

D.4. *Directed ERGM*

Let $\mathcal{T} : \widetilde{\mathcal{W}} \to \mathbb{R}$ be a bounded continuous function in space $(\widetilde{\mathcal{W}}, \delta_\square)$. In our model, $\mathcal{T}$ corresponds to the potential function Q of the network formation game after re-scaling some of the utility components (see below for details and examples). In what follows, we omit the dependence on the parameters to simplify notation. For a given n, the probability of observing network G is given by

$$\pi_n(G) = \exp\{n^2[\mathcal{T}(\widetilde{G}) - \psi_n]\},$$

where $\widetilde{G}$ is defined on $\widetilde{\mathcal{W}}$ according to the map $G \to f^G \to \widetilde{f}^G = \widetilde{G}$, and ψ_n is a normalization constant defined as

$$\psi_n = \frac{1}{n^2} \log \sum_{G \in \mathcal{G}_n} \exp\{n^2[\mathcal{T}(\widetilde{G})]\}. \tag{52}$$

This is the same as the stationary distribution of our model, after some re-scaling of the utility functions. The re-scaling by n^2 is necessary to guarantee that the limits for $n \to \infty$ converge to some nontrivial quantity. We are interested in finding the value of ψ_n as $n \to \infty$, using the same line of reasoning in Theorem 3.1 of Chatterjee and Diaconis (2013). We define a rate function

$$\mathcal{I}(u) \equiv u \log u + (1-u)\log(1-u), \tag{53}$$

which we extend to $\widetilde{\mathcal{W}}$ as

$$\begin{aligned}\mathcal{I}(\widetilde{h}) &\equiv \int_0^1 \int_0^1 I\big(h(x,y)\big)\,dx\,dy \\ &= \int_0^1 \int_0^1 \big[h(x,y)\log h(x,y) + \big(1-h(x,y)\big)\log\big(1-h(x,y)\big)\big]\,dx\,dy.\end{aligned} \tag{54}$$

THEOREM 10—Asymptotic Log-Constant for Directed ERGM: *If* $\mathcal{T} : \widetilde{\mathcal{W}} \to \mathbb{R}$ *is a bounded continuous function and* ψ_n *and* $\mathcal{I}$ *are defined as in* (52) *and* (54), *respectively, then*

$$\psi \equiv \lim_{n\to\infty} \psi_n = \sup_{\widetilde{h}\in\widetilde{\mathcal{W}}} \big\{\mathcal{T}(\widetilde{h}) - \mathcal{I}(\widetilde{h})\big\}. \tag{55}$$

PROOF: The proof of this result follows closely the proof of Theorem 3.1 in Chatterjee and Diaconis (2013), with minimal changes. Let $\widetilde{A}$ denote a Borel set $\widetilde{A} \subseteq \widetilde{\mathcal{W}}$. For each n, let $\widetilde{A}_n$ be the (finite) set

$$\widetilde{A}_n \equiv \{\widetilde{h} \in \widetilde{A} \text{ such that } \widetilde{h} = \widetilde{G} \text{ for some } G \in \mathcal{G}_n\}.$$

Let $\mathcal{P}_{n,p}$ be the probability distribution of the directed random graph $G_d(n,p)$ defined above. We have

$$|\widetilde{A}_n| = 2^{n(n-1)}\mathcal{P}_{n,1/2}(\widetilde{A}_n) = 2^{n(n-1)}\mathcal{P}_{n,1/2}(\widetilde{A}).$$

We can use the result in Theorem 8 to show that, for a closed subset $\widetilde{F}$ of $\widetilde{\mathcal{W}}$, we have

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n^2}\log \widetilde{\mathcal{P}}_{n,1/2}(\widetilde{F}_n) &= \limsup_{n\to\infty} \frac{1}{n^2}\big[\log|\widetilde{F}_n| - n(n-1)\log 2\big] \\ &= \limsup_{n\to\infty} \frac{1}{n^2}\log|\widetilde{F}_n| - \log 2 \\ &\le -\inf_{\widetilde{h}\in\widetilde{F}} \mathcal{I}_{1/2}(\widetilde{h}).\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n^2}\log|\widetilde{F}_n| &\le \log 2 - \inf_{\widetilde{h}\in\widetilde{F}} \mathcal{I}_{1/2}(\widetilde{h}) \\ &= \inf_{\widetilde{h}\in\widetilde{F}} \mathcal{I}(\widetilde{h}).\end{aligned}$$

Similarly, for an open subset $\widetilde{U}$ of $\widetilde{\mathcal{W}}$, we have

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n^2}\log|\widetilde{U}_n| &\geq \log 2 - \inf_{\widetilde{h}\in\widetilde{U}}\mathcal{I}_{1/2}(\widetilde{h})\\ &= \inf_{\widetilde{h}\in\widetilde{U}}\mathcal{I}(\widetilde{h}).\end{aligned}$$

The rest of the proof is equivalent to the undirected case (see proof of Theorem 3.1 in Chatterjee and Diaconis (2013)). *Q.E.D.*

The result of Theorem 10 shows that as n grows large, we can compute the normalizing constant of the ERGM as the result of a variational problem. The main issue is that the variational problem does not have a closed-form solution for most cases. However, there are some special cases in which the solution can be computed explicitly. Let us consider a model with utility from directed links and friends of friends. Using the notation developed above, we are considering a model with function $\mathcal{T}$:

$$\mathcal{T}(\widetilde{G}) = \theta_1\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}g_{ij}}{n^2} + \theta_2\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}g_{ij}g_{jk}}{n^3}. \tag{56}$$

For any $h\in\mathcal{W}$, we can define

$$\mathcal{T}(h) = \theta_1 t(H_1, h) + \theta_2 t(H_2, h),$$

where the limit objects are

$$t(H_1, h) = \int\int_{[0,1]^2} h(x, y)\,dx\,dy$$

and

$$t(H_2, h) = \int\int\int_{[0,1]^3} h(x, y)h(y, z)\,dx\,dy\,dz.$$

We will assume that $\theta_2 > 0$. In this case, there is an explicit solution of the variational problem. The following theorem provides a characterization of the variational problem along the same lines of Radin and Yin (2013) and Aristoff and Zhu (2014).

THEOREM 11: *Let $\theta_2 > 0$ and $\mathcal{T}$ be defined as in (56) above. Then*

$$\lim_{n\to\infty}\psi_n = \psi = \sup_{\mu\in[0,1]}\left\{\theta_1\mu + \theta_2\mu^2 - \mu\log\mu - (1-\mu)\log(1-\mu)\right\}.$$

1. *If $\theta_2 \leq 2$, the maximization problem has a unique maximizer $\mu^*\in[0,1]$.*
2. *If $\theta_2 > 2$ and $\theta_1 \geq -2$, then there is a unique maximizer $\mu^* > 0.5$.*
3. *If $\theta_2 > 2$ and $\theta_1 < -2$, then there is a V-shaped region of the parameters such that*

(a) *inside the V-shaped region, the maximization problem has two local maximizers $\mu_1^* < 0.5 < \mu_2^*$;*

(b) *outside the V-shaped region, the maximization problem has a unique maximizer μ^*.*

4. *For any θ_1 inside the V-shaped region, there exists a $\theta_2 = q(\theta_1)$, such that the two maximizers are both global, that is, $\ell(\mu_1^*) = \ell(\mu_2^*)$.*

PROOF: We need to use the Holder inequality: if p, q are such that $1/p + 1/q = 1$, then for any measurable functions f, g defined on the same domain,

$$\int f(x)g(x)\,dx \le \left(\int f(x)^p\,dx\right)^{\frac{1}{p}} \left(\int g(x)^q\,dx\right)^{\frac{1}{q}}.$$

In particular, we have in our case

$$\begin{aligned}
t(H_2, h) &= \int\int\int_{[0,1]^3} h(x,y)h(y,z)\,dx\,dy\,dz \\
&\le \left(\int\int\int_{[0,1]^3} h(x,y)^2\,dx\,dy\,dz\right)^{\frac{1}{2}} \left(\int\int\int_{[0,1]^3} h(y,z)^2\,dx\,dy\,dz\right)^{\frac{1}{2}} \\
&= \left(\int\int_{[0,1]^2} h(x,y)^2 \left[\int_{[0,1]} dz\right] dx\,dy\right)^{\frac{1}{2}} \left(\int\int_{[0,1]^2} h(y,z)^2 \left[\int_{[0,1]} dx\right] dy\,dz\right)^{\frac{1}{2}} \\
&= \left(\int\int_{[0,1]^2} h(x,y)^2\,dx\,dy\right)^{\frac{1}{2}} \left(\int\int_{[0,1]^2} h(y,z)^2\,dy\,dz\right)^{\frac{1}{2}} \\
&= \left(\int\int_{[0,1]^2} h(x,y)^2\,dx\,dy\right)^{\frac{1}{2}} \left(\int\int_{[0,1]^2} h(x,y)^2\,dx\,dy\right)^{\frac{1}{2}} \\
&= \int\int_{[0,1]^2} h(x,y)^2\,dx\,dy.
\end{aligned}$$

We have assumed that $\theta_2 > 0$. Given the results of the Holder inequality, we can say that

$$\begin{aligned}
\mathcal{T}(h) &= \theta_1 t(H_1, h) + \theta_2 t(H_1, h) \\
&= \theta_1 \int\int_{[0,1]^2} h(x,y)\,dx\,dy + \theta_2 \int\int\int_{[0,1]^3} h(x,y)h(y,z)\,dx\,dy\,dz \\
&\le \theta_1 \int\int_{[0,1]^2} h(x,y)\,dx\,dy + \theta_2 \int\int_{[0,1]^2} h(x,y)^2\,dx\,dy.
\end{aligned}$$

Suppose $h(x, y) = \mu$ is a constant. Then the equality holds, and if $\mu \in [0, 1]$ solves the variational problem

$$\lim_{n\to\infty} \psi_n(\theta) = \psi(\theta) = \sup_{\mu\in[0,1]} \theta_1\mu + \theta_2\mu^2 - \mu\log\mu - (1-\mu)\log(1-\mu),$$

then $h(x, y) = \mu$ is the limit graphon.

To show that this is the only solution, let us consider the maximization problem again. For $h(x, y)$ to be a solution, we need

$$\mathcal{T}(h) = \theta_1 \int\int_{[0,1]^2} h(x,y)\,dx\,dy + \theta_2 \int\int_{[0,1]^2} h(x,y)^2\,dx\,dy.$$

In other words, the Holder inequality must hold with equality, that is, we need

$$t(H_2, h) = \int\int\int_{[0,1]^3} h(x,y)h(y,z)\,dx\,dy\,dz$$
$$= \int\int_{[0,1]^2} h(x,y)^2\,dx\,dy.$$

This implies that

$$h(x,y) = h(y,z)$$

for almost all (x, y, z). In particular, we have that, given x and y, $\mu = h(x,y) = h(y,z)$ for any $z \in [0,1]$ because the left-hand side does not depend on z. Given y and z, we have $\mu' = h(y,z) = h(x,y)$ for any $x \in [0,1]$ because the left-hand side does not depend on x. For $x = y$ and $z = y$, we have $\mu = h(y,y) = h(y,y) = \mu'$. In addition, we have $h(x,y) = h(y,x) = \mu = h(x,z)$. It follows that $h(x,y) = \mu$ almost everywhere.

It follows that $\mathcal{T}(h) = \theta_1\mu + \theta_2\mu^2$ and $I(\mu) = \mu\log\mu + (1-\mu)\log(1-\mu)$, so we get

$$\lim_{n\to\infty}\psi_n = \psi = \sup_{\mu\in[0,1]}\left\{\theta_1\mu + \theta_2\mu^2 - \mu\log\mu - (1-\mu)\log(1-\mu)\right\}.$$

We can now characterize the maximization problem above, to obtain the rest of the results. The analysis follows the same steps of Radin and Yin (2013) and Aristoff and Zhu (2014). The first-order and second-order conditions are

$$\ell'(\mu,\theta_1,\theta_2) = \theta_1 + 2\theta_2\mu - \log\frac{\mu}{1-\mu}, \tag{57}$$

$$\ell''(\mu,\theta_1,\theta_2) = 2\theta_2 - \frac{1}{\mu(1-\mu)}. \tag{58}$$

Let us study the concavity of $\ell(\mu;\theta_1,\theta_2)$. We have that $\ell''(\mu,\theta_1,\theta_2) \le 0$ when

$$\theta_2 \le \frac{1}{2\mu(1-\mu)}.$$

Notice that $2 \le \frac{1}{2\mu(1-\mu)} \le \infty$ for any $\mu \in [0,1]$; and $\frac{1}{2\mu(1-\mu)} = 2$ if $\mu = 0.5$. As a consequence, the function $\ell(\mu;\theta_1,\theta_2)$ is concave on the whole interval $[0,1]$ for $\theta_2 \le 2$.

When $\theta_2 > 2$, the second derivative can be positive or negative, with inflection points denoted as a and b; notice that $a < 0.5 < b$.[50] Consider the first derivative $\ell'(\mu,\theta_1,\theta_2)$. For $\theta_2 \le 2$, the derivative is decreasing for any μ, because $\ell''(\mu,\theta_1,\theta_2) \le 0$ for any $\mu \in [0,1]$.

For $\theta_2 > 2$ then (see Figure 9), it is decreasing in $[0,a)$, increasing in (a,b), and decreasing in $(b,1]$.

The function $\ell(\mu,\theta_1,\theta_2)$ is bounded and continuous for any θ and $\mu \in [0,1]$, and we could find the interior maximizers by studying the first and second derivative. If we consider the case $\theta_2 \le 2$, the derivative $\ell'(\mu,\theta_1,\theta_2)$ is decreasing on the whole interval $[0,1]$.

[50]This is because, when $\theta_2 > 2$, we have $\ell''(\mu,\theta_1,\theta_2) \le 0$ when $\theta_2 \le \frac{1}{2\mu(1-\mu)}$ or $2\mu(1-\mu) \le \frac{1}{\theta_2}$. The equality is realized at two intersections of the horizontal line $1/\theta_2$ with the parabola $2\mu(1-\mu)$. We call the intersections $\frac{1}{\theta_2} = 2\mu(1-\mu)$, respectively a and b.

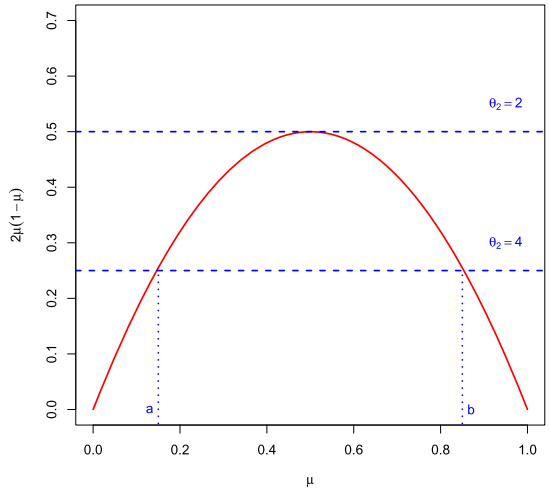

FIGURE 9.—Analysis of second derivative $\ell''(\mu, \theta_1, \theta_2)$.

It is easy to show that $\ell'(0) = \infty$ and $\ell'(1) = -\infty$. Therefore, when $\theta_2 \leq 2$, there is only one maximizer μ^* that solves $\ell'(\mu, \theta_1, \theta_2) = 0$.

If $\theta_2 > 2$, then we have three possible cases. We know that in this region $\ell'(\mu, \theta_1, \theta_2)$ is decreasing in $[0, a)$, increasing in (a, b), and decreasing in $(b, 1]$.

1. If $\ell'(a, \theta_1, \theta_2) \geq 0$, then there is a unique maximizer $\mu^* > b$.
2. If $\ell'(b, \theta_1, \theta_2) \leq 0$, then there is a unique maximizer $\mu^* < a$.
3. If $\ell'(a, \theta_1, \theta_2) < 0 < \ell'(b, \theta_1, \theta_2)$, then there are two local maximizers $\mu_1^* < a < b < \mu_2^*$.

The three cases are shown in Figure 10, where we plot $\ell'(\mu, \theta_1, \theta_2)$ against μ for several values of θ_1 and for a fixed $\theta_2 = 4 > 2$.

We indicate the maximizer with μ^* when it is unique, and with μ_1^*, μ_2^* when there are two.

Let us consider the first case, with $\ell'(a, \theta_1, \theta_2) \geq 0$. To compute $\ell'(a, \theta_1, \theta_2)$, notice that $\theta_2 = \frac{1}{2a(1-a)}$. Substituting in $\ell'(a, \theta_1, \theta_2)$, we obtain

$$\ell'(a, \theta_1, \theta_2) = \theta_1 + \frac{1}{1-a} - \log \frac{a}{1-a},$$

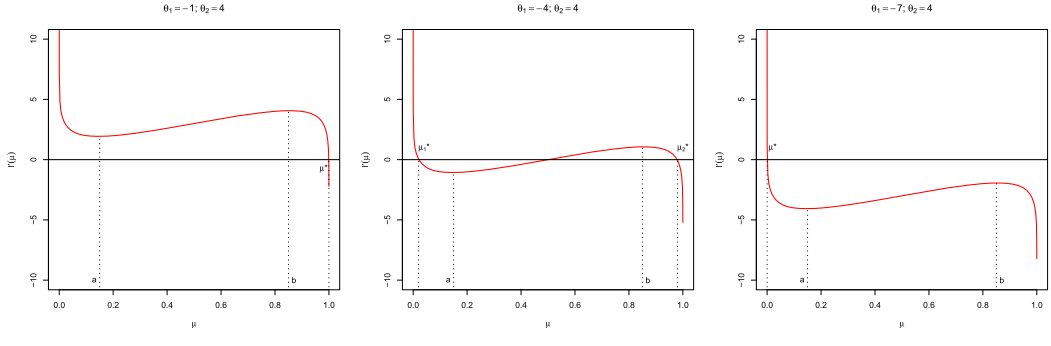

FIGURE 10.—First order conditions for maximization of $\ell(\mu, \theta_1, \theta_2)$: three cases (see text for details).

and analogously for $\theta_2 = \frac{1}{2b(1-b)}$ we have

$$\ell'(b, \theta_1, \theta_2) = \theta_1 + \frac{1}{1-b} - \log \frac{b}{1-b}.$$

So $\ell'(a, \theta_1, \theta_2) \geq 0$ implies

$$\theta_1 \geq \log \frac{a}{1-a} - \frac{1}{1-a}.$$

The function $\log \frac{a}{1-a} - \frac{1}{1-a}$ has a maximum at -2 and therefore we have[51]

$$\ell'(a, \theta_1, \theta_2) \geq 0 \quad \Leftrightarrow \quad \theta_1 \geq -2.$$

When the above condition is satisfied, there is a unique maximizer, $\mu^* > b$, as shown in the picture on the left.

When $\theta_1 < -2$, it is easier to draw a picture of the function $\log \frac{a}{1-a} - \frac{1}{1-a}$, shown in Figure 11.

Notice that when $\theta_1 < -2$, there are two intersections of the function and the horizontal line $y = \theta_1$ (in the picture $\theta_1 = -3$). We denote the intersections $\phi_1(\theta_1)$ and $\phi_2(\theta_1)$. By construction, we know that $a < 0.5 < b$. By looking at the picture, it is clear that $\ell'(a, \theta_1, \theta_2) > 0$ if $a < \phi_1(\theta_1)$ and $\ell'(a, \theta_1, \theta_2) < 0$ if $a > \phi_1(\theta_1)$. Analogously, we have $\ell'(b, \theta_1, \theta_2) > 0$ if $b > \phi_2(\theta_1)$ and $\ell'(b, \theta_1, \theta_2) < 0$ if $b < \phi_2(\theta_1)$.

For any $\theta_1 < -2$, there exist $\phi_1(\theta_1)$ and $\phi_2(\theta_1)$ which are the intersection of the function $y = \log(\frac{x}{1-x}) - \frac{1}{1-x}$ with the line $y = \theta_1$. Since the function is continuous, monotonic increasing in $[0, 0.5)$, and monotonic decreasing in $(0.5, 1]$, it follows that $\phi_1(\theta_1)$ and $\phi_2(\theta_1)$ are both continuous in θ_1. In addition, $\phi_1(\theta_1)$ is increasing in θ_1 and $\phi_2(\theta_1)$ is

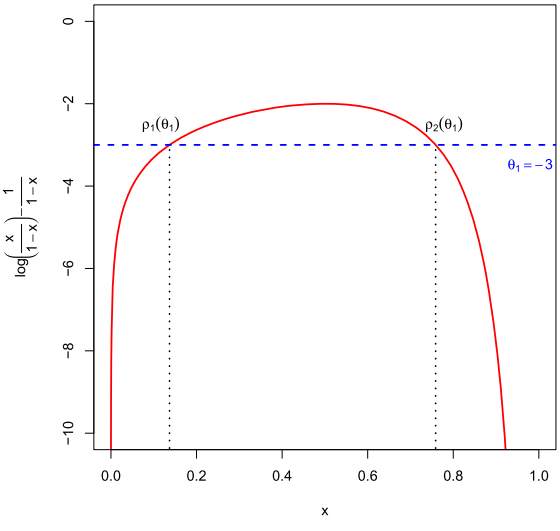

FIGURE 11.—Determination of values $\phi_1(\theta_1)$ and $\phi_2(\theta_1)$.

[51]Taking derivative $\frac{1}{a} + \frac{1}{1-a} - \frac{1}{(1-a)^2} = 0$, we obtain the maximizer $a^* = 0.5$. The function is increasing in $[0, 0.5)$ and decreasing in $(0.5, 1]$. The maximum is therefore at -2.

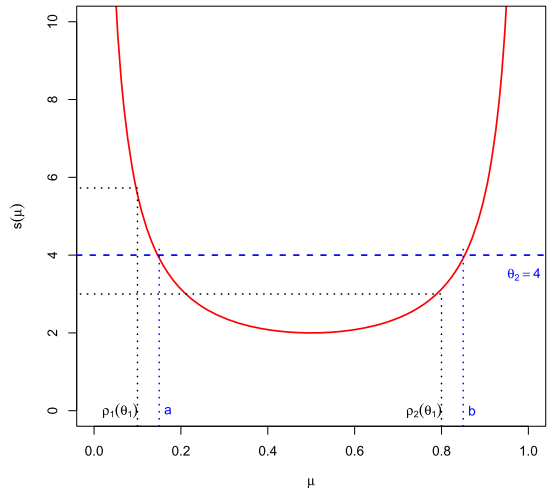

FIGURE 12.—The case with two maximizers.

decreasing in θ_1. It is trivial to show that when θ_1 decreases, $\phi_1(\theta_1)$ converges to 0 while $\phi_2(\theta_1)$ converges to 1.

Consider the case in which $\ell'(a, \theta_1, \theta_2) < 0 < \ell'(b, \theta_1, \theta_2)$ with two maximizers. Define the function

$$s(\mu) \equiv \frac{1}{2\mu(1-\mu)}.$$

Since $\ell'(a, \theta_1, \theta_2) < 0$, we have $a > \phi_1(\theta_1)$, which implies $s(a) < s(\phi_1(\theta_1))$. Therefore, $\theta_2 < s(\phi_1(\theta_1)) = \frac{1}{2\phi_1(\theta_1)(1-\phi_1(\theta_1))}$.

Since $\ell'(b, \theta_1, \theta_2) > 0$, we have $b > \phi_2(\theta_1)$, which implies $s(b) > s(\phi_2(\theta_1))$. Therefore, $\theta_2 > s(\phi_2(\theta_1)) = \frac{1}{2\phi_2(\theta_1)(1-\phi_2(\theta_1))}$.

Notice that $s(\phi_1(\theta_1)) > s(\phi_2(\theta_1))$ for any (θ_1, θ_2) in this region of the parameters (see Figure 12).

The areas are shown in Figure 13.

Within the V-shaped region, there are two solutions to the maximization problem, that is, two local maxima. Also, it is trivial to show that there exists a function q such that, for $\theta_2 = q(\theta_1)$, both solutions are global maxima. Indeed, the two local maxima are both global maxima if $\ell(\mu_2^*, \theta_1, \theta_2) - \ell(\mu_1^*, \theta_1, \theta_2) = 0$. The latter difference is negative when μ_1^* is the global maximizer, while it is positive when μ_2^* is the global maximizer. Therefore, for a given value of θ_1, there must be a unique θ_2 such that $s(\phi_1(\theta_1)) > \theta_2 > s(\phi_2(\theta_1))$ such that both μ_1^* and μ_2^* are global maximizers. Let us indicate this value of $\theta_2 = q(\theta_1)$.

Notice that the difference $\ell(\mu_2^*, \theta_1, \theta_2) - \ell(\mu_1^*, \theta_1, \theta_2)$ corresponds to the difference between the positive and negative areas between $\mu_1^*$and μ_2^* in Figure 14, that is (let $\widehat{\mu}$ indicate the intersection of $\ell'(\mu, \theta_1, \theta_2)$ and the x-axis between $\mu_1^*$and μ_2^*),

$$\begin{aligned}\ell\big(\mu_2^*, \theta_1, \theta_2\big) - \ell\big(\mu_1^*, \theta_1, \theta_2\big) &= \int_0^{\mu_2^*} \ell'(\mu, \theta_1, \theta_2)\, d\mu - \int_0^{\mu_1^*} \ell'(\mu, \theta_1, \theta_2)\, d\mu \\ &= \int_0^{\mu_1^*} \ell'(\mu, \theta_1, \theta_2)\, d\mu + \int_{\mu_1^*}^{\widehat{\mu}} \ell'(\mu, \theta_1, \theta_2)\, d\mu\end{aligned}$$

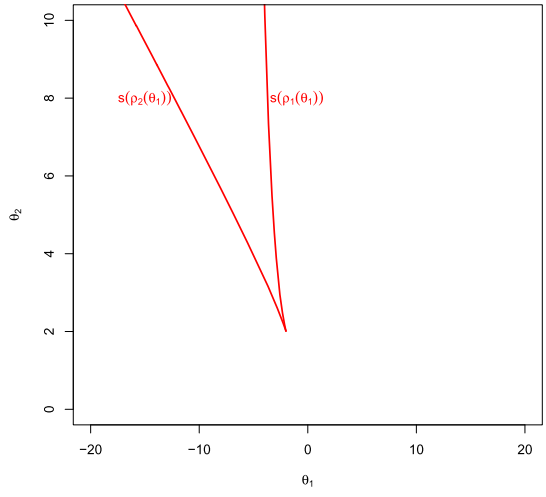

FIGURE 13.—The V-shaped region of the parameters where there are two local maximizers.

$$+\int_{\widehat{\mu}}^{\mu_2^*} \ell'(\mu,\theta_1,\theta_2)\,d\mu - \int_0^{\mu_1^*} \ell'(\mu,\theta_1,\theta_2)\,d\mu$$
$$= \int_{\mu_1^*}^{\widehat{\mu}} \ell'(\mu,\theta_1,\theta_2)\,d\mu + \int_{\widehat{\mu}}^{\mu_2^*} \ell'(\mu,\theta_1,\theta_2)\,d\mu.$$

When this difference is equal to zero, it means that the positive area and the negative area are equivalent and they cancel each other out. If we increase θ_1, then the curve $\ell'(\mu,\theta_1,\theta_2)$ will shift upwards and the negative area will decrease; therefore, we have to decrease θ_2 to counterbalance this effect. The opposite happens when we decrease θ_1.

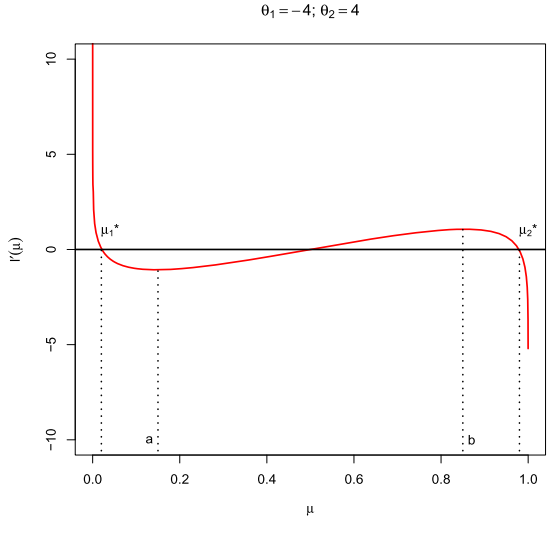

FIGURE 14.—Graphical explanation for the derivation of function $q(\theta_1)$.

Therefore, $q(\theta_1)$ is a downward-sloping curve and it is continuous because of the continuity of $\ell'(\mu, \theta_1, \theta_2)$. This completes the proof. *Q.E.D.*

This theoretical result is confirmed by simulations.

It turns out that there is a more general result. If the homomorphism density $t(H_2, G)$ associated with the parameter θ_2 is such that the resulting variational problem can be shown to be

$$\psi = \sup_{\mu\in[0,1]} \ell(\mu, \alpha, \beta) = \sup_{\mu\in[0,1]} \{\alpha\mu + \beta\mu^r - \mu \ln \mu - (1-\mu)\ln(1-\mu)\},$$

where we assume $r \geq 2$, then the same characterization applies, as shown in the next theorem. For example, this is the case if we consider

$$t(H_2, G) = \frac{\sum_i \sum_j \sum_k g_{ij} g_{jk} g_{ki}}{n^3}$$

with $r = 3$; or if we consider

$$t(H_2, G) = \frac{\sum_i \sum_j \sum_k \sum_l g_{ij} g_{jk} g_{kl} g_{li}}{n^4}$$

with $r = 4$.

The next lemma provides conditions under which the network statistics can be upper-bounded by the power of the graphon. For practical purposes, this condition is necessary to be able to rewrite the variational problem as a calculus problem, as shown in the theorems below.

LEMMA 6: *For the following homomorphism densities*:

$$t(H, G) = \frac{\sum_i \sum_j \sum_k g_{ij} g_{jk} g_{ki}}{n^3}, \tag{59}$$

$$t(H, G) = \frac{\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n g_{ij} g_{jk}}{n^3}, \tag{60}$$

$$t(H, G) = \frac{\sum_i \sum_j \sum_k \sum_l g_{ij} g_{jk} g_{kl} g_{li}}{n^4}, \tag{61}$$

$$t(H, G) = \frac{1}{n^r} \sum_{1\leq i, j_1, j_2, \ldots, j_r \leq n} g_{ij_1} g_{j_1 j_2} \cdots g_{j_r i}, \tag{62}$$

$$t(H, G) = \frac{1}{n^{r-1}} \sum_{1\leq i, j_1, j_2, \ldots, j_r \leq n} g_{ij_1} g_{ij_2} \cdots g_{ij_r}, \tag{63}$$

the following property holds:

$$t(H,h) \leq \int_0^1 \int_0^1 h(x,y)^{e(H)}\,dx\,dy,$$

where $e(H)$ is the number of directed links included in the subgraph H.

PROOF: For the homomorphism density (59), the value $e(H) = 3$ and the limit object is

$$t(H,h) = \int_{[0,1]^3} h(x,y)h(y,z)h(z,x)\,dx\,dy\,dz.$$

Using the Holder inequality and some algebra, we obtain

$$\begin{aligned}
t(H,h) &= \int_{[0,1]^3} h(x,y)h(y,z)h(z,x)\,dx\,dy\,dz \\
&\leq \left(\int_{[0,1]^3} h(x,y)^3\,dx\,dy\,dz\right)^{\frac{1}{3}} \left(\int_{[0,1]^3} h(y,z)^3\,dx\,dy\,dz\right)^{\frac{1}{3}} \\
&\quad \times \left(\int_{[0,1]^3} h(z,x)^3\,dx\,dy\,dz\right)^{\frac{1}{3}} \\
&= \left(\int_{[0,1]^2} h(x,y)^3\,dx\,dy \int_0^1 dz\right)^{\frac{1}{3}} \left(\int_{[0,1]^2} h(y,z)^3\,dy\,dz \int_0^1 dx\right)^{\frac{1}{3}} \\
&\quad \times \left(\int_{[0,1]^2} h(z,x)^3\,dx\,dz \int_0^1 dy\right)^{\frac{1}{3}} \\
&= \left(\int_{[0,1]^2} h(x,y)^3\,dx\,dy\right)^{\frac{1}{3}} \left(\int_{[0,1]^2} h(y,z)^3\,dy\,dz\right)^{\frac{1}{3}} \left(\int_{[0,1]^2} h(z,x)^3\,dx\,dz\right)^{\frac{1}{3}} \\
&= \left(\int_{[0,1]^2} h(x,y)^3\,dx\,dy\right)^{\frac{1}{3}} \left(\int_{[0,1]^2} h(x,y)^3\,dx\,dy\right)^{\frac{1}{3}} \left(\int_{[0,1]^2} h(x,y)^3\,dx\,dy\right)^{\frac{1}{3}} \\
&= \int_{[0,1]^2} h(x,y)^3\,dx\,dy \\
&= \int_0^1 \int_0^1 h(x,y)^{e(H)}\,dx\,dy.
\end{aligned}$$

For the homomorphism density in (60), $e(H) = 2$ and using the Holder inequality, we get

$$\begin{aligned}
t(H,h) &= \int\int\int_{[0,1]^3} h(x,y)h(y,z)\,dx\,dy\,dz \\
&\leq \left(\int\int\int_{[0,1]^3} h(x,y)^2\,dx\,dy\,dz\right)^{\frac{1}{2}} \left(\int\int\int_{[0,1]^3} h(y,z)^2\,dx\,dy\,dz\right)^{\frac{1}{2}}
\end{aligned}$$

$$= \left(\int\int_{[0,1]^2} h(x,y)^2 \left[\int_{[0,1]} dz \right] dx\, dy \right)^{\frac{1}{2}} \left(\int\int_{[0,1]^2} h(y,z)^2 \left[\int_{[0,1]} dx \right] dy\, dz \right)^{\frac{1}{2}}$$

$$= \left(\int\int_{[0,1]^2} h(x,y)^2\, dx\, dy \right)^{\frac{1}{2}} \left(\int\int_{[0,1]^2} h(y,z)^2\, dy\, dz \right)^{\frac{1}{2}}$$

$$= \left(\int\int_{[0,1]^2} h(x,y)^2\, dx\, dy \right)^{\frac{1}{2}} \left(\int\int_{[0,1]^2} h(x,y)^2\, dx\, dy \right)^{\frac{1}{2}}$$

$$= \int\int_{[0,1]^2} h(x,y)^2\, dx\, dy.$$

For the homomorphism density in (62), $e(H) = r$ and using the Holder inequality, we get

$$t(H,h) = \int_{[0,1]^r} h(x_i, x_{j_1}) h(x_{j_1}, x_{j_2}) \cdots h(x_{j_r}, x_i)\, dx_i\, dx_{j_1} \cdots dx_{j_r}$$

$$\leq \left(\int_{[0,1]^r} h(x_i, x_{j_1})^r\, dx_i\, dx_{j_1} \cdots dx_{j_r} \right)^{\frac{1}{r}} \left(\int_{[0,1]^r} h(x_{j_1}, x_{j_2})^r\, dx_i\, dx_{j_1} \cdots dx_{j_r} \right)^{\frac{1}{r}} \cdots$$

$$\times \left(\int_{[0,1]^r} h(x_{j_r}, x_i)^r\, dx_i\, dx_{j_1} \cdots dx_{j_r} \right)^{\frac{1}{r}}$$

$$= \left(\int_{[0,1]^2} h(x_i, x_{j_1})^r\, dx_i\, dx_{j_1} \int_{[0,1]^{r-2}} dx_{j_2} \cdots dx_{j_r} \right)^{\frac{1}{r}}$$

$$\times \left(\int_{[0,1]^2} h(x_{j_1}, x_{j_2})^r\, dx_{j_1}\, dx_{j_2} \int_{[0,1]^{r-2}} dx_i\, dx_{j_3} \cdots dx_{j_r} \right)^{\frac{1}{r}} \cdots$$

$$\times \left(\int_{[0,1]^2} h(x_{j_r}, x_i)^r\, dx_{j_r}\, dx_i \int_{[0,1]^{r-2}} dx_{j_1} \cdots dx_{j_{r-1}} \right)^{\frac{1}{r}}$$

$$= \left(\int_{[0,1]^2} h(x,y)^r\, dx\, dy \right)^{\frac{1}{r}} \left(\int_{[0,1]^2} h(x,y)^r\, dx\, dy \right)^{\frac{1}{r}} \cdots \left(\int_{[0,1]^2} h(x,y)^r\, dx\, dy \right)^{\frac{1}{r}}$$

$$= \int_{[0,1]^2} h(x,y)^r\, dx\, dy$$

$$= \int_0^1 \int_0^1 h(x,y)^{e(H)}\, dx\, dy.$$

For the homomorphism density in (63), $e(H) = r$ and using the Holder inequality, we get

$$t(H,h) = \int_{[0,1]^r} h(x_i, x_{j_1}) h(x_i, x_{j_2}) \cdots h(x_i, x_{j_r})\, dx_i\, dx_{j_1} \cdots dx_{j_r}$$

$$\leq \left(\int_{[0,1]^r} h(x_i, x_{j_1})^r\, dx_i\, dx_{j_1} \cdots dx_{j_r} \right)^{\frac{1}{r}} \left(\int_{[0,1]^r} h(x_i, x_{j_2})^r\, dx_i\, dx_{j_1} \cdots dx_{j_r} \right)^{\frac{1}{r}} \cdots$$

$$
\begin{aligned}
&\times \left(\int_{[0,1]^r} h(x_i, x_{j_r})^r \, dx_i \, dx_{j_1} \cdots dx_{j_r} \right)^{\frac{1}{r}} \\
&= \left(\int_{[0,1]^2} h(x_i, x_{j_1})^r \, dx_i \, dx_{j_1} \int_{[0,1]^{r-2}} dx_{j_2} \cdots dx_{j_r} \right)^{\frac{1}{r}} \\
&\quad \times \left(\int_{[0,1]^2} h(x_i, x_{j_2})^r \, dx_i \, dx_{j_2} \int_{[0,1]^{r-2}} dx_{j_1} \, dx_{j_3} \cdots dx_{j_r} \right)^{\frac{1}{r}} \cdots \\
&\quad \times \left(\int_{[0,1]^2} h(x_i, x_{j_r})^r \, dx_i \, dx_{j_r} \int_{[0,1]^{r-2}} dx_{j_1} \cdots dx_{j_{r-1}} \right)^{\frac{1}{r}} \\
&= \left(\int_{[0,1]^2} h(x, y)^r \, dx \, dy \right)^{\frac{1}{r}} \left(\int_{[0,1]^2} h(x, y)^r \, dx \, dy \right)^{\frac{1}{r}} \cdots \left(\int_{[0,1]^2} h(x, y)^r \, dx \, dy \right)^{\frac{1}{r}} \\
&= \int_{[0,1]^2} h(x, y)^r \, dx \, dy \\
&= \int_0^1 \int_0^1 h(x, y)^{e(H)} \, dx \, dy. \qquad \textit{Q.E.D.}
\end{aligned}
$$

The following theorem uses the result of Lemma 6 above, to show that the variational problem can be solved explicitly as a one-variable calculus problem in special cases. This result is very useful in studying the behavior of the model as the number of players grows large and it provides a way to characterize the convergence of the sampling algorithms according to the same argument of Bhamidi, Bresler, and Sly (2011) (see more detail below).

THEOREM 12: *Let* $\beta > 0$. *For the following models*:

$$
\mathcal{T}(G) = \alpha \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij}}{n^2} + \beta \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} g_{ij} g_{jk}}{n^3},
$$

$$
\mathcal{T}(G) = \alpha \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij}}{n^2} + \beta \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} g_{ij} g_{jk} g_{ki}}{n^3},
$$

$$
\mathcal{T}(G) = \alpha \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij}}{n^2} + \beta \frac{\sum_{1 \le i, j_1, j_2, \ldots, j_r \le n} g_{ij_1} g_{j_1 j_2} \cdots g_{j_r i}}{n^r},
$$

$$
\mathcal{T}(G) = \alpha \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij}}{n^2} + \beta \frac{\sum_{1 \le i, j_1, j_2, \ldots, j_r \le n} g_{ij_1} g_{ij_2} \cdots g_{ij_r}}{n^{r-1}},
$$

the log-partition asymptotic variational problem becomes a calculus problem. Let $\ell(\mu, \alpha, \beta)$ *be the following function*:

$$\ell(\mu, \alpha, \beta) = \alpha\mu + \beta\mu^r - \mu \log \mu - (1-\mu)\log(1-\mu).$$

Then, as $n \to \infty$, *the log-partition is the solution of the following*:

$$\lim_{n\to\infty} \psi_n(\theta) = \psi(\theta) = \sup_{\mu\in[0,1]} \ell(\mu, \alpha, \beta).$$

For the following model with $\beta > 0$ *and* $\gamma > 0$:

$$\mathcal{T}(G) = \alpha \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}}{n^2} + \beta \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} g_{ij}g_{jk}}{n^3} + \gamma \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} g_{ij}g_{jk}g_{ki}}{n^3},$$

the log-partition asymptotic variational problem is

$$\lim_{n\to\infty} \psi_n(\theta) = \psi(\theta) = \sup_{\mu\in[0,1]} \left\{\alpha\mu + \beta\mu^2 + \gamma\mu^3 - \mu\log\mu - (1-\mu)\log(1-\mu)\right\}.$$

PROOF: Consider the first model. We have assumed that $\beta > 0$. Given the results of the Holder inequality in Lemma 6, we can say that

$$\begin{aligned}\mathcal{T}(h) &= \alpha t(H_1, h) + \beta t(H_2, h)\\ &\le \alpha \int\int_{[0,1]^2} h(x,y)\,dx\,dy + \beta \int\int_{[0,1]^2} h(x,y)^2\,dx\,dy.\end{aligned}$$

Suppose $h(x, y) = \mu$ is a constant. Then the equality holds, and if $\mu \in [0, 1]$ solves the variational problem

$$\lim_{n\to\infty} \psi_n(\theta) = \psi(\theta) = \sup_{\mu\in[0,1]} \alpha\mu + \beta\mu^2 - \mu\log\mu - (1-\mu)\log(1-\mu),$$

then $h(x, y) = \mu$ is the limit graphon.

To show that this is the only solution, let us consider the maximization problem again. For $h(x, y)$ to be a solution, we need

$$\mathcal{T}(h) = \alpha \int\int_{[0,1]^2} h(x,y)\,dx\,dy + \beta \int\int_{[0,1]^2} h(x,y)^2\,dx\,dy.$$

In other words, the Holder inequality must hold with equality, that is, we need

$$\begin{aligned}t(H_2, h) &= \int\int\int_{[0,1]^3} h(x,y)h(y,z)\,dx\,dy\,dz\\ &= \int\int_{[0,1]^2} h(x,y)^2\,dx\,dy.\end{aligned}$$

This implies that

$$h(x, y) = h(y, z)$$

for almost all (x, y, z). In particular, we have that, given x and y, $\mu = h(x, y) = h(y, z)$ for any $z \in [0, 1]$ because the left-hand side does not depend on z. Given y and z, we have $\mu' = h(y, z) = h(x, y)$ for any $x \in [0, 1]$ because the left-hand side does not depend on x. For $x = y$ and $z = y$, we have $\mu = h(y, y) = h(y, y) = \mu'$. In addition, we have $h(x, y) = h(y, x) = \mu = h(x, z)$. It follows that $h(x, y) = \mu$ almost everywhere.

It follows that $\mathcal{T}(h) = \alpha\mu + \beta\mu^2$ and $I(\mu) = \mu \log \mu + (1 - \mu) \log(1 - \mu)$, so we get

$$\lim_{n\to\infty} \psi_n = \psi = \sup_{\mu\in[0,1]} \left\{\alpha\mu + \beta\mu^2 - \mu \log \mu - (1 - \mu) \log(1 - \mu)\right\}.$$

The proof for the remaining models follows similar steps and reasoning and it is omitted for brevity. *Q.E.D.*

The next theorem contains a complete characterization of the maximization problem considered in the previous theorem.

THEOREM 13: *Assume that $\beta > 0$ and $r \geq 2$. If the variational problem can be shown to be*

$$\lim_{n\to\infty} \psi_n(\theta) = \psi(\theta) = \sup_{\mu\in[0,1]} \left\{\alpha\mu + \beta\mu^r - \mu \log \mu - (1 - \mu) \log(1 - \mu)\right\},$$

then we have the following:

1. *If $\beta \leq \frac{r^{r-1}}{(r-1)^r}$, the maximization problem has a unique maximizer $\mu^* \in [0, 1]$.*
2. *If $\beta > \frac{r^{r-1}}{(r-1)^r}$ and $\alpha \geq \log(r - 1) - \frac{r}{r-1}$, then there is a unique maximizer $\mu^* > 0.5$.*
3. *If $\beta > \frac{r^{r-1}}{(r-1)^r}$ and $\alpha < \log(r - 1) - \frac{r}{r-1}$, then there is a V-shaped region of parameters such that*

(a) *inside the V-shaped region, the maximization problem has two local maximizers $\mu_1^* < 0.5 < \mu_2^*$;*

(b) *outside the V-shaped region, the maximization problem has a unique maximizer μ^*.*

4. *For any α inside the V-shaped region, there exists a $\beta = \zeta(\alpha)$ such that the two maximizers are both global, that is, $\ell(\mu_1^*) = \ell(\mu_2^*)$.*

PROOF: The first- and second-order conditions are

$$\ell'(\mu, \alpha, \beta) = \alpha + \beta r \mu^{r-1} - \ln\left(\frac{\mu}{1 - \mu}\right),$$

$$\ell''(\mu, \alpha, \beta) = \beta r(r - 1)\mu^{r-2} - \frac{1}{\mu(1 - \mu)}.$$

The function $\ell(\mu, \alpha, \beta)$ is concave if $\ell''(\mu, \alpha, \beta) < 0$, that is, when

$$\beta < \frac{1}{r(r - 1)\mu^{r-1}(1 - \mu)} \equiv s(\mu).$$

The function $s(\mu)$ has a minimum at $\frac{r}{r-1}$, where $s(\frac{r}{r-1}) = \frac{r^{r-1}}{(r-1)^r}$; it is decreasing in the interval $[0, \frac{r}{r-1})$ and increasing in the interval $(\frac{r}{r-1}, 1]$. Therefore, the function $\ell(\mu, \alpha, \beta)$

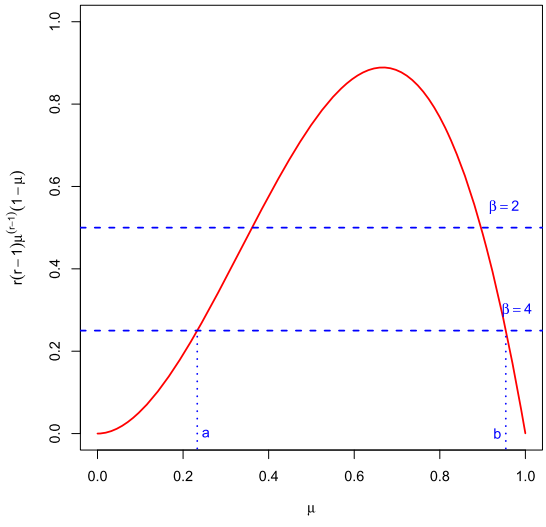

FIGURE 15.—Using function $s(\mu)$ to determine concavity.

is concave on the whole interval $[0,1]$ if $\beta < \frac{r^{r-1}}{(r-1)^r}$.[52] In this region, there is a unique maximizer μ^* of $\ell(\mu,\alpha,\beta)$.

If $\beta > \frac{r^{r-1}}{(r-1)^r}$, there are three possible cases. We know that in this region the second derivative $\ell''(\mu,\alpha,\beta)$ can be positive or negative, with inflection points denoted as a and b, found by solving the equation $\beta = s(\mu)$. An example for $r=3$ and $\beta=4$ is shown in Figure 15 (notice that we are plotting the function $1/s(\mu)$ against the line $1/\beta$).

In particular, the first derivative $\ell'(\mu,\alpha,\beta)$ is decreasing in $[0,a)$, increasing in (a,b), and decreasing in $(b,1]$.

1. If $\ell'(a,\alpha,\beta) \geq 0$, then there is a unique maximizer $\mu^* > b$.
2. If $\ell'(b,\alpha,\beta) \leq 0$, then there is a unique maximizer $\mu^* < a$.
3. If $\ell'(a,\alpha,\beta) < 0 < \ell'(b,\alpha,\beta)$, then there are two local maximizers $\mu_1^* < a < b < \mu_2^*$.

The three cases are shown in Figure 16, where we plot $\ell'(\mu,\alpha,\beta)$ against μ for several values of α and for a fixed $\beta=4$. In the pictures, $r=3$.

[52]Consider the function $1/s(\mu) = r(r-1)\mu^{r-1}(1-\mu) = r(r-1)(\mu^{r-1}-\mu^r)$. This function has derivative

$$\frac{\partial[1/s(\mu)]}{\partial\mu} = r(r-1)^2\mu^{r-2} - r^2(r-1)\mu^{r-1} = r(r-1)\mu^{r-2}\big[(r-1)-r\mu\big],$$

$$\frac{\partial^2[1/s(\mu)]}{\partial\mu\partial\mu} = r(r-1)^2(r-2)\mu^{r-3} - r^2(r-1)^2\mu^{r-2} = r(r-1)^2\mu^{r-3}\big[(r-2)-r\mu\big].$$

So, solving the FOCs, we obtain the maximizer of $1/s(\mu)$,

$$r(r-1)\mu^{r-2}\big[(r-1)-r\mu\big] = 0 \quad \Leftrightarrow \quad \mu = \frac{r-1}{r},$$

and the maximum is

$$1/s\left(\frac{r-1}{r}\right) = r(r-1)\left(\frac{r-1}{r}\right)^{r-1}\left(1-\frac{r-1}{r}\right) = \frac{(r-1)^r}{r^{r-1}}.$$

Therefore, the minimum of $s(\mu)$ is $\frac{r^{r-1}}{(r-1)^r}$.

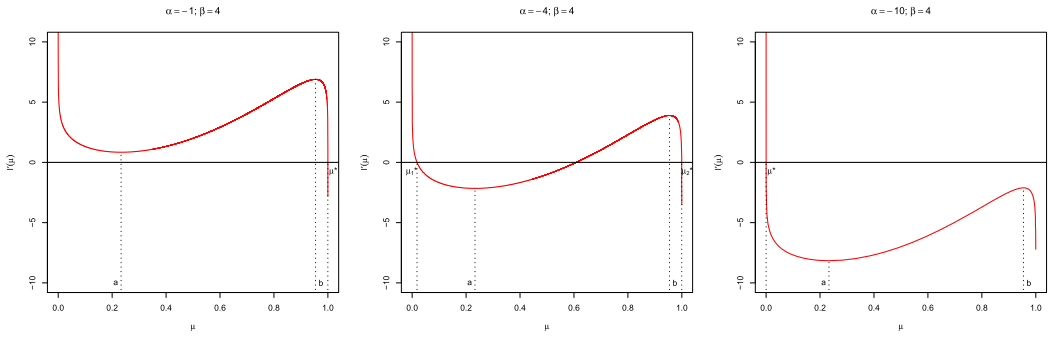

FIGURE 16.—Three possible cases to find the maximizers of $\ell(\mu, \theta_1, \theta_2)$.

We indicate the maximizer with μ^* when it is unique, and with μ_1^*, μ_2^* when there are two.

Let us consider the first case, with $\ell'(a, \alpha, \beta) \geq 0$. To compute $\ell'(a, \alpha, \beta)$, notice that $\beta = s(a) = \frac{1}{r(r-1)a^{r-1}(1-a)}$. Substituting in $\ell'(a, \alpha, \beta)$, we obtain

$$\ell'(a, \alpha, \beta) = \alpha + \frac{1}{(r-1)(1-a)} - \log \frac{a}{1-a},$$

and analogously for $\beta = s(b) = \frac{1}{r(r-1)b^{r-1}(1-b)}$ we have

$$\ell'(b, \alpha, \beta) = \alpha + \frac{1}{(r-1)(1-b)} - \log \frac{b}{1-b}.$$

So $\ell'(a, \alpha, \beta) \geq 0$ implies

$$\alpha \geq \log \frac{a}{1-a} - \frac{1}{(r-1)(1-a)}.$$

The function $\log \frac{a}{1-a} - \frac{1}{(r-1)(1-a)}$ has a maximum at $\log(r-1) - \frac{r}{r-1}$ and therefore we have[53]

$$\ell'(a, \alpha, \beta) \geq 0 \quad \Leftrightarrow \quad \theta_1 \geq \log(r-1) - \frac{r}{r-1}.$$

When the above condition is satisfied, there is a unique maximizer, $\mu^* > b$, as shown in the picture on the left.

When $\theta_1 < \log(r-1) - \frac{r}{r-1}$, it is easier to draw a picture of the function $\log \frac{a}{1-a} - \frac{1}{(r-1)(1-a)}$, shown Figure 17.

Notice that when $\theta_1 < \log(r-1) - \frac{r}{r-1}$, there are two intersections of the function and the horizontal line $y = \alpha$ (in the picture, $\alpha = -3$). We denote the intersections $\phi_1(\alpha)$ and $\phi_2(\alpha)$. By construction, we know that $a < 0.5 < b$. By looking at the picture, it is clear that $\ell'(a, \alpha, \beta) > 0$ if $a < \phi_1(\alpha)$ and $\ell'(a, \alpha, \beta) < 0$ if $a > \phi_1(\alpha)$. Analogously, we have $\ell'(b, \alpha, \beta) > 0$ if $b > \phi_2(\alpha)$ and $\ell'(b, \alpha, \beta) < 0$ if $b < \phi_2(\alpha)$.

[53]Taking derivative $\frac{1}{a} + \frac{1}{1-a} - \frac{1}{(r-1)(1-a)^2} = 0$, we obtain the maximizer $a^* = \frac{r-1}{r}$. The function is increasing in $[0, \frac{r-1}{r})$ and decreasing in $(\frac{r-1}{r}, 1]$. The maximum is therefore at $\log(r-1) - \frac{r}{r-1}$.

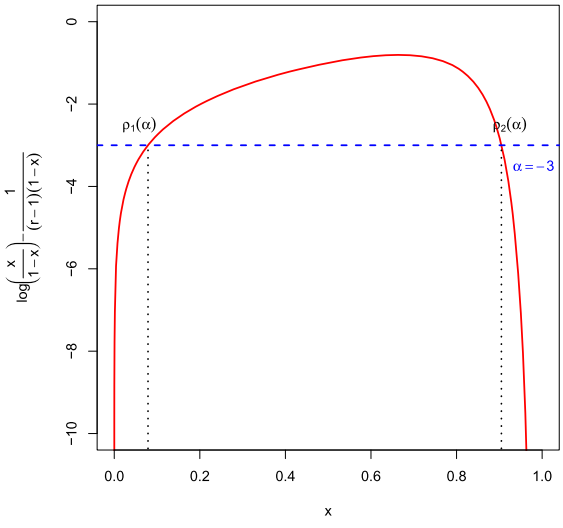

FIGURE 17.—How to compute $\phi_1(\theta_1)$ and $\phi_2(\theta_2)$.

For any $\alpha < -2$, there exist $\phi_1(\alpha)$ and $\phi_2(\alpha)$ which are the intersection of the function $y = \log(\frac{x}{1-x}) - \frac{1}{(r-1)(1-x)}$ with the line $y = \alpha$. Since the function is continuous, monotonic increasing in $[0, \frac{r-1}{r})$, and monotonic decreasing in $(\frac{r-1}{r}, 1]$, it follows that $\phi_1(\alpha)$ and $\phi_2(\alpha)$ are both continuous in α. In addition, $\phi_1(\alpha)$ is increasing in α and $\phi_2(\alpha)$ is decreasing in α. It is trivial to show that when α decreases, $\phi_1(\alpha)$ converges to 0 while $\phi_2(\alpha)$ converges to 1.

Consider the case in which $\ell'(a, \alpha, \beta) < 0 < \ell'(b, \alpha, \beta)$ with two maximizers of $\ell(\mu, \alpha, \beta)$. Consider the function $s(\mu)$ defined above.

Since $\ell'(a, \alpha, \beta) < 0$, we have $a > \phi_1(\alpha)$, which implies $s(a) < s(\phi_1(\alpha))$. Therefore, $\beta < s(\phi_1(\alpha, \beta)) = \frac{1}{r(r-1)\phi_1(\alpha)^{r-1}(1-\phi_1(\alpha))}$.

Since $\ell'(b, \alpha, \beta) > 0$, we have $b > \phi_2(\alpha)$, which implies $s(b) > s(\phi_2(\alpha))$. Therefore, $\beta > s(\phi_2(\alpha)) = \frac{1}{r(r-1)\phi_2(\alpha)^{r-1}(1-\phi_2(\alpha))}$.

Notice that $s(\phi_1(\alpha)) > s(\phi_2(\alpha))$ for any (α, β) in this region of the parameters (see Figure 18 for an example with $\beta = 4$, $\alpha = -2$, and $r = 3$).

The areas are shown in Figure 19 and the rest of the proof follows. The existence of $\zeta(\alpha)$ is shown using similar arguments as in the proof of Theorem 11, so it is omitted for brevity. *Q.E.D.*

The next result is analogous to Theorem 6.3 in Chatterjee and Diaconis (2013), adapted to the directed network model. It shows that not all the specifications of the model generate directed Erdős–Rényi networks. We show this by focusing on a special case.

THEOREM 14: *Consider the model with re-scaled potential $\mathcal{T}(G)$ and with $\beta < 0$,*

$$\mathcal{T}(G) = \alpha \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}}{n^2} + \beta \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} g_{ij}g_{jk}}{n^3}.$$

Then, for any value of α, there exists a positive constant $C(\alpha)$ such that, for $\beta < -C(\alpha)$, the variational problem is not solved at a constant graphon.

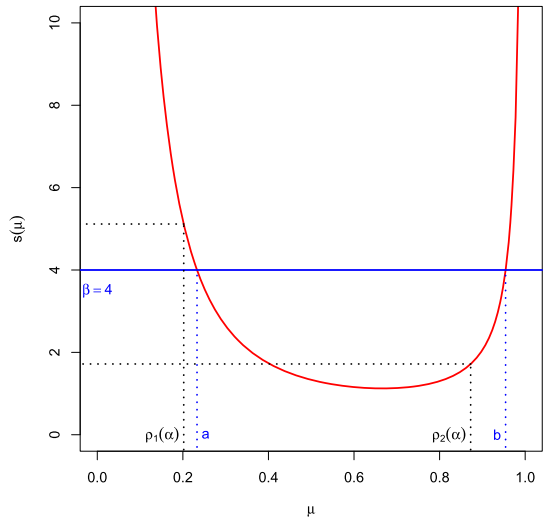

FIGURE 18.—The case with two maximizers.

PROOF: Fix the value of α and let $p = \frac{e^\alpha}{1+e^\alpha}$, and $\lambda = -\beta$. For any h, we have

$$\begin{aligned}\mathcal{T}(h) - \mathcal{I}(h) &= \alpha \int h(x,y)\,dx\,dy + \beta \int h(x,y)h(y,z)\,dx\,dy\,dz \\ &\quad - \int h(x,y)\ln h(x,y) + \big(1-h(x,y)\big)\ln\big(1-h(x,y)\big)\,dx\,dy \\ &= \alpha \int h(x,y)\,dx\,dy + \beta \int h(x,y)h(y,z)\,dx\,dy\,dz\end{aligned}$$

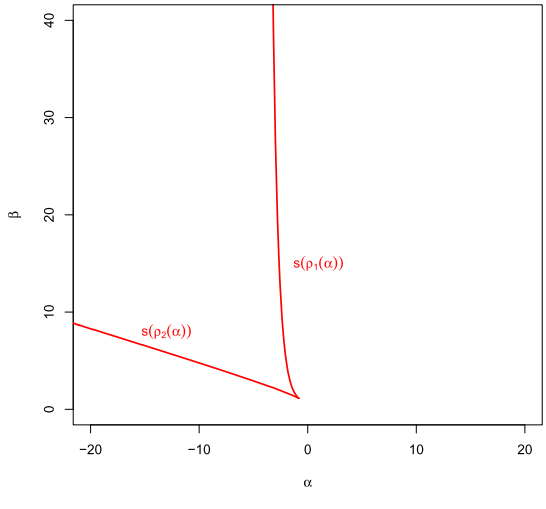

FIGURE 19.—The V-shaped region with two maximizers.

$$
\begin{aligned}
&+ \int h(x,y)\ln\bigl(1+e^{\alpha}\bigr)\,dx\,dy - \int h(x,y)\ln\bigl(1+e^{\alpha}\bigr)\,dx\,dy \\
&- \int h(x,y)\ln h(x,y) + \bigl(1-h(x,y)\bigr)\ln\bigl(1-h(x,y)\bigr)\,dx\,dy \\
&= \beta\int h(x,y)h(y,z)\,dx\,dy\,dz + \int h(x,y)\ln p\,dx\,dy \\
&+ \int h(x,y)\ln(1-p)\,dx\,dy \\
&- \int h(x,y)\ln h(x,y) + \bigl(1-h(x,y)\bigr)\ln\bigl(1-h(x,y)\bigr)\,dx\,dy \\
&= \beta\int h(x,y)h(y,z)\,dx\,dy\,dz + \int h(x,y)\ln p\,dx\,dy \\
&\int h(x,y)\ln(1-p)\,dx\,dy \\
&+ \int \ln(1-p)\,dx\,dy - \int \ln(1-p)\,dx\,dy \\
&- \int h(x,y)\ln h(x,y) + \bigl(1-h(x,y)\bigr)\ln\bigl(1-h(x,y)\bigr)\,dx\,dy \\
&= \beta\int h(x,y)h(y,z)\,dx\,dy\,dz + \ln(1-p) \\
&- \int h(x,y)\ln\frac{h(x,y)}{p} + \bigl(1-h(x,y)\bigr)\ln\frac{1-h(x,y)}{1-p}\,dx\,dy \\
&= -\lambda t(H_2,h) + \ln(1-p) - \mathcal{I}_p(h).
\end{aligned}
$$

We have assumed that $\beta < 0$. Assume that the quantity $\mathcal{T}(h) - \mathcal{I}(h)$ is maximized at a constant graphon $h(x,y) = \mu$. As a consequence, μ *minimizes* the function

$$\lambda t(H_2,h) + \mathcal{I}_p(h) = \lambda\mu^2 + \mathcal{I}_p(\mu).$$

Since μ is the graphon that maximizes $\mathcal{T}(h) - \mathcal{I}(h)$, then we have that, for any $x \in [0,1]$, the following holds: $\lambda\mu^2 + \mathcal{I}_p(\mu) \le \lambda x^2 + \mathcal{I}_p(x)$. The first-order conditions for minimization give

$$v(x) = \frac{d}{dx}\bigl[\lambda x^2 + \mathcal{I}_p(x)\bigr] = 2\lambda x + \ln\frac{x}{1-x} - \ln\frac{p}{1-p}.$$

Notice that $v(0) = -\infty$ and $v(1) = +\infty$; therefore, μ must be an interior minimum. By solving the first-order conditions

$$2\lambda\mu + \ln\frac{\mu}{1-\mu} - \ln\frac{p}{1-p} = 0,$$

it is easy to see that there exists a function $c(\lambda)$ such that

$$\mu = \frac{\exp\left[-2\lambda\mu + \ln\dfrac{p}{1-p}\right]}{1+\exp\left[-2\lambda\mu + \ln\dfrac{p}{1-p}\right]} \leq c(\lambda).$$

So we get $\mu \leq c(\lambda)$, where $c(\lambda)$ is a function such that

$$\lim_{\lambda\to\infty} c(\lambda) = 0,$$

and therefore, it follows that

$$\lim_{\lambda\to\infty} \min_{x\in[0,1]} \lambda x^2 + \mathcal{I}_p(x) = \mathcal{I}_p(0) = \ln\frac{1}{1-p}.$$

We will now show that there exists a graphon $\nu(x, y)$ which is not a constant and gives a lower value of the expression above.

Let $\nu(x, y)$ be the function

$$\nu(x,y) = \begin{cases} p, & \text{if } x \in [0, 0.5] \text{ and } y \in [0.5, 1], \\ 0, & \text{otherwise.} \end{cases}$$

It follows that, for almost all (x, y, z) triplets, $\nu(x, y)\nu(y, z) = 0$ and thus, $t(H_2, \nu) = 0$. If we compute the value of $\mathcal{I}_p(\nu)$, we obtain

$$\begin{aligned}
\mathcal{I}_p(\nu) = &\int_{[\frac{1}{2},1]\times[0,\frac{1}{2}]} 0\ln\frac{0}{p} + \ln\frac{1}{1-p}\,dx\,dy \\
&+ \int_{[0,\frac{1}{2}]\times[0,\frac{1}{2}]} 0\ln\frac{0}{p} + \ln\frac{1}{1-p}\,dx\,dy \\
&+ \int_{[0,\frac{1}{2}]\times[\frac{1}{2},1]} p\ln\frac{p}{p} + (1-p)\ln\frac{1-p}{1-p}\,dx\,dy \\
&+ \int_{[\frac{1}{2},1]\times[\frac{1}{2},1]} 0\ln\frac{0}{p} + \ln\frac{1}{1-p}\,dx\,dy \\
= &\frac{3}{4}\ln\frac{1}{1-p}.
\end{aligned}$$

Therefore, we have shown that for λ large enough (i.e., for β negative and large enough), $\mathcal{T}(\nu) - \mathcal{I}(\nu) \geq \mathcal{T}(\mu) - \mathcal{I}(\mu)$. So, given a value for α, there exists a $C(\alpha)$ large enough such that, for any $\beta < -C(\alpha)$, a constant graphon is not solution to the variational problem. *Q.E.D.*

This result extends to models with two parameters and higher-order dependencies, as shown in the next theorem.

THEOREM 15: *For the models in the first part of Theorem 12, the result of Theorem 14 holds.*

PROOF: The proof is equivalent to the proof of Theorem 14, replacing μ^2 with μ^r, where r is the order of dependence of the second homomorphism density $t(H_2, h)$. *Q.E.D.*

THEOREM 16: *Consider the model with re-scaled potential $\mathcal{T}(G)$ and with $\beta < 0$,*

$$\mathcal{T}(G) = \alpha \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}}{n^2} + \beta \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} g_{ij}g_{jk}}{n^3} + \gamma \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} g_{ij}g_{jk}g_{ki}}{n^3}. \tag{64}$$

Then, for any value of $\alpha \in \mathbb{R}$ and $\gamma > 0$, there exists a positive constant $C(\alpha, \gamma) > 0$ such that, for $\beta < -C(\alpha, \gamma)$, the variational problem is not solved at a constant graphon. Analogously, if $\gamma < 0$, then, for any value of $\alpha \in \mathbb{R}$ and $\beta > 0$, there exists a positive constant $C(\alpha, \beta) > 0$ such that, for $\gamma < C(\alpha, \gamma)$, the variational problem is not solved at a constant graphon.

PROOF: Fix the value of α and $\gamma > 0$. Let $p = \frac{e^\alpha}{1+e^\alpha}$, and $\lambda = -\beta$. For any h, we have

$$\begin{aligned}
\mathcal{T}(h) - \mathcal{I}(h) &= \alpha \int h(x,y)\,dx\,dy + \beta \int h(x,y)h(y,z)\,dx\,dy\,dz \\
&\quad + \gamma \int h(x,y)h(y,z)h(z,x)\,dx\,dy\,dz \\
&\quad - \int h(x,y)\ln h(x,y) + \big(1-h(x,y)\big)\ln\big(1-h(x,y)\big)\,dx\,dy \\
&= \alpha \int h(x,y)\,dx\,dy + \beta \int h(x,y)h(y,z)\,dx\,dy\,dz \\
&\quad + \gamma \int h(x,y)h(y,z)h(z,x)\,dx\,dy\,dz \\
&\quad + \int h(x,y)\ln\big(1+e^\alpha\big)\,dx\,dy - \int h(x,y)\ln\big(1+e^\alpha\big)\,dx\,dy \\
&\quad - \int h(x,y)\ln h(x,y) + \big(1-h(x,y)\big)\ln\big(1-h(x,y)\big)\,dx\,dy \\
&= \beta \int h(x,y)h(y,z)\,dx\,dy\,dz + \gamma \int h(x,y)h(y,z)h(z,x)\,dx\,dy\,dz \\
&\quad + \int h(x,y)\ln p\,dx\,dy + \int h(x,y)\ln(1-p)\,dx\,dy \\
&\quad - \int h(x,y)\ln h(x,y) + \big(1-h(x,y)\big)\ln\big(1-h(x,y)\big)\,dx\,dy \\
&= \beta \int h(x,y)h(y,z)\,dx\,dy\,dz + \gamma \int h(x,y)h(y,z)h(z,x)\,dx\,dy\,dz \\
&\quad + \int h(x,y)\ln p\,dx\,dy + \int h(x,y)\ln(1-p)\,dx\,dy \\
&\quad + \int \ln(1-p)\,dx\,dy - \int \ln(1-p)\,dx\,dy
\end{aligned}$$

$$
\begin{aligned}
&- \int h(x,y)\ln h(x,y) + \big(1-h(x,y)\big)\ln\big(1-h(x,y)\big)\,dx\,dy \\
&= \beta \int h(x,y)h(y,z)\,dx\,dy\,dz + \gamma \int h(x,y)h(y,z)h(z,x)\,dx\,dy\,dz \\
&\quad + \ln(1-p) - \int h(x,y)\ln\frac{h(x,y)}{p} + \big(1-h(x,y)\big)\ln\frac{1-h(x,y)}{1-p}\,dx\,dy \\
&= \beta t(H_2,h) + \gamma t(H_3,h) + \ln(1-p) - \mathcal{I}_p(h).
\end{aligned}
$$

We have assumed that $\beta < 0$. Assume that the quantity $\mathcal{T}(h) - \mathcal{I}(h)$ is maximized at a constant graphon $h(x,y) = \mu$. As a consequence, μ *maximizes* the function

$$\beta t(H_2,h) + \gamma t(H_3,h) - \mathcal{I}_p(h) = \beta\mu^2 + \gamma\mu^3 - \mathcal{I}_p(\mu).$$

Since μ is the graphon that maximizes $\mathcal{T}(h) - \mathcal{I}(h)$, then we have that for any $x \in [0,1]$, the following holds: $\beta\mu^2 + \gamma\mu^3 - \mathcal{I}_p(\mu) \geq \beta x^2 + \gamma x^3 - \mathcal{I}_p(x)$. The first-order conditions for maximization give

$$v(x) = \frac{d}{dx}\big[\beta x^2 + \gamma x^3 - \mathcal{I}_p(x)\big] = 2\beta x + 3\gamma x^2 - \ln\frac{x}{1-x} + \ln\frac{p}{1-p}.$$

Notice that $v(0) = +\infty$ and $v(1) = -\infty$; therefore, μ must be an interior maximum. By solving the first-order conditions

$$2\beta\mu + 3\gamma\mu^2 - \ln\frac{\mu}{1-\mu} + \ln\frac{p}{1-p} = 0,$$

it is easy to see that there exists a function $c(\beta,\gamma)$ such that

$$\mu = \frac{\exp\left[2\beta\mu + 3\gamma\mu^2 - \ln\dfrac{p}{1-p}\right]}{1+\exp\left[2\beta\mu + 3\gamma\mu^2 - \ln\dfrac{p}{1-p}\right]} \leq c(\beta,\gamma).$$

So we get $\mu \leq c(\beta,\gamma)$, and $c(\beta,\gamma)$ is a function such that

$$\lim_{\beta\to-\infty} c(\beta,\gamma) = 0,$$

and therefore, it follows that

$$\lim_{\beta\to-\infty} \min_{x\in[0,1]} \beta x^2 + \gamma x^3 - \mathcal{I}_p(x) = -\mathcal{I}_p(0) = -\ln\frac{1}{1-p}.$$

We will now show that there exists a graphon $\nu(x,y)$ which is not a constant and gives a lower value of the expression above.

Let $\nu(x,y)$ be the function

$$\nu(x,y) = \begin{cases} p, & \text{if } x \in \left[0,\dfrac{1}{2}\right] \text{ and } y \in \left[\dfrac{1}{2},1\right], \\ 0, & \text{otherwise.} \end{cases}$$

It follows that for almost all (x, y, z) triplets, $\nu(x, y)\nu(y, z) = 0$ and $\nu(x, y)\nu(y, z) \times \nu(z, x) = 0$. As a consequence, $t(H_2, \nu) = 0$ and $t(H_3, \nu) = 0$. If we compute the value of $\mathcal{I}_p(\nu)$, we obtain

$$\begin{aligned}\mathcal{I}_p(\nu) = &\int_{[\frac{1}{2},1]\times[0,\frac{1}{2}]} 0 \ln \frac{0}{p} + \ln \frac{1}{1-p}\, dx\, dy \\ &+ \int_{[0,\frac{1}{2}]\times[0,\frac{1}{2}]} 0 \ln \frac{0}{p} + \ln \frac{1}{1-p}\, dx\, dy \\ &+ \int_{[0,\frac{1}{2}]\times[\frac{1}{2},1]} p \ln \frac{p}{p} + (1-p) \ln \frac{1-p}{1-p}\, dx\, dy \\ &+ \int_{[\frac{1}{2},1]\times[\frac{1}{2},1]} 0 \ln \frac{0}{p} + \ln \frac{1}{1-p}\, dx\, dy \\ = &\frac{3}{4} \ln \frac{1}{1-p}.\end{aligned}$$

Therefore, we have shown that for $\beta < 0$ large enough in magnitude, $\mathcal{T}(\nu) - \mathcal{I}(\nu) \geq \mathcal{T}(\mu) - \mathcal{I}(\mu)$. So, given a value of $\alpha \in \mathbb{R}$ and $\gamma > 0$, there exists a positive constant $C(\alpha, \gamma) > 0$ such that, for $\beta < -C(\alpha, \gamma)$, a constant graphon is not solution to the variational problem (55) for the model in (64). The proof for $\gamma < 0$ follows the same steps. *Q.E.D.*

THEOREM 17: *Fix parameter $\gamma > 0$. Let the variational problem be described as*

$$\lim_{n\to\infty} \psi_n(\theta) = \psi(\theta) = \sup_{\mu\in[0,1]} \left\{\alpha\mu + \beta\mu^2 + \gamma\mu^3 - \mu \log \mu - (1-\mu)\log(1-\mu)\right\}.$$

Let μ_0 be (uniquely) determined by

$$6\gamma = \frac{2\mu_0 - 1}{\mu_0^2(1-\mu_0)^2},$$

and let α_0,β_0 be defined as follows:

$$\begin{aligned}\beta_0 &= \frac{1}{2\mu_0(1-\mu_0)} - 3\gamma\mu_0, \\ \alpha_0 &= \log \frac{\mu_0}{1-\mu_0} - \frac{1}{(1-\mu_0)} + \frac{2\mu_0 - 1}{2(1-\mu_0)^2}.\end{aligned}$$

1. *If $\beta \leq \beta_0$, the maximization problem has a unique maximizer $\mu^* \in [0, 1]$.*
2. *If $\beta > \beta_0$ and $\alpha \geq \alpha_0$, then there is a unique maximizer $\mu^* > 0.5$.*
3. *If $\beta > \beta_0$ and $\alpha < \alpha_0$, then there are two functions $S_\gamma(\phi_1(\alpha))$ and $S_\gamma(\phi_2(\alpha))$ that define a V-shaped region of parameters (α, β) such that*

(a) *inside the V-shaped region, the maximization problem has two local maximizers $\mu_1^* < 0.5 < \mu_2^*$;*

(b) *outside the V-shaped region, the maximization problem has a unique maximizer μ^*.*

4. *For any $\alpha < \alpha_0$ inside the V-shaped region, there exists a function $\beta = \zeta_\gamma(\alpha)$, such that $S_\gamma(\phi_1(\alpha)) < \zeta_\gamma(\alpha) < S_\gamma(\phi_2(\alpha))$ and the two maximizers are both global.*

PROOF: Fix $\gamma > 0$ and consider the function

$$\ell_\gamma(\mu, \alpha, \beta) = \alpha\mu + \beta\mu^2 + \gamma\mu^3 - \mu\log\mu - (1-\mu)\log(1-\mu).$$

For the moment, we do not constrain β to be positive. The first- and second-order derivatives w.r.t. μ are

$$\ell'_\gamma(\mu, \alpha, \beta) = \alpha + 2\beta\mu + 3\gamma\mu^2 - \ln\left(\frac{\mu}{1-\mu}\right),$$

$$\ell''_\gamma(\mu, \alpha, \beta) = 2\beta + 6\gamma\mu - \frac{1}{\mu(1-\mu)}.$$

The function $\ell_\gamma(\mu, \alpha, \beta)$ is concave if $\ell''_\gamma(\mu, \alpha, \beta) < 0$, that is, when

$$2\beta + 6\gamma\mu < \frac{1}{\mu(1-\mu)} \equiv s(\mu).$$

The function $s(\mu)$ is decreasing in $[0, 0.5)$ and increasing in $(0.5, 1]$, and it has a minimum at $\mu = 0.5$, where $s(0.5) = 4$.

Let μ_0 be the value of μ at which the line $2\beta + 6\gamma\mu$ is tangent to $s(\mu)$, defined as the solution of

$$6\gamma = \frac{2\mu - 1}{\mu^2(1-\mu)^2}.$$

Notice that μ_0 is unique, since the right-hand side of the equation is a monotone increasing function. Given μ_0, we can find β_0 by solving

$$\beta_0 = \frac{1}{2}\left[-6\gamma\mu_0 + \frac{1}{\mu_0(1-\mu_0)}\right].$$

Therefore, the function $\ell_\gamma(\mu, \alpha, \beta)$ is concave on the whole interval $[0, 1]$ if $\beta \le \beta_0$. In this region, there is a unique maximizer μ^* of $\ell_\gamma(\mu, \alpha, \beta)$.

If $\beta > \beta_0$, the line $2\beta + 6\gamma\mu$ has two intersections with $s(\mu)$, and there are three possible cases. We know that in this region the second derivative $\ell''_\gamma(\mu, \alpha, \beta)$ can be positive or negative, with inflection points denoted as a and b, found by solving the equation $2\beta + 6\gamma\mu = s(\mu)$. In Figure 20, we plot $s(\mu)$ (in red), the line $2\beta + 6\gamma\mu$ (blue dashed) that define the points a and b, and the tangent line (black solid) that defines μ_0.

By looking at the picture, it is clear that the first derivative $\ell'_\gamma(\mu, \alpha, \beta)$ is decreasing for $\mu \in [0, a)$, increasing in $\mu \in (a, b)$, and decreasing in $\mu \in (b, 1]$.

1. If $\ell'_\gamma(a, \alpha, \beta) \ge 0$, then there is a unique maximizer $\mu^* > b$.
2. If $\ell'_\gamma(b, \alpha, \beta) \le 0$, then there is a unique maximizer $\mu^* < a$.
3. If $\ell'_\gamma(a, \alpha, \beta) < 0 < \ell'_\gamma(b, \alpha, \beta)$, then there are two local maximizers $\mu_1^* < a < b < \mu_2^*$.

The three cases are shown in Figure 21, where we plot $\ell'_\gamma(\mu, \alpha, \beta)$ against μ for several values of α and for a fixed $\beta = 1$ and $\gamma = 1.5$.

We indicate the maximizer with μ^* when it is unique, and with μ_1^*, μ_2^* when there are two.

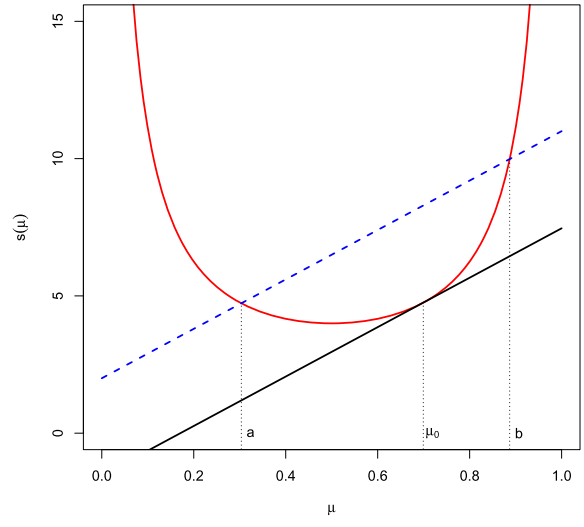

FIGURE 20.—Graphical explanation of the concavity for $\ell_\gamma(\mu, \alpha, \beta)$.

Let us consider the first case, with $\ell'_\gamma(a, \alpha, \beta) \geq 0$. To compute $\ell'_\gamma(a, \alpha, \beta)$, notice that

$$\beta = \frac{1}{2a(1-a)} - \frac{2\mu_0 - 1}{2\mu_0^2(1-\mu_0)^2}a.$$

Substituting in $\ell'_\gamma(a, \alpha, \beta)$, we obtain

$$\begin{aligned}\ell'_\gamma(a, \alpha, \beta) &= \alpha + \frac{a}{a(1-a)} - \frac{2\mu_0 - 1}{\mu_0^2(1-\mu_0)^2}a^2 + \frac{2\mu_0 - 1}{2\mu_0^2(1-\mu_0)^2}a^2 - \log\frac{a}{1-a} \\ &= \alpha + \frac{1}{(1-a)} - \frac{2\mu_0 - 1}{2\mu_0^2(1-\mu_0)^2}a^2 - \log\frac{a}{1-a},\end{aligned}$$

and analogously we have for b

$$\ell'_\gamma(b, \alpha, \beta) = \alpha + \frac{1}{(1-b)} - \frac{2\mu_0 - 1}{2\mu_0^2(1-\mu_0)^2}b^2 - \log\frac{b}{1-b}.$$

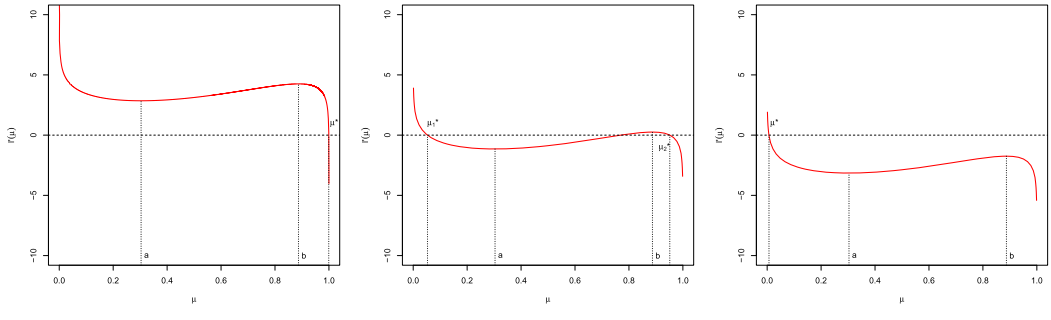

FIGURE 21.—The three possible cases for the solution of the first order condition $\ell'(\mu, \alpha, \beta)$.

Notice that we can write $\ell'_\gamma(a,\alpha,\beta) = \alpha + \eta(a)$, where $\eta(a) = \frac{1}{(1-a)} - \frac{2\mu_0-1}{2\mu_0^2(1-\mu_0)^2}a^2 - \log\frac{a}{1-a}$. Consider the derivative of $\eta(a)$:

$$\begin{aligned}\eta'(a) &= \frac{1}{(1-a)^2} - \frac{2\mu_0-1}{\mu_0^2(1-\mu_0)^2}a - \frac{1}{a(1-a)} \\ &= a\left[\frac{2a-1}{a^2(1-a)^2} - \frac{2\mu_0-1}{\mu_0^2(1-\mu_0)^2}\right].\end{aligned}$$

We know that the function $\mathfrak{h}(a) = \frac{2a-1}{a^2(1-a)^2}$ is monotone increasing, with $\mathfrak{h}(0) = -\infty$ and $\mathfrak{h}(1) = \infty$. Therefore, the minimum of $\eta(a)$ is found at $a = \mu_0$, where we have

$$\eta(\mu_0) = \frac{1}{(1-\mu_0)} - \frac{2\mu_0-1}{2(1-\mu_0)^2} - \log\frac{\mu_0}{1-\mu_0}.$$

This means that $\ell'_\gamma(a,\alpha,\beta) \geq 0$ only if

$$\alpha \geq \alpha_0 = -\eta(\mu_0) = \log\frac{\mu_0}{1-\mu_0} - \frac{1}{(1-\mu_0)} + \frac{2\mu_0-1}{2(1-\mu_0)^2}.$$

When the above condition is satisfied, there is a unique maximizer, $\mu^* > b$, as shown in the picture on the left.

When $\alpha < \alpha_0$ and $\beta > \beta_0$, we have $\ell'_\gamma(a,\alpha,\beta) < 0 < \ell'(b,\alpha,\beta)$. We draw a picture of $-\eta(\mu)$ to help with the reasoning (Figure 22).

Notice that when $\alpha < \alpha_0$, there are two intersections of the function and the horizontal line $y = \alpha$ (in the picture, $\alpha = -3$). We denote the intersections $\phi_1(\alpha)$ and $\phi_2(\alpha)$. By construction, we know that $a < 0.5 < b$. By looking at the picture, it is clear that $\ell'_\gamma(a,\alpha,\beta) > 0$ if $a < \phi_1(\alpha)$ and $\ell'_\gamma(a,\alpha,\beta) < 0$ if $a > \phi_1(\alpha)$. Analogously, we have $\ell'_\gamma(b,\alpha,\beta) > 0$ if $b > \phi_2(\alpha)$ and $\ell'_\gamma(b,\alpha,\beta) < 0$ if $b < \phi_2(\alpha)$.

For any $\alpha < \alpha_0$, there exist $\phi_1(\alpha)$ and $\phi_2(\alpha)$ which are the intersections of the function $-\eta(\mu)$ with the line α. Since the function is continuous, monotonic increasing in $[0,\mu_0)$, and monotonic decreasing in $(\mu_0,1]$, it follows that $\phi_1(\alpha)$ and $\phi_2(\alpha)$ are both continuous

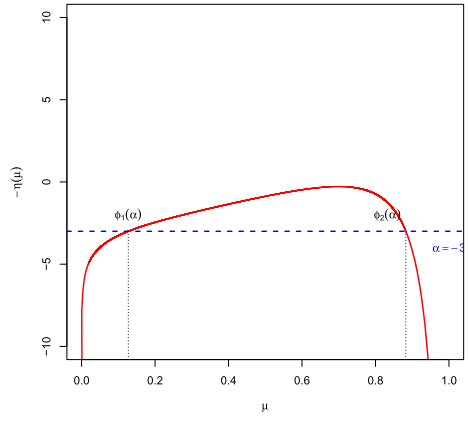

FIGURE 22.—Computation of $\phi_1(\alpha)$ and $\phi_2(\alpha)$.

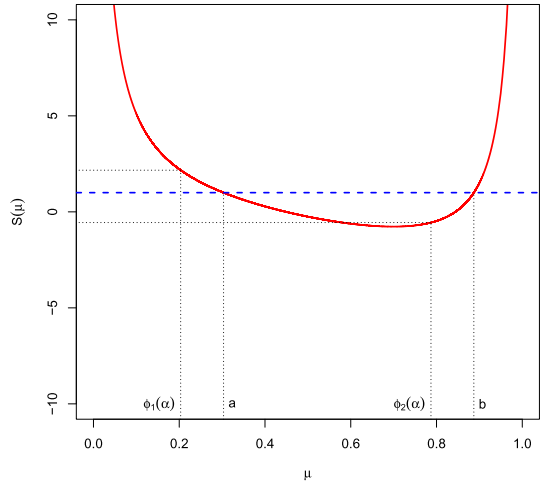

FIGURE 23.—Proof that $S(\phi_1(\alpha)) > S(\phi_2(\alpha))$ when there are two maximizers.

in α. In addition, $\phi_1(\alpha)$ is increasing in α and $\phi_2(\alpha)$ is decreasing in α. It is trivial to show that when α decreases, $\phi_1(\alpha)$ converges to 0 while $\phi_2(\alpha)$ converges to 1.

Consider the case in which $\ell'_\gamma(a, \alpha, \beta) < 0 < \ell'_\gamma(b, \alpha, \beta)$ with two maximizers of $\ell_\gamma(\mu, \alpha, \beta)$. Consider the function

$$S(\mu) = \frac{1}{2\mu(1-\mu)} - \frac{2\mu_0 - 1}{2\mu_0^2(1-\mu_0)^2}\mu.$$

Since $\ell'_\gamma(a, \alpha, \beta) < 0$, we have $a > \phi_1(\alpha)$, which implies $S(a) < S(\phi_1(\alpha))$. Therefore, $\beta < S(\phi_1(\alpha)) = \frac{1}{2\phi_1(\alpha)(1-\phi_1(\alpha))} - \frac{2\mu_0-1}{2\mu_0^2(1-\mu_0)^2}\phi_1(\alpha)$.

Since $\ell'_\gamma(b, \alpha, \beta) > 0$, we have $b > \phi_2(\alpha)$, which implies $S(b) > S(\phi_2(\alpha))$. Therefore, $\beta > S(\phi_2(\alpha)) = \frac{1}{2\phi_2(\alpha)(1-\phi_2(\alpha))} - \frac{2\mu_0-1}{2\mu_0^2(1-\mu_0)^2}\phi_2(\alpha)$.

Notice that $S(\phi_1(\alpha)) > S(\phi_2(\alpha))$ for any (α, β) in this region of the parameters (see Figure 23 for an example with $\beta = 1$, $\alpha = -3$, and $\gamma = 1.5$).

In Figure 24, we show the function $S(\phi_1(\alpha))$ and $S(\phi_2(\alpha))$ in the (α, β) space, for a given $\gamma > 0$. Notice that for our models, we are only interested in the part of the graph where $\beta > 0$. The graphs show that when we increase the value of γ, the area in which the model has multiple local maxima increases.

The existence of $\zeta_\gamma(\alpha)$ is shown using similar arguments as in the proof of Theorem 11, so it is omitted for brevity. *Q.E.D.*

The last set of results extends the analysis of sampling algorithms in Bhamidi, Bresler, and Sly (2011) to directed graphs. In particular, the solution to the variational problems in the previous theorems provides a characterization for the convergence of the MCMC samplers commonly used to simulate samples of ERGMs from the model. The set of parameters that lie within the V-shaped region correspond to what Bhamidi, Bresler, and Sly (2011) called the *low temperature phase*. The set of parameters lying outside the V-shaped region correspond to the *high temperature phase*.

To be precise, let $\widetilde{M}^* \subset \widetilde{\mathcal{W}}$ be the set of maximizers of the variational problem and let G_n be a graph on n vertices drawn from the ERGM model implied by function $\mathcal{T}$. The

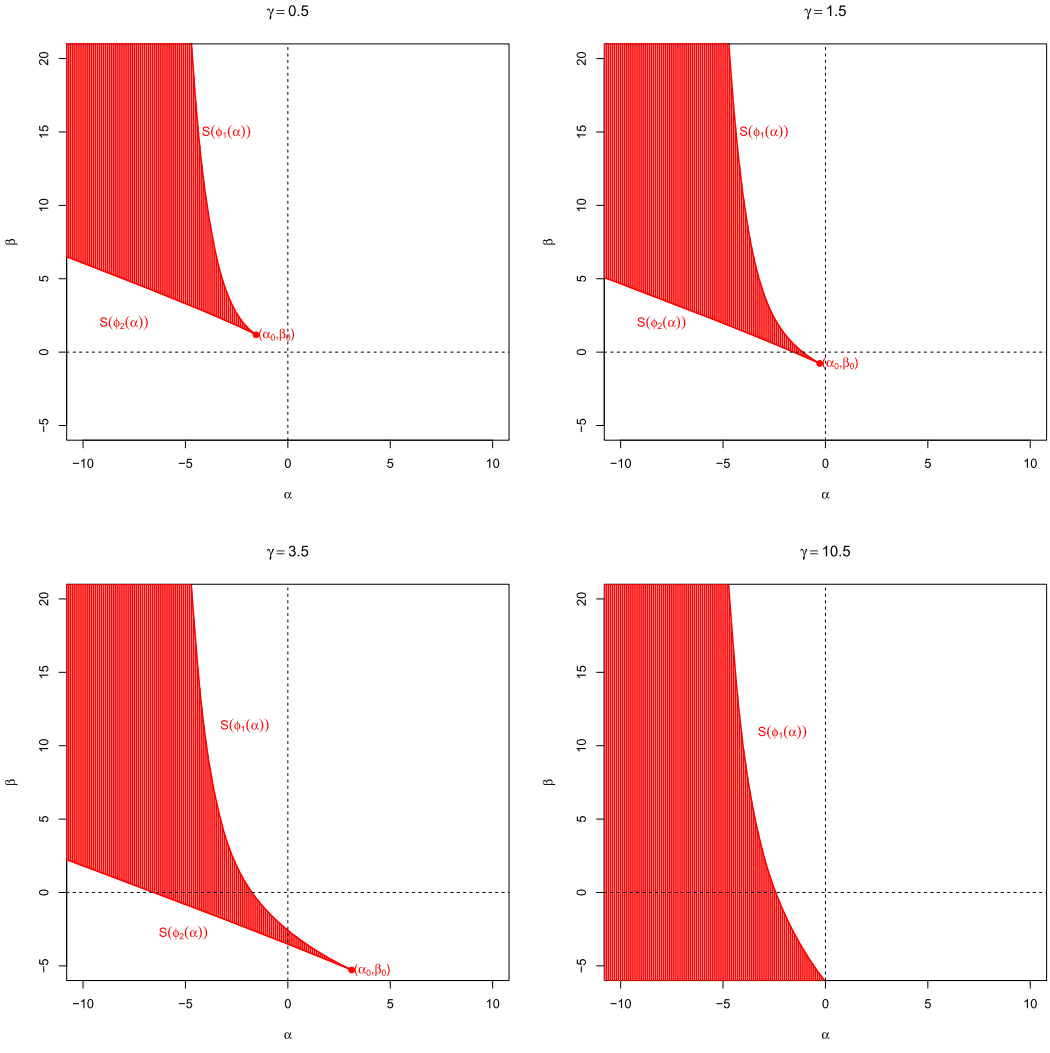

FIGURE 24.—The V-shaped region of the parameters space with two local maximizers, as a function of $\gamma > 0$.

next theorem shows that as n grows large, the network $\widetilde{G}_n$ must be close to the set $\widetilde{M}^*$. If the set consists of a single graph, then this is equivalent to a weak law of large numbers for G_n.

THEOREM 18: *Let $\widetilde{M}^*$ be the set of maximizers of the variational problem (55). Let G_n be a graph on n vertices drawn from the model implied by function $\mathcal{T}$. Then, for any $\eta > 0$, there exist $C, \kappa > 0$ such that, for any n,*

$$\mathbb{P}\big(\delta_{\square}\big(\widetilde{G}_n, \widetilde{M}^*\big) > \eta\big) \leq Ce^{-n^2\kappa},$$

where $\mathbb{P}$ denotes the probability measure implied by the model.

PROOF: The proof is identical to the proof of Theorem 3.2 in Chatterjee and Diaconis (2013). *Q.E.D.*

For the model we analyze in this paper, the result specializes to the following theorem.

THEOREM 19: *Consider the model above in* (56) *and assume* $\theta_2 > 0$. *Let* G_n *be the directed graph implied by the model*.

1. *If the maximization problem in Theorem* 11 *has a unique solution* μ^*, *then* $G_n \to G_d(n, \mu^*)$ *in probability as* $n \to \infty$.

2. *If the maximization problem in Theorem* 11 *has two solutions* $\mu_1^* < \frac{1}{2} < \mu_2^*$, *then* G_n *is drawn from a mixture of directed Erdős–Rényi graphs* $G_d(n, \mu_1^*)$ *and* $G_d(n, \mu_2^*)$, *as* $n \to \infty$.

PROOF: It is an application of Theorem 18. *Q.E.D.*

The previous results consider the limit as $n \to \infty$. However, for fixed n, the speed of convergence of the model to the stationary distribution π_n can be studied using the previous results. The model evolves according to a Glauber dynamics: essentially it behaves like a random Gibbs sampler.

In particular, when the maximization problem in Theorem 11 has a unique solution, the Markov chain of networks converges in an order $n^2 \log n$ steps. However, when the maximization problem in Theorem 11 has two solutions $\mu_1^* < \frac{1}{2} < \mu_2^*$, the convergence is exponentially slow, that is, there exists a constant $C > 0$ such that the number of steps needed to reach stationarity is $O(e^{Cn})$. This is true for any local chain, that is, a chain that updates $o(n)$ links per iteration.

The main convergence result that is proven in Bhamidi, Bresler, and Sly (2011) is extended to our directed network formation model in the following proposition.

PROPOSITION 3—Convergence Rates: *Assume* $\beta, \gamma > 0$ *in any of the models in Theorem* 12.

1. *If the variational problem has a unique solution*, *we say that the parameters belong to the* high temperature *region*. *The chain of networks generated by the model mixes in order* $n^2 \log n$ *steps*.

2. *If the variational problem has two local maxima*, *we say that the parameters belong to the* low temperature *region*. *The chain of networks generated by the model mixes in order* e^{n^2} *steps*. *This holds for any local dynamics*, *that is*, *a dynamics that updates an* $o(n)$ *number of links per period*.

PROOF: See Bhamidi, Bresler, and Sly (2011), Theorems 5 and 6. *Q.E.D.*

The main reason for the slow convergence in the bi-modal regime is that a local chain makes small steps. The solution to this problem is to allow the sampler to perform larger steps. However, large steps are not sufficient. Indeed, we need to be able to make large steps of order n; in other words, we need a large step whose size is a function of n.

The result of asymptotically independent edges (Theorem 7 in Bhamidi, Bresler, and Sly (2011)) is proven above in our Theorem 19.

APPENDIX E: ADDITIONAL SIMULATION RESULTS

The performance of the estimation method is tested using artificial data. All the computations with artificial data are performed in a standard desktop Dell Precision T7620 with two Intel Xeon CPUs E5-2697 v2 with 12 Dual core processors at 2.7 GHZ each and 64 GB of RAM. For replication purposes, there is a package in Github at

https://github.com/meleangelo/netnew.[54] Ideally, we want to compare the results of the approximate exchange algorithm with the exact algorithm. This is feasible for a special case, where preferences depend only on direct and mutual links (i.e., excluding friends of friends and popularity effects):

$$Q(g,\alpha,\beta)=\alpha\sum_{i=1}^{n}\sum_{j=1}^{n}g_{ij}+\beta\sum_{i=1}^{n}\sum_{j>i1}^{n}g_{ij}g_{ji}. \tag{65}$$

For this model, described by equation (65), we can show that the constant is

$$c(\theta)=\left(1+2e^{\alpha}+e^{2\alpha+\beta}\right)^{\frac{n(n-1)}{2}};$$

thus, we can compute the exact likelihood and we can perform inference using the *exact* Metropolis–Hastings sampler. We then compare the results of the exact algorithm with the approximate exchange algorithm.

The results of the simulations are shown in Table I. The data were generated by parameters $(\alpha,\beta)=(-2.0,0.5)$. The numbers of network simulations per each proposed parameter are $R=\{1000,\ 5000,\ 10{,}000,\ 50{,}000,\ 100{,}000,\ 1{,}000{,}000,\ 10{,}000{,}000\}$. We run each algorithm for $S=10{,}000$ parameter iterations, and we use the output to measure the Kolmogorov–Smirnov distance and the Kullback–Leibler divergence between the posterior estimated with the exact metropolis sampler $p(\theta|g,X)$ and the posterior estimated with the approximated algorithm with R network simulations $p_R(\theta|g,X)$:

$$\mathrm{KS}=\sup_{\theta_i\in\Theta_i}\left|\int_{-\infty}^{\theta_i}p_R(\theta_i|g,X)-\int_{-\infty}^{\theta_i}p(\theta_i|g,X)\right|,$$

$$\mathrm{KL}=\int_{\Theta_i}\log\left[\frac{p_R(\theta_i|g,X)}{p(\theta_i|g,X)}\right]p_R(\theta_i|g,X)\,d\theta_i.$$

The table reports posterior mean, median, standard deviation, Monte Carlo standard errors for the posterior mean (mcse), 95% credibility intervals, Kolmogorov–Smirnov statistics, Kullback–Leibler divergence, and time for computation.

The exact Metropolis–Hastings is reported in the first column of the table. The approximate exchange algorithm works very well for small to moderate networks. For a small network with $n=100$ players, a reasonable degree of accuracy can be reached with as low as $R=5000$ network simulations per parameter. Simulations from over-dispersed starting values converge to the same posterior distribution. Convergence is quite fast to the high density region of the posterior.[55]

[54]In all estimation exercises, we use independent normal priors $\mathcal{N}(0,10)$. The proposal of the exchange algorithm is a random walk $\mathcal{N}(0,\Sigma)$. We repeat the estimation twice: the first time, we use a diagonal Σ; in the second round, we use the covariance from the first round as baseline. In all simulations, the probability of large steps is 0.001 and a large step updates $0.1n$ links.

[55]This result is common with the class of exchange algorithms. See Caimo and Friel (2011), Atchade and Wang (2014) for examples. Computations can be faster if we embed sparse matrix algebra routines in the codes. The results in Table I are obtained with codes that do not use sparse matrix algebra, thus representing a worst case scenario in computational time.

TABLE I

CONVERGENCE OF ESTIMATED POSTERIORS, MODEL (65)

	Exact Metropolis		$R = 1000$		$R = 5000$		$R = 10{,}000$		$R = 50{,}000$		$R = 100{,}000$		$R = 1$ mil		$R = 10$ mil	
	α	β	α	β	α	β	α	β	α	β	α	β	α	β	α	β
							$n = 100$									
Mean	−1.923	0.286	−1.915	0.286	−1.925	0.285	−1.922	0.275	−1.921	0.284	−1.919	0.286				
Median	−1.923	0.288	−1.919	0.296	−1.923	0.292	−1.921	0.273	−1.921	0.287	−1.919	0.286				
Std. dev.	0.034	0.114	0.105	0.263	0.054	0.141	0.042	0.123	0.034	0.115	0.034	0.111				
Mcse	0.000	0.002	0.007	0.033	0.001	0.005	0.001	0.004	0.000	0.003	0.000	0.003				
Pctile 2.5%	−1.992	0.058	−2.115	−0.257	−2.034	−0.007	−2.006	0.034	−1.987	0.039	−1.985	0.069				
Pctile 97.5%	−1.857	0.506	−1.705	0.767	−1.820	0.553	−1.842	0.514	−1.853	0.505	−1.851	0.512				
KS	NA	NA	0.275	0.205	0.114	0.057	0.066	0.057	0.032	0.015	0.060	0.022				
KL	NA	NA	0.041	0.027	0.013	0.186	0.039	0.075	0.040	0.062	0.006	0.088				
							$n = 200$									
Mean	−1.988	0.463	−1.975	0.463	−1.964	0.463	−1.979	0.465	−1.989	0.455	−1.988	0.463	−1.988	0.466	−1.987	0.459
Median	−1.989	0.467	−1.974	0.509	−1.968	0.468	−1.978	0.465	−1.989	0.454	−1.989	0.464	−1.988	0.467	−1.987	0.457
Std. dev.	0.017	0.061	0.048	0.275	0.042	0.113	0.033	0.073	0.019	0.053	0.017	0.059	0.017	0.057	0.016	0.055
Mcse	0	0.003	0.002	0.075	0.002	0.012	0.001	0.005	0	0.002	0	0.003	0	0.003	0	0.002
Pctile 2.5%	−2.021	0.335	−2.071	−0.21	−2.044	0.186	−2.042	0.32	−2.024	0.353	−2.024	0.339	−2.024	0.358	−2.017	0.354
Pctile 97.5%	−1.954	0.572	−1.89	0.889	−1.872	0.687	−1.921	0.614	−1.949	0.56	−1.955	0.571	−1.955	0.582	−1.955	0.577
KS	NA	NA	0.343	0.34	0.381	0.135	0.25	0.057	0.067	0.105	0.015	0.039	0.039	0.045	0.062	0.086
KL	NA	NA	0.1	0.178	0.105	0.079	0.129	0.099	0.05	0.05	0.041	0.058	0.092	0.044	0.173	0.234
Time	0.124 s		14.539 s		21.808 s		30.451 s		100.761 s		193.722 s		1762.202 s		17,370.945 s	
							$n = 500$									
Mean	−2.018	0.551	−1.941	0.337	−2.014	0.562	−2.017	0.561	−2.017	0.552	−2.018	0.552	−2.018	0.55	−2.018	0.55
Median	−2.018	0.552	−1.922	0.369	−2.012	0.562	−2.019	0.562	−2.016	0.553	−2.018	0.552	−2.018	0.551	−2.018	0.55
Std. dev.	0.007	0.024	0.071	0.218	0.045	0.107	0.036	0.074	0.016	0.036	0.012	0.028	0.007	0.022	0.007	0.022
Mcse	0	0	0.005	0.047	0.002	0.011	0.001	0.005	0	0.001	0	0	0	0	0	0
Pctile 2.5%	−2.032	0.501	−2.074	−0.106	−2.105	0.335	−2.085	0.424	−2.05	0.479	−2.041	0.497	−2.032	0.508	−2.031	0.507
Pctile 97.5%	−2.004	0.596	−1.838	0.666	−1.931	0.755	−1.942	0.707	−1.988	0.621	−1.994	0.606	−2.004	0.596	−2.005	0.592
KS	NA	NA	0.743	0.703	0.408	0.363	0.341	0.31	0.229	0.117	0.107	0.066	0.027	0.034	0.033	0.032
KL	NA	NA	0.466	0.196	0.121	0.081	0.081	0.019	0.026	0.009	0.02	0.008	0.061	0.036	0.049	0.041
Time	0.187 s		87.344 s		95.831 s		105.955 s		181.413 s		275.357 s		2010.322 s		19,319.663 s	

(*Continues*)

TABLE I—*Continued*

	Exact Metropolis		$R = 1000$		$R = 5000$		$R = 10{,}000$		$R = 50{,}000$		$R = 100{,}000$		$R = 1$ mil		$R = 10$ mil	
	α	β	α	β	α	β	α	β	α	β	α	β	α	β	α	β
							$n = 1000$									
Mean	−2.001	0.481	−1.986	0.456	−1.974	0.459	−1.995	0.486	−1.999	0.479	−2.001	0.479	−2.001	0.481	−2.002	0.481
Median	−2.001	0.48	−1.991	0.501	−1.979	0.461	−1.993	0.479	−2.001	0.48	−2.001	0.479	−2.001	0.48	−2.002	0.482
Std. dev.	0.003	0.011	0.081	0.247	0.05	0.091	0.031	0.07	0.017	0.037	0.011	0.026	0.004	0.012	0.003	0.012
Mcse	0	0	0.007	0.06	0.002	0.007	0.001	0.004	0	0.001	0	0	0	0	0	0
Pctile 2.5%	−2.008	0.459	−2.148	−0.007	−2.047	0.284	−2.057	0.361	−2.029	0.404	−2.024	0.425	−2.01	0.457	−2.008	0.459
Pctile 97.5%	−1.995	0.503	−1.835	0.813	−1.825	0.642	−1.938	0.64	−1.966	0.55	−1.979	0.529	−1.993	0.505	−1.995	0.504
KS	NA	NA	0.506	0.484	0.63	0.47	0.502	0.355	0.351	0.268	0.271	0.216	0.078	0.018	0.027	0.036
KL	NA	NA	0.3	0.261	0.528	0.041	0.297	0.083	0.45	0.21	0.137	0.047	0.021	0.031	0.034	0.014
Time	0.234 s		364.761 s		371.563 s		381.172 s		459.578 s		556.330 s		2304.228 s		19,730.772 s	

TABLE II

ESTIMATED STRUCTURAL PARAMETERS FOR MODEL (66), $n = 100$

	$R = 1000$		$R = 10{,}000$		$R = 100{,}000$		$R = 1{,}000{,}000$	
$n = 100$	α	β	α	β	α	β	α	β
true $= (-3, 0.01)$	True parameters $(\alpha, \beta) = (-3, 0.01)$							
Mean	−2.7450	−0.0233	−2.9095	−0.0035	−2.9407	−0.0009	−2.9206	−0.0025
Median	−2.7575	−0.0185	−2.9171	−0.0021	−2.9498	0.0003	−2.9288	−0.0014
Std. dev.	0.4782	0.0460	0.2032	0.0201	0.1860	0.0183	0.1916	0.0189
Mcse	0.0975	0.0010	0.0111	0.0001	0.0094	0.0001	0.0104	0.0001
Pctile 2.5%	−3.6862	−0.1208	−3.2660	−0.0462	−3.2614	−0.0400	−3.2468	−0.0412
Pctile 97.5%	−1.7789	0.0545	−2.4916	0.0305	−2.5452	0.0303	−2.5158	0.0297
Time (s)	25.3800		236.2100		2485.5200		24,658.1500	
true $= (-3, 0.03)$	True parameters $(\alpha, \beta) = (-3, 0.03)$							
Mean	−2.6075	0.0002	−2.7578	0.0124	−2.7618	0.0126	−2.7720	0.0134
Median	−2.6425	0.0036	−2.7804	0.0140	−2.7812	0.0140	−2.7917	0.0148
Std.dev.	0.4396	0.0306	0.1757	0.0122	0.1663	0.0116	0.1671	0.0116
Mcse	0.0819	0.0004	0.0075	0.0000	0.0073	0.0000	0.0080	0.0000
Pctile 2.5%	−3.4144	−0.0682	−3.0320	−0.0150	−3.0185	−0.0132	−3.0165	−0.0129
Pctile 97.5%	−1.6856	0.0526	−2.3671	0.0299	−2.4054	0.0299	−2.3897	0.0297
Time (s)	27.3900		256.2500		2647.7600		26,277.7000	
true $= (5, -0.1)$	True parameters $(\alpha, \beta) = (5, -0.1)$							
Mean	4.8397	−0.0964	4.8722	−0.0968	4.8743	−0.0968	4.8856	−0.0970
Median	4.8265	−0.0963	4.8674	−0.0968	4.8682	−0.0968	4.8846	−0.0970
Std. dev.	0.4031	0.0067	0.1550	0.0026	0.1188	0.0018	0.1137	0.0018
Mcse	0.0427	0.0000	0.0064	0.0000	0.0041	0.0000	0.0039	0.0000
Pctile 2.5%	4.0677	−0.1101	4.5688	−0.1020	4.6493	−0.1005	4.6645	−0.1006
Pctile 97.5%	5.6615	−0.0836	5.1707	−0.0920	5.1202	−0.0933	5.1156	−0.0936
Time (s)	49.3300		433.5100		4254.5800		41,218.3700	

Let us now consider a model with homogeneous players where there is no utility from reciprocated links, but only from indirect connections and popularity, that is,

$$Q(g, \alpha, \beta) = \alpha \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij} + \beta \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} g_{ij} g_{jk}. \tag{66}$$

The estimated parameters for this model are shown in Tables II and III. We generate the network data using different parameter vectors. The first panel corresponds to parameters $(\alpha, \beta) = (-3, 1/n)$. This is a model that generates a sparse network and the likelihood has a unique mode. The second panel shows estimates for a model with parameters $(\alpha, \beta) = (-3, 3/n)$, with a variational problem with two local solutions that generates problems of convergence with a local sampler. The last panel is a model with negative externalities $(\alpha, \beta) = (5, -10/n)$ that does not converge to an Erdős–Rényi model. We also simulated a model with parameters $(\alpha, \beta) = (-3, 7/n)$. However, if we solve the variational problem with these parameters, we can show that the solution is an Erdős–Rényi model with probability of linking $\mu^* = 1$, that is, the full network. Therefore, a model with parameters $\alpha = -3$, for any $\beta > 7/n$, would also generate a full network. Any attempt to estimate β with data consisting of a full network is futile.

The estimates using the non-local sampler are precise for a moderate amount of network simulations. Clearly, estimates in Table II are less precise than the ones in Table III,

TABLE III

ESTIMATED STRUCTURAL PARAMETERS FOR MODEL (66), $n = 200$

	$R = 1000$		$R = 10{,}000$		$R = 100{,}000$		$R = 1{,}000{,}000$	
$n = 200$	α	β	α	β	α	β	α	β
true $= (-3, 0.005)$			True parameters $(\alpha, \beta) = (-3, 0.005)$					
Mean	−2.6694	−0.0112	−2.9707	0.0042	−3.0529	0.0083	−3.0603	0.0086
Median	−2.7045	−0.0086	−2.9972	0.0056	−3.0675	0.0089	−3.0784	0.0095
Std. dev.	0.5631	0.0254	0.1841	0.0083	0.1137	0.0053	0.1113	0.0052
Mcse	0.1341	0.0003	0.0089	0.0000	0.0035	0.0000	0.0032	0.0000
Pctile 2.5%	−3.7109	−0.0665	−3.2574	−0.0148	−3.2202	−0.0035	−3.2244	−0.0036
Pctile 97.5%	−1.5002	0.0332	−2.5689	0.0159	−2.8044	0.0159	−2.8007	0.0158
Time (s)	173.7800		1651.4400		16,248.9400		149,962.1800	
true $= (-3, 0.015)$			True parameters $(\alpha, \beta) = (-3, 0.015)$					
Mean	−2.4770	−0.0033	−2.7773	0.0075	−2.8601	0.0104	−2.8518	0.0101
Median	−2.5002	−0.0019	−2.8042	0.0083	−2.8785	0.0111	−2.8703	0.0108
Std. dev.	0.5828	0.0200	0.1627	0.0055	0.1012	0.0035	0.1028	0.0035
Mcse	0.1012	0.0001	0.0078	0.0000	0.0028	0.0000	0.0028	0.0000
Pctile 2.5%	−3.6184	−0.0474	−3.0206	−0.0054	−3.0026	0.0024	−2.9961	0.0020
Pctile 97.5%	−1.2515	0.0346	−2.4080	0.0148	−2.6267	0.0150	−2.6149	0.0149
Time (s)	190.5800		1783.1900		17,496.5900		161,462.5300	
true $= (5, -0.05)$			True parameters $(\alpha, \beta) = (5, -0.05)$					
Mean	5.0734	−0.0504	5.0782	−0.0503	5.0528	−0.0501	5.0477	−0.0501
Median	5.0475	−0.0502	5.0718	−0.0503	5.0539	−0.0501	5.0478	−0.0501
Std. dev.	0.4791	0.0039	0.1535	0.0012	0.0713	0.0005	0.0644	0.0005
Mcse	0.0765	0.0000	0.0068	0.0000	0.0017	0.0000	0.0011	0.0000
Pctile 2.5%	4.1775	−0.0587	4.7926	−0.0529	4.9129	−0.0512	4.9220	−0.0511
Pctile 97.5%	6.0765	−0.0431	5.3987	−0.0480	5.1904	−0.0490	5.1781	−0.0491
Time (s)	361.5600		2996.3300		29,001.6900		257,572.3700	

since the number of links is smaller. Because of our modified network sampler, there is no need to have a large number of network simulations. The reason is that the non-local sampler can jump quickly to the correct mode(s) of the likelihood: once it reaches an area close to the global maximum, convergence is in quadratic time, since it will reject jumps to local maxima of the variational problem that are not global maxima. While the number of iterations may be lower, the computational time of each iteration is higher, because the large steps are computationally expensive. In Table II, the precision gain from additional network simulations is negligible when $R > 10{,}000$. Notice that the computational time is higher for the model with negative externality. The reason is that the equilibrium network generated at the true parameters $(\alpha, \beta) = (5, -10/n)$ is denser than the ones in the previous panels, and therefore the large steps are more computationally expensive than for the other two models.

In the next table, we consider a model where players are homogeneous and they receive utility from direct links, reciprocated links, indirect links, and popularity. The potential function of such model is defined as

$$Q(g, \alpha, \beta, \gamma) = \alpha \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij} + \beta \sum_{i=1}^{n} \sum_{j>i}^{n} g_{ij} g_{ji} + \gamma \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} g_{ij} g_{jk}. \tag{67}$$

TABLE IV

ESTIMATED STRUCTURAL PARAMETERS FOR MODEL (67), $n = 100$

	True Parameters $(\alpha, \beta, \gamma) = (-2.00, 0.50, 0.01)$											
	$R = 1000$			$R = 10{,}000$			$R = 100{,}000$			$R = 1{,}000{,}000$		
$n = 100$	α	β	γ	α	β	γ	α	β	γ	α	β	γ
Mean	−1.9321	0.5098	0.0074	−2.1182	0.5168	0.0133	−2.1034	0.5115	0.0129	−2.0938	0.5196	0.0126
Median	−1.9756	0.5080	0.0089	−2.1382	0.5214	0.0139	−2.1251	0.5134	0.0136	−2.1066	0.5207	0.0131
Std. dev.	0.4677	0.2330	0.0135	0.1899	0.0997	0.0054	0.1877	0.0894	0.0054	0.1832	0.0882	0.0053
Mcse	0.0967	0.0209	0.0001	0.0121	0.0037	0.0000	0.0110	0.0028	0.0000	0.0132	0.0027	0.0000
Pctile 2.5%	−2.7241	0.0459	−0.0224	−2.4259	0.3186	0.0014	−2.4002	0.3341	0.0012	−2.3871	0.3416	0.0013
Pctile 97.5%	−0.9492	0.9699	0.0300	−1.7149	0.7115	0.0216	−1.6983	0.6896	0.0213	−1.7014	0.6894	0.0211
Time (s)		42			355			3545			35,806	

TABLE V

ESTIMATED STRUCTURAL PARAMETERS FOR MODEL (68), $n = 100$

	True Parameters $(\alpha, \beta, \gamma) = (2.00, 0.11, -0.05)$								
	$R = 1000$			$R = 10{,}000$			$R = 100{,}000$		
$n = 100$	α	β	γ	α	β	γ	α	β	γ
Mean	2.2144	0.1056	−0.0530	2.0718	0.1053	−0.0503	2.0636	0.1052	−0.0501
Median	2.0828	0.1064	−0.0506	2.0443	0.1054	−0.0498	2.0396	0.1055	−0.0497
Std. dev.	0.8575	0.0227	0.0155	0.3348	0.0084	0.0061	0.2723	0.0066	0.0049
Mcse	0.4485	0.0002	0.0001	0.0711	0.0000	0.0000	0.0487	0.0000	0.0000
Pctile 2.5%	0.8456	0.0598	−0.0881	1.4803	0.0875	−0.0636	1.5959	0.0914	−0.0608
Pctile 97.5%	4.1495	0.1473	−0.0286	2.8115	0.1222	−0.0397	2.6445	0.1174	−0.0418
Time (s)		97			314			2913	

The data are generated by parameters $(\alpha, \beta, \gamma) = (-2.00, 0.50, 0.01)$. The pattern of Table IV is similar to the previous analysis: the increase in precision for $R > 10{,}000$ is minimal with respect to the increased cost of sampling networks.

Finally, we estimate a simple model with heterogeneous players. There is only one binary covariate X and the players receive utility from direct links, and indirect links and popularity. The covariate is generated as a Bernoulli variable with $P(X_i = 1) = 0.3$. The utility from indirect links/popularity is positive if both i and k belong to type-1; and it is negative if they belong to different types. The potential of this model is

$$Q(g, \alpha, \beta, \gamma) = \alpha \sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij} + \beta \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} g_{ij}g_{jk}\mathbb{1}_{\{X_i = X_k = 1\}} + \gamma \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} g_{ij}g_{jk}\mathbb{1}_{\{X_i \neq X_k\}} \tag{68}$$

and the data are generated with parameters $(2, 11/n, -5/n)$.

The estimation results in Table V are for $n = 100$. The estimates are again very precise for a moderate amount of simulations.

REFERENCES

ARISTOFF, D., AND L. ZHU (2014): "On the Phase Transition Curve in a Directed Exponential Random Graph Model". [10,12]

ATCHADE, Y., AND J. WANG (2014): "Approximate Bayesian Computation for Exponential Random Graph Models for Large Social Networks," *Communications in Statistics—Simulation and Computation*, 43 (2), 359–377. [38]

ATHEY, S., AND G. IMBENS (2007): "Discrete Choice Models With Multiple Unobserved Choice Characteristics," *International Economic Review*, 48 (4), 1159–1192. [1]

BHAMIDI, S., G. BRESLER, AND A. SLY (2011): "Mixing Time of Exponential Random Graphs," *The Annals of Applied Probability*, 21 (6), 2146–2170. [20,35,37]

BOECKNER, D. (2013): "Directed Graph Limits and Directed Threshold Graphs," PhD Thesis. [3,5,7]

BORGS, C., J. T. CHAYES, L. LOVÁSZ, V. T. SÓS, AND K. VESZTERGOMBI (2008): "Convergent Sequences of Dense Graphs I: Subgraph Frequencies, Metric Properties and Testing," *Advances in Mathematics*, 219 (6), 1801–1851. [3]

CAIMO, A., AND N. FRIEL (2011): "Bayesian Inference for Exponential Random Graph Models," *Social Networks*, 33 (1), 41–55. [38]

CHATTERJEE, S., AND P. DIACONIS (2013): "Estimating and Understanding Exponential Random Graph Models," *Annals of Statistics*, 41 (5), 2428–2461. [2,3,8-10,25,36]

CHATTERJEE, S., AND S. R. S. VARADHAN (2011): "The Large Deviation Principle for the Erdős-Rényi Random Graph," *European Journal of Combinatorics*, 32 (7), 1000–1017. Homomorphisms and Limits. [3,6, 7]

GRAHAM, B. (2016): "An Econometric Model of Network Formation With Degree Heterogeneity," *Econometrica* (forthcoming). [2]

LOVASZ, L. (2012): *Large Networks and Graph Limits*. American Mathematical Society Colloquium Publications. American Mathematical Society. [3]

LOVASZ, L., AND B. SZEGEDY (2007): "Szemeredi's Lemma for the Analyst," *GAFA Geometric And Functional Analysis*, 17 (1), 252–270. [3,5]

RADIN, C., AND M. YIN (2013): "Phase Transitions in Exponential Random Graphs," *The Annals of Applied Probability*, 23 (6), 2458–2471. [10,12]

ROSSI, P., R. MCCULLOCH, AND G. ALLENBY (1996): "The Value of Purchase History Data in Target Marketing," *Marketing Science*, 15 (4), 321–340. [1]

Carey Business School, Johns Hopkins University, 100 International Drive, Baltimore, MD 21202, U.S.A.; angelo.mele@jhu.edu.

Co-editor Matthew O. Jackson handled this manuscript.

Manuscript received 4 November, 2011; final version accepted 31 October, 2016; available online 18 January, 2017.