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 this 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 Diaconis and Chatterjee (2011), we rescale the terms as

$$\pi_n(g; \alpha, \beta, \gamma) = \frac{\exp\left\{n^2\left[\alpha\frac{\sum_{i=1}^n \sum_{j=1}^n g_{ij}}{n^2} + \frac{\beta}{2}\frac{\sum_{i=1}^n \sum_{j=1}^n g_{ij}g_{ji}}{n^2} + \gamma\frac{\sum_{i=1}^n \sum_{j=1}^n \sum_{k\neq i}^n g_{ij}g_{jk}}{n^3}\right]\right\}}{c(\alpha, \beta, \gamma, \mathcal{G}_n)} \tag{46}$$

Notice that γ needs to be rescaled (i.e. divided by n) when we run the simulations using the usual ERGM form, i.e. $\gamma^o = \frac{\gamma}{n}$ 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 Diaconis and Chatterjee (2011), for completeness.

D.1 A crash course on graph limits

Most of this brief digression follows the overview in Diaconis and Chatterjee (2011), 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

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

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 Diaconis and Chatterjee (2011) because we are considering directed graphs, 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. $V(H) = \{1, 2, ..., k\}$) the limit object for $t(H,G_n)$ is

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

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

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

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

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

with limit object

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

[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.

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 rescale 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 $\{1, ..., n\}$ we can always define the graph limit representation f^G as

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

where the symbol $\lceil a \rceil$ indicates the ceiling of a, i.e. 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*

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

where f and g are functions in $\mathcal{W}$. However, there is some non-trivial 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

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

This makes $\left(\widetilde{\mathcal{W}}, \delta_\square\right)$ 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* $\left(\widetilde{\mathcal{W}}, \delta_\square\right)$ *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 $\mathcal{P} = \{P_1, P_2, ..., P_k\}$ of $[0, 1]$ into measurable sets, we define $h_\mathcal{P} : [0, 1]^2 \to [0, 1]$ to be the stepfunction 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, ... 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 $\mathcal{P}_{n,k} = \{P_{1,n,k}, ..., P_{m_k,n,k}\}$ 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 re-arrange 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 i-th 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's 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 $k = 2, 3, ...$ 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, i.e. 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 $U_1(X,Y)$, $U_2(X,Y)$, ... is martingale, and each element of the sequence is bounded. Using the Martingale Convergence Theorem we can show that the sequence $U_1(X,Y)$, $U_2(X,Y)$, ... 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 $\left(\widetilde{\mathcal{W}}, \delta_\square\right)$. ■

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 Erdos-Renyi graph. Let $G(n,p)$ indicate the the random *undirected* graph with n vertices where each link is formed independently with probability p. Define a function $I_p : [0,1] \to \mathbb{R}$

$$I_p(u) \equiv \frac{1}{2}u\log\frac{u}{p} + \frac{1}{2}(1-u)\log\frac{1-u}{1-p} \tag{47}$$

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

$$\begin{aligned} I_p(h) &= \int_0^1\int_0^1 I_p\left(h(x,y)\right)dxdy \\ &= \frac{1}{2}\int_0^1\int_0^1 \left[h(x,y)\log\frac{h(x,y)}{p} + (1-h(x,y))\log\frac{1-h(x,y)}{p}\right]dxdy \end{aligned} \tag{48}$$

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) state a large deviation principle for the Erdos Renyi 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 Erdos-Renyi 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 (48). 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* Erdos-Renyi 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 $\mathcal{I}_p : [0,1] \to \mathbb{R}$

$$\mathcal{I}_p(u) \equiv u \log \frac{u}{p} + (1-u) \log \frac{1-u}{1-p} \tag{49}$$

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

$$\begin{aligned} \mathcal{I}_p(h) &= \int_0^1 \int_0^1 I_p\left(h(x,y)\right) dxdy \\ &= \int_0^1 \int_0^1 \left[h(x,y) \log \frac{h(x,y)}{p} + (1-h(x,y)) \log \frac{1-h(x,y)}{p} \right] dxdy \end{aligned} \tag{50}$$

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) prove 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 Erdos-Renyi 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 (50). 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 (50). 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 simmetry. ■

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

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

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

$$\psi_n = \frac{1}{n^2}\log\sum_{G\in\mathcal{G}_n}\exp\left\{n^2\left[T(\widetilde{G})\right]\right\} \tag{51}$$

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

$$I(u) \equiv \frac{1}{2}u\log u + \frac{1}{2}(1-u)\log(1-u) \tag{52}$$

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

$$I(\widetilde{h}) \equiv \frac{1}{2}\int_0^1\int_0^1 I(h(x,y))dxdy$$

$$\begin{aligned} I(\widetilde{h}) &\equiv \frac{1}{2}\int_0^1\int_0^1 I(h(x,y))dxdy \\ &= \frac{1}{2}\int_0^1\int_0^1 \left[h(x,y)\log h(x,y) + (1-h(x,y))\log(1-h(x,y))\right]dxdy \end{aligned} \tag{53}$$

THEOREM 9 *(Theorem 3.1 for ERGM in Chatterjee-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 (51) and (53) respectively, then*

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

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 rescaling 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

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

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

$$\psi_n = \frac{1}{n^2}\log\sum_{G\in\mathcal{G}_n}\exp\left\{n^2\left[\mathcal{T}(\widetilde{G})\right]\right\} \tag{54}$$

This is the same as the stationary distribution of our model, after some re-scaling of the utility functions. The rescaling by n^2 is necessary to guarantee that the limits for $n \to \infty$ converge to some non-trivial 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 Diaconis and Chatterjee (2011). We define a rate function

$$\mathcal{I}(u) \equiv u\log u + (1-u)\log(1-u) \tag{55}$$

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

$$\begin{aligned}\mathcal{I}(\widetilde{h}) &\equiv \int_0^1\int_0^1 I(h(x,y))dxdy \\ &= \int_0^1\int_0^1 \left[h(x,y)\log h(x,y) + (1-h(x,y))\log(1-h(x,y))\right]dxdy\end{aligned} \tag{56}$$

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 (54) and (56) respectively, then*

$$\psi \equiv \lim_{n\to\infty}\psi_n = \sup_{\widetilde{h}\in\widetilde{\mathcal{W}}}\left\{\mathcal{T}(\widetilde{h}) - \mathcal{I}(\widetilde{h})\right\} \tag{57}$$

Proof. The proof of this result follows closely the proof of Theorem 3.1 in Diaconis and Chatterjee (2011), 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

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

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

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

Therefore we obtain

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

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 Diaconis and Chatterjee (2011). ■

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's consider a model with utility from directed links and friends of friends. Using the notation developed above, we are considering a model with function $\mathcal{T}$

$$\mathcal{T}(\widetilde{G}) = \theta_1\frac{\sum_{i=1}^n\sum_{j=1}^n g_{ij}}{n^2} + \theta_2\frac{\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n g_{ij}g_{jk}}{n^3} \tag{58}$$

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

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

and

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

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 (58) 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, i.e. $\ell(\mu_1^*) = \ell(\mu_2^*)$.*

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

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

In particular we have in our case

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

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

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

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's consider the maximization problem again. For $h(x,y)$ to be a solution, we need

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

In other words, the Holder inequality must hold with equality, i.e. we need

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

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 $\mathcal{T}(h) = \theta_1\mu + \theta_2\mu^2$ 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), Aristoff and Zhu (2014). The first order and second order conditions are

$$\ell'(\mu, \theta_1, \theta_2) = \theta_1 + 2\theta_2\mu - \log\frac{\mu}{1-\mu} \tag{59}$$

$$\ell''(\mu, \theta_1, \theta_2) = 2\theta_2 - \frac{1}{\mu(1-\mu)} \tag{60}$$

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

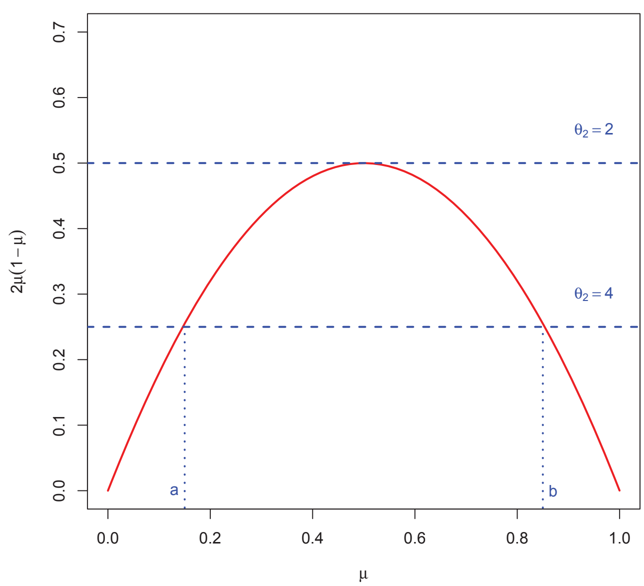
$$\theta_2 \le \frac{1}{2\mu(1-\mu)}$$

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 picture of parabola), 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 $[0, 1]$. It is easy to

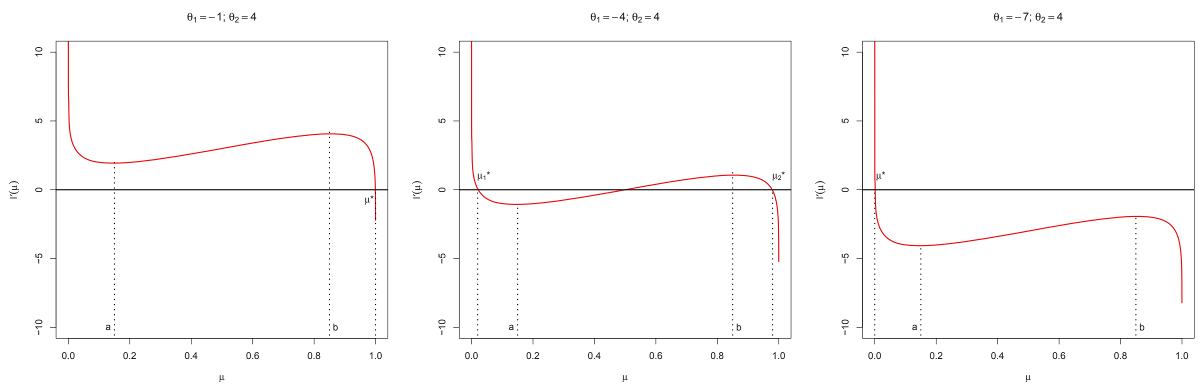
[50]This is because, when $\theta_2 > 2$, we have $\ell''(\mu, \theta_1, \theta_2) \le 0$ when $\theta_2 \le \frac{1}{2\mu(1-\mu)}$ or $2\mu(1-\mu) \le \frac{1}{\theta_2}$. The equality is realized at two intersections of the horizontal line $1/\theta_2$ with the parabola $2\mu(1-\mu)$. We call the intersections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

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 3 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 2 local maximizers $\mu_1^* < a < b < \mu_2^*$

The three cases are shown in the following pictures, 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's 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

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

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

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

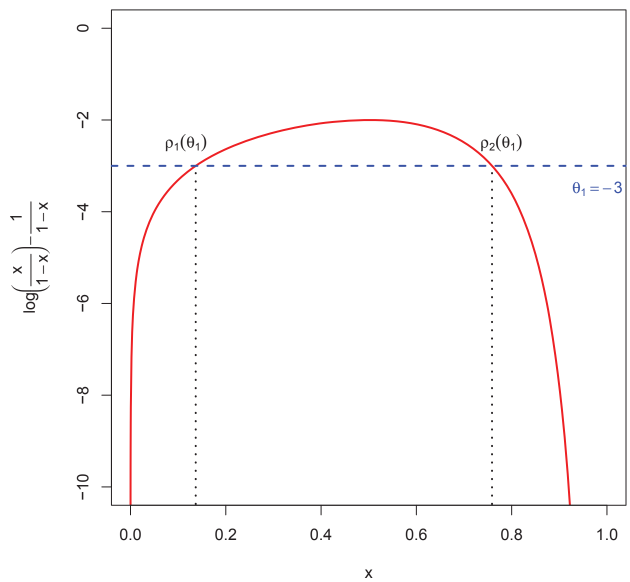
The function $\log \frac{a}{1-a} - \frac{1}{1-a}$ has a maximum at -2 and therefore we have [51]

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

[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.

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 below.



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 $y = \log\left(\frac{x}{1-x}\right) - \frac{1}{1-x}$ 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 decreasing in θ_1. It's trivial to show that when θ_1 decreases, $\phi_1(\theta_1)$ converges to 0 while $\phi_2(\theta_1)$ converges to 1.

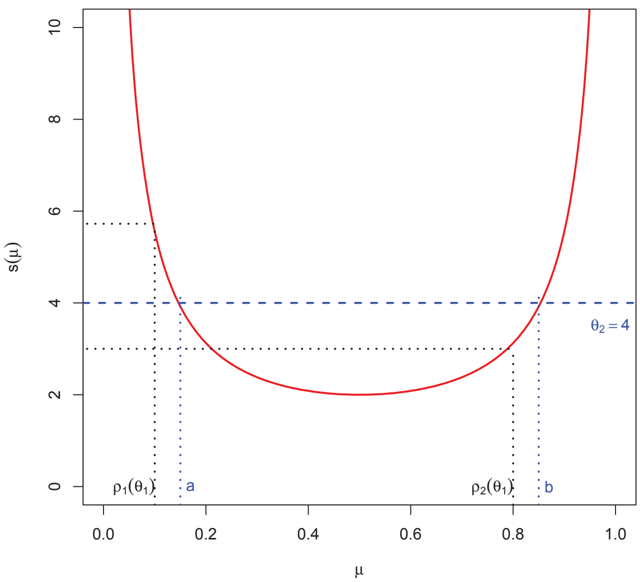
Consider the case in which $\ell'(a, \theta_1, \theta_2) < 0 < \ell'(b, \theta_1, \theta_2)$ with two maximizers. Define the function

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

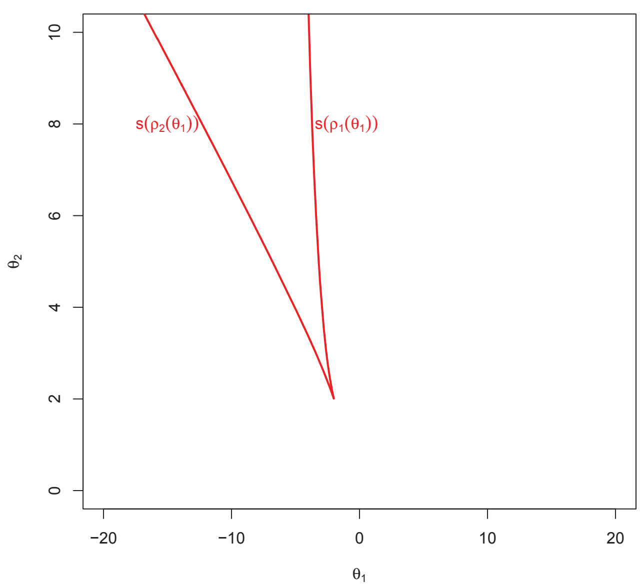
Since $\ell'(a, \theta_1, \theta_2) < 0$ we have $a > \phi_1(\theta_1)$, which implies $s(a) < s(\phi_1(\theta_1))$. Therefore $\theta_2 < s(\phi_1(\theta_1)) = \frac{1}{2\phi_1(\theta_1)(1-\phi_1(\theta_1))}$.

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 $\theta_2 > s(\phi_2(\theta_1)) = \frac{1}{2\phi_2(\theta_1)(1-\phi_2(\theta_1))}$.

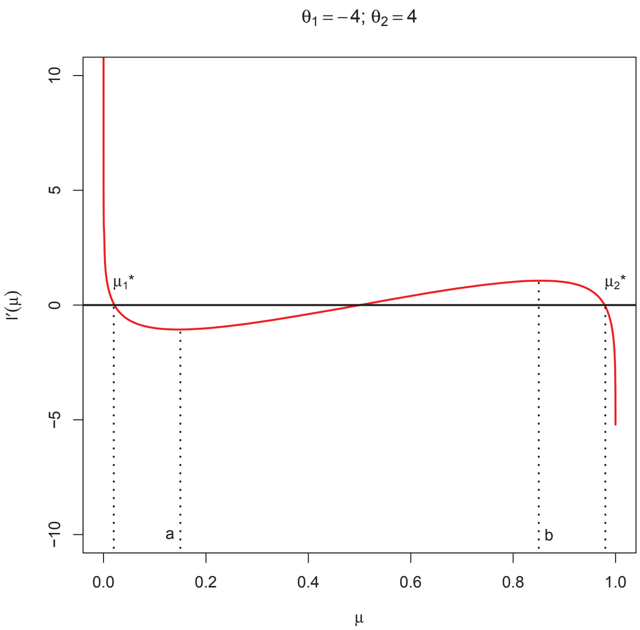
Notice that $s(\phi_1(\theta_1)) > s(\phi_2(\theta_1))$ for any (θ_1, θ_2) in this region of the parameters (see picture below).



The areas are shown in the following picture



Within the V-shaped region there are 2 solutions to the maximization problem, i.e. 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. Therefeore 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 maximizer. Let's 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 the graph above, i.e. (let $\widehat{\mu}$ indicate the intersection of $\ell'(\mu, \theta_1, \theta_2)$ and the x-axis between $\mu_1^*$and μ_2^*)

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

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. 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. ■

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 re-write the variational problem as a calculus problem, as shown in the Theorems below.

LEMMA 6 *For the following homomorphism densities:*

$$t(H, G) = \frac{\sum_i \sum_j \sum_k g_{ij} g_{jk} g_{ki}}{n^3} \tag{61}$$

$$t(H, G) = \frac{\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n g_{ij} g_{jk}}{n^3} \tag{62}$$

$$t(H, G) = \frac{\sum_i \sum_j \sum_k \sum_l g_{ij} g_{jk} g_{kl} g_{li}}{n^4} \tag{63}$$

$$t(H, G) = \frac{1}{n^r} \sum_{1 \leq i, j_1, j_2, .., j_r \leq n} g_{ij_1} g_{j_1 j_2} \cdot \cdot g_{j_r i} \tag{64}$$

$$t(H, G) = \frac{1}{n^{r-1}} \sum_{1 \leq i, j_1, j_2, .., j_r \leq n} g_{ij_1} g_{ij_2} \cdot \cdot g_{ij_r} \tag{65}$$

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 (61) the value $e(H) = 3$ and the limit object is

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

Using the Holder inequality and some algebra, we obtain

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

For the homomorphism density in (62), $e(H) = 2$ and using Holder inequality we get

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

For the homomorphism density in (64), $e(H) = r$ and using Holder inequality we get

$$
\begin{aligned}
t(H,h) &= \int_{[0,1]^r} h(x_i,x_{j_1})h(x_{j_1},x_{j_2})\cdots h(x_{j_r},x_i)dx_i dx_{j_1}\cdots dx_{j_r} \\
&\leq \left(\int_{[0,1]^r} h(x_i,x_{j_1})^r dx_i dx_{j_1}\cdots dx_{j_r}\right)^{\frac{1}{r}} \left(\int_{[0,1]^r} h(x_{j_1},x_{j_2})^r dx_i dx_{j_1}\cdots dx_{j_r}\right)^{\frac{1}{r}} \\
&\quad \cdots \left(\int_{[0,1]^r} h(x_{j_r},x_i)^r dx_i dx_{j_1}\cdots dx_{j_r}\right)^{\frac{1}{r}} \\
&= \left(\int_{[0,1]^2} h(x_i,x_{j_1})^r dx_i dx_{j_1} \int_{[0,1]^{r-2}} dx_{j_2}\cdots dx_{j_r}\right)^{\frac{1}{r}} \\
&\quad \times \left(\int_{[0,1]^2} h(x_{j_1},x_{j_2})^r dx_{j_1} dx_{j_2} \int_{[0,1]^{r-2}} dx_i dx_{j_3}\cdots dx_{j_r}\right)^{\frac{1}{r}} \\
&\quad \cdots \left(\int_{[0,1]^2} h(x_{j_r},x_i)^r dx_{j_r} dx_i \int_{[0,1]^{r-2}} dx_{j_1}\cdots dx_{j_{r-1}}\right)^{\frac{1}{r}} \\
&= \left(\int_{[0,1]^2} h(x,y)^r dxdy\right)^{\frac{1}{r}} \left(\int_{[0,1]^2} h(x,y)^r dxdy\right)^{\frac{1}{r}} \cdots \left(\int_{[0,1]^2} h(x,y)^r dxdy\right)^{\frac{1}{r}} \\
&= \int_{[0,1]^2} h(x,y)^r dxdy = \int_0^1\int_0^1 h(x,y)^{e(H)} dxdy
\end{aligned}
$$

For the homomorphism density in (65), $e(H) = r$ and using Holder inequality we get

$$\begin{aligned}
t(H,h) &= \int_{[0,1]^r} h(x_i,x_{j_1})h(x_i,x_{j_2})\cdots h(x_i,x_{j_r})dx_idx_{j_1}\cdots dx_{j_r} \\
&\leq \left(\int_{[0,1]^r} h(x_i,x_{j_1})^r dx_idx_{j_1}\cdots dx_{j_r}\right)^{\frac{1}{r}} \left(\int_{[0,1]^r} h(x_i,x_{j_2})^r dx_idx_{j_1}\cdots dx_{j_r}\right)^{\frac{1}{r}} \\
&\quad \cdots \left(\int_{[0,1]^r} h(x_i,x_{j_r})^r dx_idx_{j_1}\cdots dx_{j_r}\right)^{\frac{1}{r}} \\
&= \left(\int_{[0,1]^2} h(x_i,x_{j_1})^r dx_idx_{j_1}\int_{[0,1]^{r-2}} dx_{j_2}\cdots dx_{j_r}\right)^{\frac{1}{r}} \\
&\quad \times \left(\int_{[0,1]^2} h(x_i,x_{j_2})^r dx_idx_{j_2}\int_{[0,1]^{r-2}} dx_{j_1}dx_{j_3}\cdots dx_{j_r}\right)^{\frac{1}{r}} \\
&\quad \cdots \left(\int_{[0,1]^2} h(x_i,x_{j_r})^r dx_idx_{j_r}\int_{[0,1]^{r-2}} dx_{j_1}\cdots dx_{j_{r-1}}\right)^{\frac{1}{r}} \\
&= \left(\int_{[0,1]^2} h(x,y)^r dxdy\right)^{\frac{1}{r}} \left(\int_{[0,1]^2} h(x,y)^r dxdy\right)^{\frac{1}{r}} \cdots \left(\int_{[0,1]^2} h(x,y)^r dxdy\right)^{\frac{1}{r}} \\
&= \int_{[0,1]^2} h(x,y)^r dxdy = \int_0^1\int_0^1 h(x,y)^{e(H)}dxdy
\end{aligned}$$

■

The following theorem uses the result of the 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 et al. (2011) (see more detail below).

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

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

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

$$\mathcal{T}(G) = \alpha\frac{\sum_{i=1}^n\sum_{j=1}^n g_{ij}}{n^2} + \beta\frac{\sum_{1\leq i,j_1,j_2,..,j_r\leq n} g_{ij_1}g_{j_1j_2}\cdot\cdot g_{j_ri}}{n^r}$$

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

the log-partitition 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-partitition 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's inequality in Lemma 6 we can say that

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

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's consider the maximization problem again. For $h(x, y)$ to be a solution, we need

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

In other words, the Holder inequality must hold with equality, i.e. we need

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

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. ■

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*

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

then we have

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, i.e. $\ell(\mu_1^*) = \ell(\mu_2^*)$.*

Proof. The first and second order conditions are

$$\begin{aligned}
\ell'(\mu,\alpha,\beta) &= \alpha + \beta r \mu^{r-1} - \ln\left(\frac{\mu}{1-\mu}\right) \\
\ell''(\mu,\alpha,\beta) &= \beta r(r-1)\mu^{r-2} - \frac{1}{\mu(1-\mu)}
\end{aligned}$$

The function $\ell(\mu, \alpha, \beta)$ is concave if $\ell''(\mu, \alpha, \beta) < 0$, i.e. 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 i, i.e. $\ell(\mu_1^*) = \ell(\mu_2^*)$n the interval $\left[0, \frac{r}{r-1}\right)$ and increasing in the interval $\left(\frac{r}{r-1}, 1\right]$. Therefore the function $\ell(\mu, \alpha, \beta)$ 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 the figure below (notice that we are plotting the function $1/s(\mu)$ against the line $1/\beta$).

[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

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

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

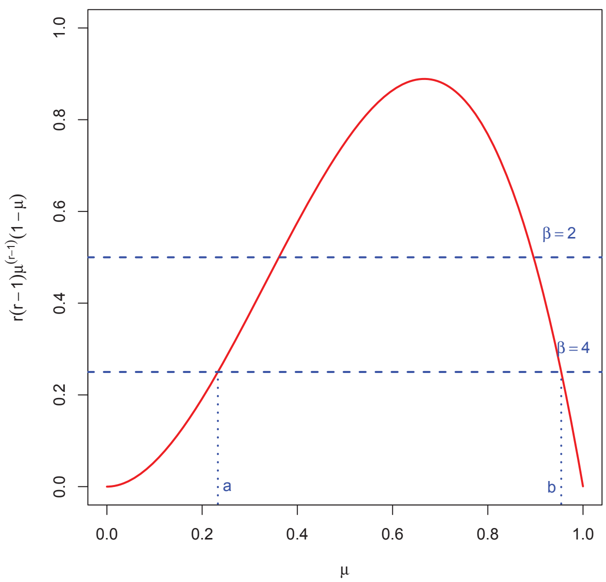
So solving the FOCs we obtain the maximizer of $1/s(\mu)$

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

and the maximum is

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

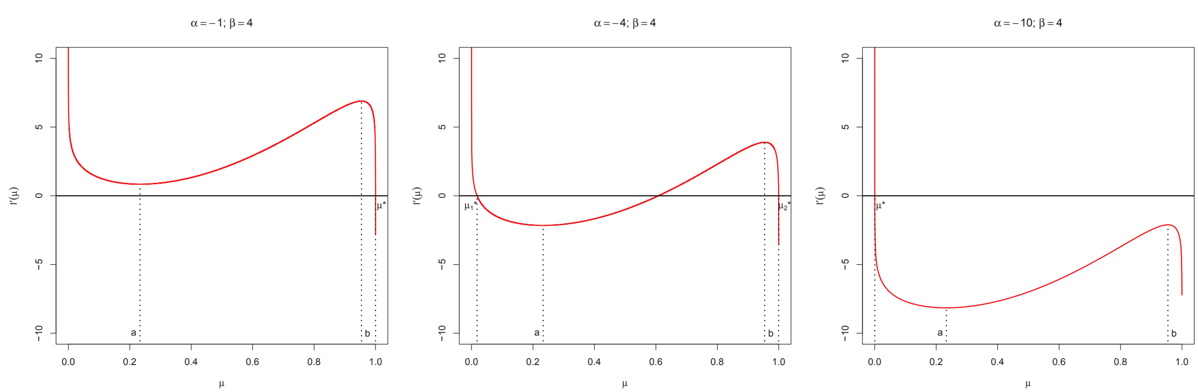
Therefore the minimum of $s(\mu)$ is $\frac{r^{r-1}}{(r-1)^r}$.



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 2 local maximizers $\mu_1^* < a < b < \mu_2^*$

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



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

Let's 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

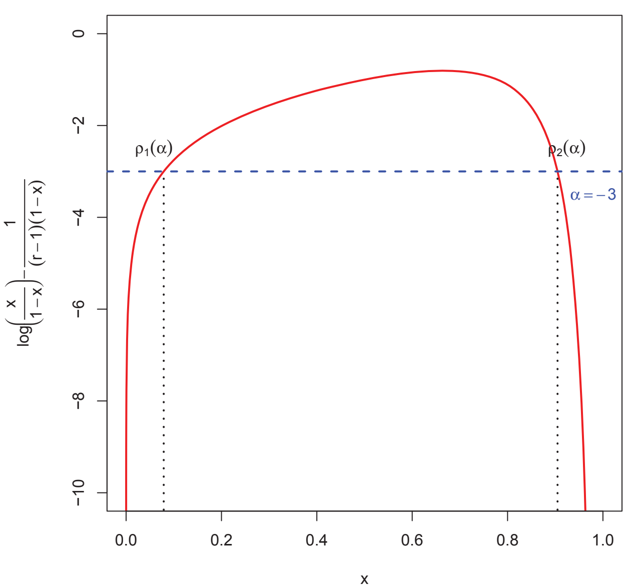
$$\alpha \geq \log \frac{a}{1-a} - \frac{1}{(r-1)(1-a)}$$

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]

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

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 below.



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

[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 $\log(r-1) - \frac{r}{r-1}$.

$\phi_2(\alpha)$. 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 $b < \phi_2(\alpha)$.

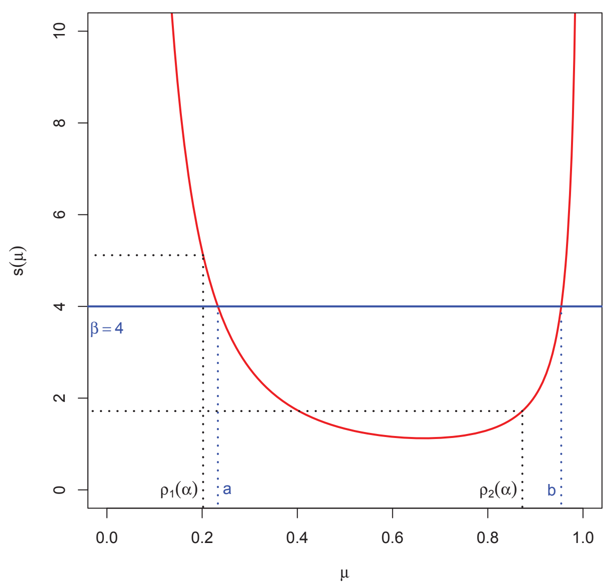
For any $\alpha < -2$, there exist $\phi_1(\alpha)$ and $\phi_2(\alpha)$ which are the intersection of the function $y = \log\left(\frac{x}{1-x}\right) - \frac{1}{(r-1)(1-x)}$ with the line $y = \alpha$. Since the function is continuous, monotonic increasing in $[0, \frac{r-1}{r})$ and monotonic decreasing 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's 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.

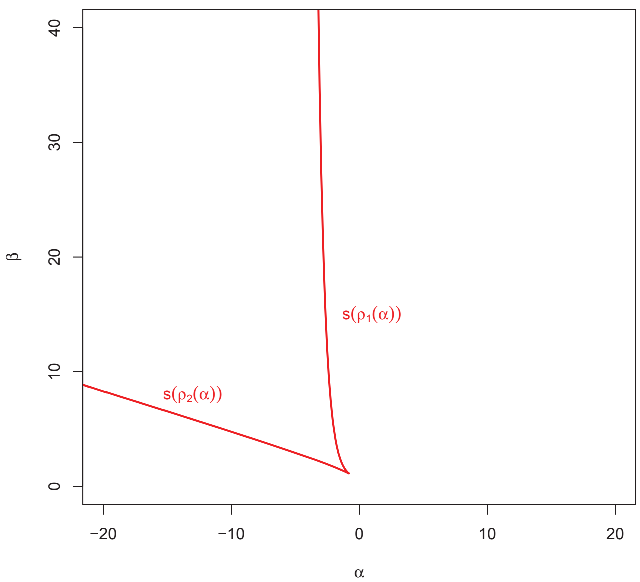
Since $\ell'(a, \alpha, \beta) < 0$ we have $a > \phi_1(\alpha)$, which implies $s(a) < s(\phi_1(\alpha))$. Therefore $\beta < s(\phi_1(\alpha, \beta)) = \frac{1}{r(r-1)\phi_1(\alpha)^{r-1}(1-\phi_1(\alpha))}$.

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 picture below for an example with $\beta = 4$, $\alpha = -2$, and $r = 3$).



The areas are shown in the following picture



and the rest of the proof follows. The existence of $\zeta(\alpha)$ is shown using similar argument as in the proof of Theorem 11, so it is omitted for brevity. ■

The next result is analogous to Theorem 6.3 in Diaconis and Chatterjee (2011), adapted to the directed network model. It shows that not all the specifications of the model generate directed Erdos-Renyi 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.

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

$$
\begin{aligned}
\mathcal{T}(h) - \mathcal{I}(h) &= \alpha \int h(x,y)dxdy + \beta \int h(x,y)h(y,z)dxdydz \\
&\quad - \int h(x,y)\ln h(x,y) + (1-h(x,y))\ln(1-h(x,y))dxdy \\
&= \alpha \int h(x,y)dxdy + \beta \int h(x,y)h(y,z)dxdydz \\
&\quad + \int h(x,y)\ln(1+e^{\alpha})dxdy - \int h(x,y)\ln(1+e^{\alpha})dxdy \\
&\quad - \int h(x,y)\ln h(x,y) + (1-h(x,y))\ln(1-h(x,y))dxdy \\
&= \beta \int h(x,y)h(y,z)dxdydz + \int h(x,y)\ln p dxdy + \int h(x,y)\ln(1-p)dxdy \\
&\quad - \int h(x,y)\ln h(x,y) + (1-h(x,y))\ln(1-h(x,y))dxdy \\
&= \beta \int h(x,y)h(y,z)dxdydz + \int h(x,y)\ln p dxdy + \int h(x,y)\ln(1-p)dxdy \\
&\quad + \int \ln(1-p)dxdy - \int \ln(1-p)dxdy \\
&\quad - \int h(x,y)\ln h(x,y) + (1-h(x,y))\ln(1-h(x,y))dxdy \\
&= \beta \int h(x,y)h(y,z)dxdydz + \ln(1-p) \\
&\quad - \int h(x,y)\ln \frac{h(x,y)}{p} + (1-h(x,y))\ln \frac{1-h(x,y)}{1-p} dxdy \\
&= -\lambda t(H_2,h) + \ln(1-p) - \mathcal{I}_p(h)
\end{aligned}
$$

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: $\lambda\mu^2 + \mathcal{I}_p(\mu) \leq \lambda x^2 + \mathcal{I}_p(x)$. The first order conditions for minimization give

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

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

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

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

$$\nu(x,y) = \begin{cases} p & \text{if } x \in [0,.5] \text{and } y \in [.5,1] \\ 0 & \text{otherwise} \end{cases}$$

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

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

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. ∎

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 hold.*

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 $t(H_2, h)$.

■

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

$$\mathcal{T}(G) = \alpha \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij}}{n^2} + \beta \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} g_{ij} g_{jk}}{n^3} + \gamma \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} g_{ij} g_{jk} g_{ki}}{n^3} \tag{66}$$

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

$$
\begin{aligned}
\mathcal{T}(h)-\mathcal{I}(h) &= \alpha\int h(x,y)dxdy+\beta\int h(x,y)h(y,z)dxdydz+\gamma\int h(x,y)h(y,z)h(z,x)dxdydz\\
&\quad -\int h(x,y)\ln h(x,y)+(1-h(x,y))\ln(1-h(x,y))dxdy\\
&= \alpha\int h(x,y)dxdy+\beta\int h(x,y)h(y,z)dxdydz+\gamma\int h(x,y)h(y,z)h(z,x)dxdydz\\
&\quad +\int h(x,y)\ln(1+e^{\alpha})dxdy-\int h(x,y)\ln(1+e^{\alpha})dxdy\\
&\quad -\int h(x,y)\ln h(x,y)+(1-h(x,y))\ln(1-h(x,y))dxdy\\
&= \beta\int h(x,y)h(y,z)dxdydz+\gamma\int h(x,y)h(y,z)h(z,x)dxdydz\\
&\quad +\int h(x,y)\ln p dxdy+\int h(x,y)\ln(1-p)dxdy\\
&\quad -\int h(x,y)\ln h(x,y)+(1-h(x,y))\ln(1-h(x,y))dxdy\\
&= \beta\int h(x,y)h(y,z)dxdydz+\gamma\int h(x,y)h(y,z)h(z,x)dxdydz\\
&\quad +\int h(x,y)\ln p dxdy+\int h(x,y)\ln(1-p)dxdy\\
&\quad +\int \ln(1-p)dxdy-\int \ln(1-p)dxdy\\
&\quad -\int h(x,y)\ln h(x,y)+(1-h(x,y))\ln(1-h(x,y))dxdy\\
&= \beta\int h(x,y)h(y,z)dxdydz+\gamma\int h(x,y)h(y,z)h(z,x)dxdydz+\ln(1-p)\\
&\quad -\int h(x,y)\ln\frac{h(x,y)}{p}+(1-h(x,y))\ln\frac{1-h(x,y)}{1-p}dxdy\\
&= \beta t(H_2,h)+\gamma t(H_3,h)+\ln(1-p)-\mathcal{I}_p(h)
\end{aligned}
$$

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

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

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

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

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

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

It follows that for almost all (x, y, z) triplets, $\nu(x, y)\nu(y, z) = 0$ and $\nu(x, y)\nu(y, z)\nu(z, x) = 0$. 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

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

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 (57) for the model in 66). The proof for $\gamma < 0$ follows the same steps. ■

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 letα_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

$$\begin{aligned}\ell'_\gamma(\mu, \alpha, \beta) &= \alpha + 2\beta\mu + 3\gamma\mu^2 - \ln\left(\frac{\mu}{1-\mu}\right) \\ \ell''_\gamma(\mu, \alpha, \beta) &= 2\beta + 6\gamma\mu - \frac{1}{\mu(1-\mu)}\end{aligned}$$

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

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

The function $s(\mu)$ is decreasing in $[0, .5)$ and increasing in $(.5, 1]$, and it has a minimum at $\mu = .5$, 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

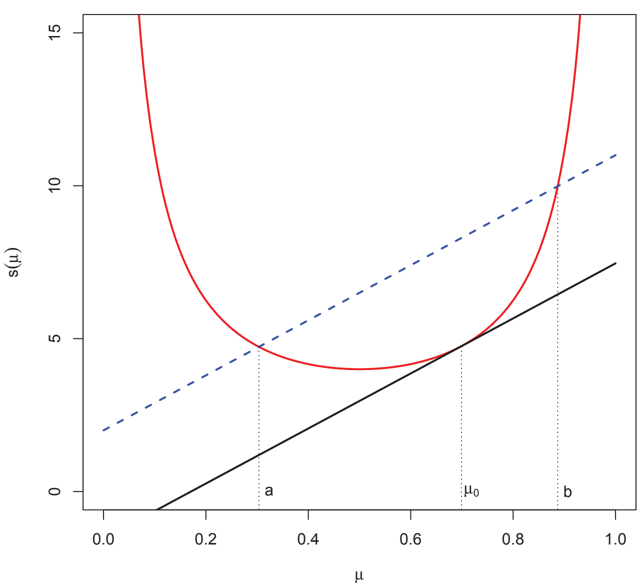
$$6\gamma = \frac{2\mu - 1}{\mu^2(1-\mu)^2}$$

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 $\beta \leq \beta_0$. In this region, there is a unique maximizer μ^* of $\ell_\gamma(\mu, \alpha, \beta)$.

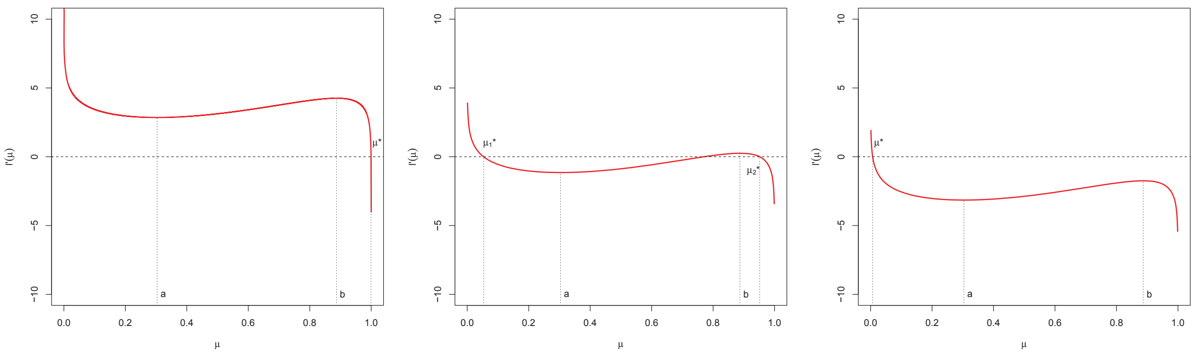
If $\beta > \beta_0$ the line $2\beta + 6\gamma\mu$ has two intersections with $s(\mu)$, and there are three possible cases. We know that in this region the second derivative $\ell''_\gamma(\mu, \alpha, \beta)$ can be positive or negative, with inflection points denoted as a and b, found by solving the equation $2\beta + 6\gamma\mu = s(\mu)$. In the picture below, 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 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 $\ell'_\gamma(a, \alpha, \beta) \geq 0$, then there is a unique maximizer $\mu^* > b$
2. If $\ell'_\gamma(b, \alpha, \beta) \leq 0$, then there is a unique maximizer $\mu^* < a$
3. If $\ell'_\gamma(a, \alpha, \beta) < 0 < \ell'_\gamma(b, \alpha, \beta)$, then there are 2 local maximizers $\mu_1^* < a < b < \mu_2^*$

The three cases are shown in the following pictures, 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.

Let's 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

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

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

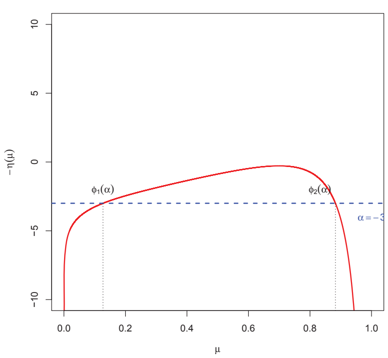
$$\eta(\mu_0) = \frac{1}{(1-\mu_0)} - \frac{2\mu_0 - 1}{2(1-\mu_0)^2} - \log\frac{\mu_0}{1-\mu_0}$$

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

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

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



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

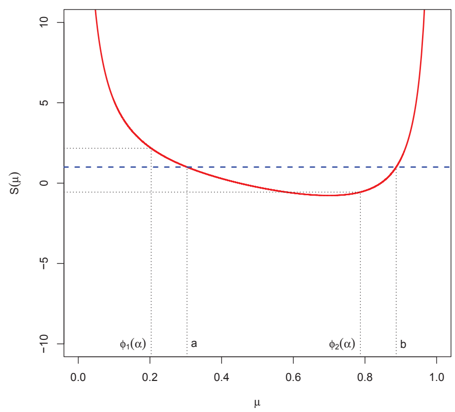
Consider the case in which $\ell'_\gamma(a, \alpha, \beta) < 0 < \ell'_\gamma(b, \alpha, \beta)$ with two maximizers of $\ell_\gamma(\mu, \alpha, \beta)$. Consider the function

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

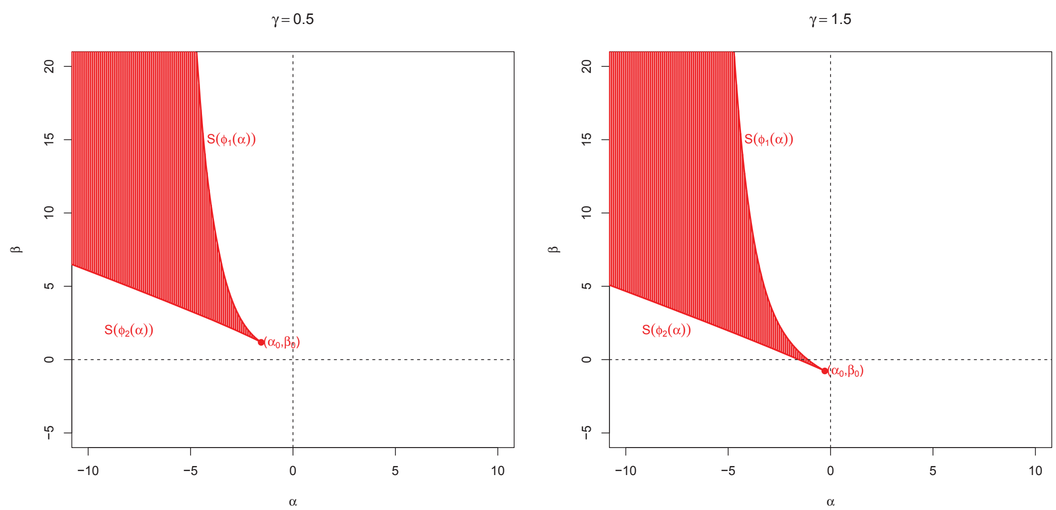
Since $\ell'_\gamma(a, \alpha, \beta) < 0$ we have $a > \phi_1(\alpha)$, which implies $S(a) < S(\phi_1(\alpha))$. Therefore $\beta < S(\phi_1(\alpha)) = \frac{1}{2\phi_1(\alpha)(1-\phi_1(\alpha))} - \frac{2\mu_0-1}{2\mu_0^2(1-\mu_0)^2}\phi_1(\alpha)$.

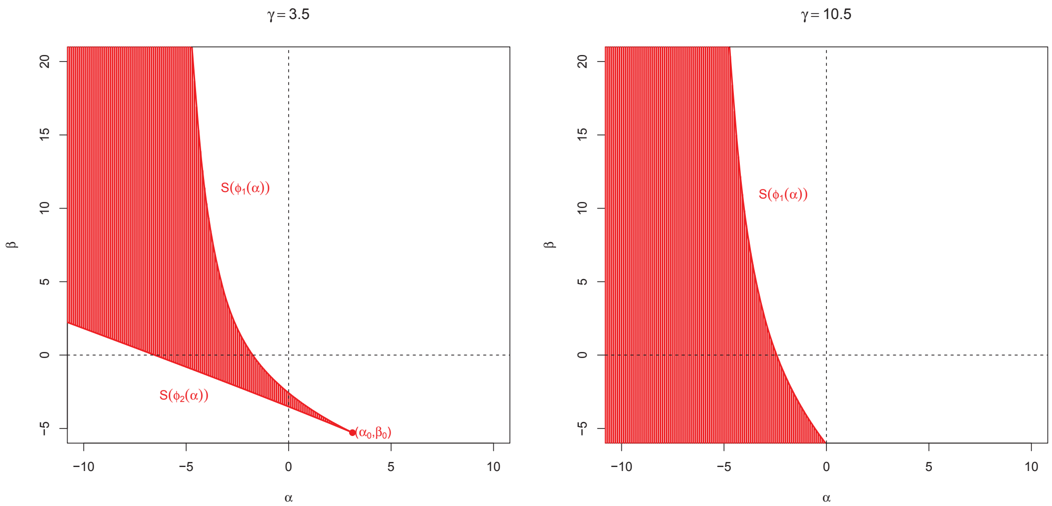
Since $\ell'_\gamma(b, \alpha, \beta) > 0$ we have $b > \phi_2(\alpha)$, which implies $S(b) > S(\phi_2(\alpha))$. Therefore $\beta > S(\phi_2(\alpha)) = \frac{1}{2\phi_2(\alpha)(1-\phi_2(\alpha))} - \frac{2\mu_0-1}{2\mu_0^2(1-\mu_0)^2}\phi_2(\alpha)$.

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



In the following pictures 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 argument as in the proof of Theorem 11, so it is omitted for brevity. ■

The last set of results extends the analysis of sampling algorithms in Bhamidi et al. (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 et al. (2011) call 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 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 (57). 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*

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

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

Proof. The proof is identical to the proof of Theorem 3.2 in Diaconis and Chatterjee (2011) ■

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

THEOREM 19 *Consider the model above in (58) 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 Erdos-Renyi graphs* $G_d(n, \mu_1^*)$ *and* $G_d(n, \mu_2^*)$*, as* $n \to \infty$.

Proof. It is an application of Theorem 18. ■

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, i.e. there exists a constant $C > 0$ such that the number of steps needed to reach stationarity are $O(e^{Cn})$. This is true for any local chain, i.e. a chain that updates $o(n)$ links per iteration.

The main convergence result that is proven in Bhamidi et al. (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, i.e. a dynamics that updates an* $o(n)$ *number of links per period.*

Proof. See Bhamidi et al. (2011), Thm. 5 and 6 ■

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 et al. (2011)) is proven above in our Theorem 19.