Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications 125 (2015) 4154–4177

stochastic processes and their applications

www.elsevier.com/locate/spa

Asymptotic structure and singularities in constrained directed graphs

David Aristoff, Lingjiong Zhu*

School of Mathematics, University of Minnesota-Twin Cities, 206 Church Street S.E., Minneapolis, MN-55455, United States

Received 3 June 2014; received in revised form 1 June 2015; accepted 6 June 2015
Available online 14 June 2015

Abstract

We study the asymptotics of large directed graphs, constrained to have certain densities of edges and/or outward p-stars. Our models are close cousins of exponential random graph models, in which edges and certain other subgraph densities are controlled by parameters. We find that large graphs have either uniform or bipodal structure. When edge density (resp. p-star density) is fixed and p-star density (resp. edge density) is controlled by a parameter, we find phase transitions corresponding to a change from uniform to bipodal structure. When both edge and p-star density are fixed, we find only bipodal structures and no phase transition.
© 2015 Elsevier B.V. All rights reserved.

MSC: 05C80; 82B26; 05C35

Keywords: Dense random graphs; Exponential random graphs; Graph limits; Entropy; Phase transitions

1. Introduction

In this article we study the asymptotics of large directed graphs with constrained densities of edges and outward directed p-stars. Large graphs are often modeled by probabilistic ensembles

* Corresponding author.
E-mail addresses: daristof@umn.edu (D. Aristoff), zhul@umn.edu (L. Zhu).

http://dx.doi.org/10.1016/j.spa.2015.06.004
0304-4149/© 2015 Elsevier B.V. All rights reserved.

with one or more adjustable parameters; see e.g. Fienberg [7,8], Lovász [18] and Newman [20]. The *exponential random graph models* (ERGMs), in which parameters are used to tune the densities of edges and other subgraphs, are one such class of models; see e.g. Besag [2], Frank and Strauss [9], Holland and Leinhardt [12], Newman [20], Rinaldo et al. [27], Robins et al. [28], Snijders et al. [29], Strauss [30], and Wasserman and Faust [32].

It has been shown that in many ERGMs the subgraph densities actually *cannot* be tuned. For example, for the class of ERGMs parametrized by edges and p-stars, large graphs are essentially Erdős–Rényi for all values of the parameters. (See Chatterjee and Diaconis [3] for more complete and precise statements.) An alternative to ERGMs was introduced by Radin and Sadun [24], where instead of using parameters to control subgraph counts, the subgraph densities are controlled directly; see also Radin et al. [23], Radin and Sadun [25] and Kenyon et al. [14]. This is the approach we take in this article, with edges and (outward) p-stars as the subgraphs controlled directly. We also consider models which split the difference between the two approaches: edges (resp. p-stars) are controlled directly, while p-stars (resp. edges) are tuned with a parameter. See [15] for the undirected graph version of the latter.

We find that, in all our models, graphs have either uniform or bipodal structure as the number of nodes becomes infinite. Our approach, following Refs. [14,24,23,25], is to study maximizers of the entropy or free energy of the model as the number of nodes becomes infinite. When we constrain both edge and p-star densities, we find only bipodal structure (except when the p-star density is fixed to be exactly equal to the pth power of the edge density). When we constrain either edge or p-star densities (but not both), we find both uniform and bipodal structures, with a sharp change at the interface. This is in contrast with the situation in the ERGM version of the model, in which one finds only uniform structures, albeit with sharp changes in their densities along a certain curve in parameter space; see Aristoff and Zhu [1].

These sharp changes along interfaces are called *phase transitions*. Phase transitions have recently been proved rigorously for ERGMs; see e.g. Yin [33] and especially Radin and Yin [26] for a precise definition of the term. Some earlier works using mean-field analysis and other approximations include Häggström and Jonasson [11] and Park and Newman [21,22]. The terminology is apt, in that ERGMs and our models are inspired by certain counterparts in statistical physics: respectively, the *grand canonical ensemble*, and the *microcanonical* and *canonical ensembles*. See the discussion in Section 2.

Our directed graph models are simpler than their undirected counterparts. In particular, we can rigorously identify the asymptotic structures at all parameters, while in analogous models for undirected graphs, only partial results are known [14]. Our analysis, however, does not easily extend to directed random graph models where other subgraphs, like triangles, cycles, etc., instead of outward directed p-stars, are constrained.

This article is organized as follows. In Section 2, we describe our models and compare them with their statistical physics counterparts. In Section 3, we state our main results. In Section 4, we prove a large deviations principle for edge and p-star densities. We use the large deviations principle to give proofs, in Section 5, of our main results.

2. Description of the models

A directed graph on n nodes will be represented by a matrix $X = (X_{ij})_{1\le i,j\le n}$, where $X_{ij} = 1$ if there is a directed edge from node i to node j, and $X_{ij} = 0$ otherwise. For simplicity, we allow for $X_{ii} = 1$, though this will not affect our results. Let $e(X)$ (resp. $s(X)$) be the directed edge and outward p-star homomorphism densities of X:

$$\begin{aligned} e(X) &:= n^{-2} \sum_{1 \le i,j \le n} X_{ij}, \\ s(X) &:= n^{-p-1} \sum_{1 \le i, j_1, j_2, \ldots, j_p \le n} X_{ij_1} X_{ij_2} \cdots X_{ij_p}. \end{aligned} \tag{2.1}$$

Here, p is an integer ≥ 2. The reason for the term homomorphism density is as follows. For a given graph X, if $\hom_e(X)$ (resp. $\hom_s(X)$) are the number of homomorphisms – edge-preserving maps – from a directed edge (resp. outward p-star) into X, then

$$e(X) = \frac{\underset{e}{\hom}(X)}{n^2}, \qquad s(X) = \frac{\underset{s}{\hom}(X)}{n^{p+1}},$$

with the denominators giving the total number of maps from a directed edge (resp. outward p-star) into X.

Let $\mathbb{P}_n$ be the uniform probability measure on the set of directed graphs on n nodes. Thus, $\mathbb{P}_n$ is the uniform probability measure on the set of $n \times n$ matrices with entries in $\{0, 1\}$. For $e, s \in [0, 1]$ and $\delta > 0$, define

$$\psi_n^\delta(e, s) = \frac{1}{n^2} \log \mathbb{P}_n \left(e(X) \in (e - \delta, e + \delta),\ s(X) \in (s - \delta, s + \delta) \right).$$

Throughout, log is the natural logarithm, and we use the convention $0 \log 0 = 0$. We are interested in the limit

$$\psi(e, s) := \lim_{\delta \to 0^+} \lim_{n \to \infty} \psi_n^\delta(e, s). \tag{2.2}$$

The function in (2.2) will be called the *limiting entropy density*. This is the directed graph version of the quantity studied in [14]. See also [24,23,25] for related work where triangles are constrained instead of p-stars.

For $\beta_1, \beta_2 \in \mathbb{R}$, $e, s \in [0, 1]$ and $\delta > 0$, define

$$\begin{aligned} \psi_n^\delta(e, \beta_2) &= \frac{1}{n^2} \log \mathbb{E}_n \left[e^{n^2 \beta_2 s(X)}\, 1_{\{e(X) \in (e-\delta, e+\delta)\}} \right], \\ \psi_n^\delta(\beta_1, s) &= \frac{1}{n^2} \log \mathbb{E}_n \left[e^{n^2 \beta_1 e(X)}\, 1_{\{s(X) \in (s-\delta, s+\delta)\}} \right]. \end{aligned}$$

We will also be interested in the limits

$$\begin{aligned} \psi(e, \beta_2) &:= \lim_{\delta \to 0^+} \lim_{n \to \infty} \psi_n^\delta(e, \beta_2), \\ \psi(\beta_1, s) &:= \lim_{\delta \to 0^+} \lim_{n \to \infty} \psi_n^\delta(\beta_1, s). \end{aligned} \tag{2.3}$$

The quantities in (2.3) will be called *limiting free energy densities*. They are the directed graph versions of the quantities studied in [15]. We abuse notation by using the same symbols ψ and ψ_n^δ to represent different functions in (2.2) and (2.3) (and in (2.4)), but the meaning will be clear from the arguments of these functions, which will be written (e, s), (e, β_2), (β_1, s) (or (β_1, β_2) as below). When it is clear which function we refer to, we may simply write ψ without any arguments.

We show that the limits in (2.2) and (2.3) exist by appealing to the variational principles in Theorems 6 and 7. Maximizers of these variational problems are associated with the asymptotic structure of the associated constrained directed graphs. More precisely, as $n \to \infty$ a typical sample X from the uniform measure $\mathbb{P}_n$ conditioned on the event $e(X) = e$, $s(X) = s$ looks like the maximizer of the variational formula for (2.2) in Theorem 6. Similarly, as $n \to \infty$, a typical sample from the probability measure which gives weight $\exp[n^2\beta_2 s]$ (resp. $\exp[n^2\beta_1 e]$) to graphs X on n nodes with $s(X) = s$ (resp. $e(X) = e$), when conditioned on the event $e(X) = e$ (resp. $s(X) = s$), looks like the maximizer of the variational formula for (2.3) in Theorem 7. Thus, in (2.2) we have constrained $e(X)$ and $s(X)$ directly, while in (2.3) we control one of $e(X)$ or $s(X)$ with a parameter β_1 or β_2. See [14,15,26,24,23,25] for discussions and related work in the undirected graph setting. Closely related to (2.2) and (2.3) is the limit

$$\psi(\beta_1, \beta_2) := \lim_{n\to\infty} \frac{1}{n^2} \log \mathbb{E}\left[e^{n^2(\beta_1 e(X)+\beta_2 s(X))}\right], \tag{2.4}$$

which was studied extensively in Aristoff and Zhu [1]. In this *exponential random graph model* (ERGM) both $e(X)$ and $s(X)$ are controlled by parameters β_1 and β_2. There it was shown that $\psi(\beta_1, \beta_2)$ is analytic except along a certain phase transition curve $\beta_2 = q(\beta_1)$, corresponding to an interface across which the edge density changes sharply. See Radin and Yin [26] for similar results, which use a similar variational characterization, in the undirected graph setting. The results in [1,26] will be useful in our analysis. This is because the Euler–Lagrange equations associated with the variational formulas for (2.2), (2.3) are the same as the analogous variational formula for (2.4). The crucial difference in the optimization problems is that for the ERGM, β_1 and β_2 are fixed parameters, while in our models one or both of β_1, β_2 is a free variable (i.e., a Lagrange multiplier). Thus, some solutions to the ERGM variational problem may have no relevance to our models. However, many of the calculations can be carried over, which is why we can make extensive use of the results in [1,26].

This article focuses on computing the limiting entropy density and free energy densities in (2.2) and (2.3), along with the corresponding maximizers in the variational formulas of Theorems 6 and 7. In statistical physics modeling there is a hierarchy analogous to (2.2)–(2.3)–(2.4), with (for example) particle density and energy density in place of $e(X)$ and $s(X)$, and temperature and chemical potential in place of β_1 and β_2. The statistical physics versions of (2.2)–(2.3)–(2.4) correspond to the *microcanonical*, *canonical* and *grand canonical ensembles*, respectively. In that setting, there are curves like q along which the free energy densities are not analytic, and these correspond to physical phase transitions, for example the familiar solid/liquid and liquid gas transitions; see Gallavotti [10]. (There is no proof of this statement, though it is widely believed; see however Lebowitz et al. [17].) We find singularities in $\psi(e, \beta_2)$ and $\psi(\beta_1, s)$, but not in $\psi(e, s)$. (In [1] we have shown $\psi(\beta_1, \beta_2)$ has singularities as well.) See Radin [26] and Radin and Sadun [25] for discussions of the relationship between these statistical physics models and some random graph models that closely parallel ours.

3. Results

To state our results we need the following. For $x \in [0, 1]$, define

$$\ell(x) = \beta_1 x + \beta_2 x^p - x \log x - (1-x)\log(1-x),$$
$$I(x) = x \log x + (1-x)\log(1-x) + \log 2,$$

with the understanding that $I(0) = I(1) = \log 2$. Of course ℓ depends on β_1 and β_2, but we omit this to simplify notation. Clearly, ℓ and I are analytic in $(0, 1)$ and continuous on $[0, 1]$. The function ℓ is essential to understanding the ERGM limiting free energy density [1,26].

Theorem 1 (*Radin and Yin [26]*)**.** *For each (β_1, β_2) the function ℓ has either one or two local maximizers. There is a curve $\beta_2 = q(\beta_1)$, $\beta_1 \le \beta_1^c$, with the endpoint*

$$(\beta_1^c, \beta_2^c) = \left(\log(p-1) - \frac{p}{p-1}, \frac{p^{p-1}}{(p-1)^p}\right),$$

such that off the curve and at the endpoint, ℓ has a unique global maximizer, while on the curve away from the endpoint, ℓ has two global maximizers $0 < x_1 < x_2 < 1$. We consider x_1 and x_2 as functions of β_1 (or β_2) for $\beta_1 < \beta_1^c$ (or $\beta_2 > \beta_2^c$); x_1 (resp. x_2) is increasing (resp. decreasing) in β_1, with

$$\begin{aligned} &\lim_{\beta_1 \to -\infty} x_1 = 0, \qquad \lim_{\beta_1 \to -\infty} x_2 = 1, \\ &\lim_{\beta_1 \to \beta_1^c} x_1 = \frac{p-1}{p} = \lim_{\beta_1 \to \beta_1^c} x_2. \end{aligned} \tag{3.1}$$

Theorem 2 (*Aristoff and Zhu [1]*)**.** *The curve q in Theorem 1 is continuous, decreasing, convex, and analytic for $\beta_1 < \beta_1^c$, with*

$$q'(\beta_1) = -\frac{x_2 - x_1}{x_2^p - x_1^p}.$$

Moreover, x_1 and x_2 are analytic in β_1 and β_2.

The curve in Theorem 1 will be called the *phase transition curve*, and its endpoint the *critical point*. Theorems 1 and 2 will be used extensively in most of our proofs; because of this, we will often not refer to it explicitly. (We comment that the last statement of Theorem 1, not made explicit in [26,1], is proved in Proposition 8 in Section 5.) We will sometimes write $x_1 = x_2 = (p-1)/p$ for the local maximizer of ℓ at the critical point. Note that the last part of the theorem implies

$$0 < x_1 \le \frac{p-1}{p} \le x_2 < 1.$$

Our first main result concerns the limiting free energy density $\psi(e, s)$.

Theorem 3. *The limiting entropy density $\psi = \psi(e, s)$ is analytic in*

$$D := \{(e, s) : e^p < s < e,\ 0 \le e \le 1\}$$

and equals $-\infty$ outside $\bar{D}$. With

$$(e^c, s^c) = \left(\frac{p-1}{p}, \left(\frac{p-1}{p}\right)^p\right)$$

and

$$D^{up} = \{(e, s) : s = e,\ 0 \le e \le 1\},$$

we have the semi-explicit formula

$$\psi(e, s) = \begin{cases} -\dfrac{x_2 - e}{x_2 - x_1} I(x_1) - \dfrac{e - x_1}{x_2 - x_1} I(x_2), & (e, s) \in D \\ -I(e), & (e, s) \in \partial D \setminus D^{up} \\ -\log 2, & (e, s) \in D^{up}, \end{cases} \tag{3.2}$$

where $0 < x_1 < x_2 < 1$ *are the unique global maximizers of* ℓ *along the phase transition curve* q *which satisfy*

$$\frac{x_2 - e}{x_2 - x_1} x_1^p + \frac{e - x_1}{x_2 - x_1} x_2^p = s.$$

Moreover, ψ *is continuous on* $\bar{D}$*, and the first order partial derivatives of* ψ *are continuous on* $\bar{D} \setminus D^{up}$ *but diverge on* D^{up}*. When* $p = 2$*, we have the explicit formula*

$$\psi(e, s) = -I\left(\frac{1}{2} + \sqrt{s - e + \frac{1}{4}}\right), \quad (e, s) \in \bar{D}.$$

In Theorem 3, the formula is only semi-explicit because x_1 and x_2 depend on e and s. Note that the formula depends *only* on e and s. (To understand the dependence better, see the subsection *uniqueness of the optimizer: geometric proof* in the proof of Theorem 3.)

Note that we have not found any interesting singular behavior of ψ, in contrast with the recent results of Kenyon et al. [14] for the undirected version of the model. On the other hand, we are able to prove that our graphs are bipodal, whereas in [14] it was proved only that the graphs are multipodal (though simulation evidence from the paper suggests bipodal structure).

Our next main result concerns $\psi(e, \beta_2)$ and $\psi(\beta_1, s)$.

Theorem 4. (i) *There is a U-shaped region*

$$U_e = \{(e, \beta_2) : x_1 < e < x_2,\ \beta_2 > \beta_2^c\}$$

whose closure has lowest point

$$\left(e^c, \beta_2^c\right) = \left(\frac{p-1}{p}, \frac{p^{p-1}}{(p-1)^p}\right)$$

such that the limiting free energy density $\psi = \psi(e, \beta_2)$ *is analytic outside* ∂U_e*. The limiting free energy density has the formula*

$$\psi(e, \beta_2) = \begin{cases} \beta_2 e^p - I(e), & (e, \beta_2) \in U_e^c \\ \beta_2 \left[\dfrac{x_2 - e}{x_2 - x_1} x_1^p + \dfrac{e - x_1}{x_2 - x_1} x_2^p\right] & \\ \quad - \left[\dfrac{x_2 - e}{x_2 - x_1} I(x_1) + \dfrac{e - x_1}{x_2 - x_1} I(x_2)\right], & (e, \beta_2) \in U_e \end{cases} \tag{3.3}$$

where $0 < x_1 < x_2 < 1$ *are the global maximizers of* ℓ *at the point* $(q^{-1}(\beta_2), \beta_2)$ *on the phase transition curve. In particular,* $\partial^2\psi/\partial e^2$ *has jump discontinuities across* ∂U_e *away from* (e^c, β_2^c)*, and* $\partial^4\psi/\partial e^4$ *is discontinuous at* (e^c, β_2^c)*.*

(ii) *There is a U-shaped region*

$$U_s = \{(\beta_1, s) : x_1^p < s < x_2^p,\ \beta_1 < \beta_1^c\}$$

whose closure has rightmost point

$$\left(\beta_1^c, s^c\right) = \left(\log(p-1) - \frac{p}{p-1}, \left(\frac{p-1}{p}\right)^p\right)$$

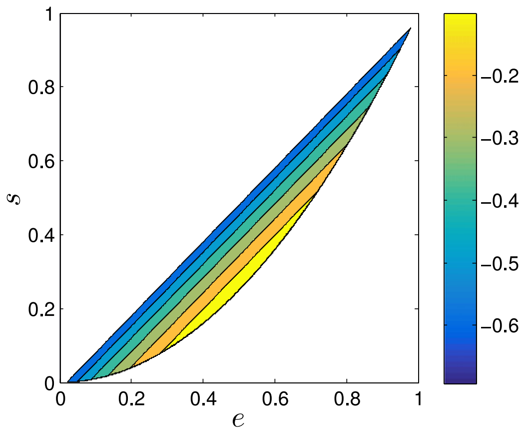

Fig. 1. Contour plot of $\psi(e,s)$ for $p=2$. The non-shaded portion represents the complement of $\bar{D}$ (where by convention ψ equals $-\infty$). Here, $(e^c,s^c)=(1/2,1/4)$.

such that the limiting free energy density $\psi=\psi(\beta_1,s)$ is analytic outside ∂U_s. The limiting free energy density has the formula

$$\psi(\beta_1,s)=\begin{cases}\beta_1 s^{\frac{1}{p}}-I\left(s^{\frac{1}{p}}\right), & (\beta_1,s)\in U_s^c\\ \beta_1\left[\dfrac{x_2^p-s}{x_2^p-x_1^p}x_1+\dfrac{s-x_1^p}{x_2^p-x_1^p}x_2\right] & \\ \quad-\left[\dfrac{x_2^p-s}{x_2^p-x_1^p}I(x_1)+\dfrac{s-x_1^p}{x_2^p-x_1^p}I(x_2)\right], & (\beta_1,s)\in U_s\end{cases}\tag{3.4}$$

where $0<x_1<x_2<1$ are the global maximizers of ℓ at the point $(\beta_1,q(\beta_1))$ on the phase transition curve. In particular, $\partial^2\psi/\partial s^2$ has jump discontinuities across ∂U_e away from (β_1^c,s^c), and $\partial^4\psi/\partial s^4$ is discontinuous at (β_1^c,s^c).

The sharp change in Theorem 4 is called a *phase transition* (because of the singularity in the derivatives of ψ). The phase transition corresponds to a sharp change from uniform to bipodal structure of the optimizers of the variational problem for (2.3); see the discussion below Theorem 7. The bipodal structure in the regions U_e and U_s is sometimes called a *replica symmetry breaking phase*.

In contrast with Theorem 3, the x_1 and x_2 in Theorem 4 are functions of β_1 or β_2 only and do not depend on e or s. See Fig. 1 for the region D from Theorem 3, and Fig. 2 for the U-shaped regions U_e and U_s in Theorem 4.

4. Large deviations

We will need the following terminology before proceeding. A sequence $(\mathbb{Q}_n)_{n\in\mathbb{N}}$ of probability measures on a topological space X is said to satisfy a *large deviation principle* with speed a_n and rate function $J:X\to\mathbb{R}$ if J is non-negative and lower semicontinuous, and for any measurable set A,

$$-\inf_{x\in A^o}J(x)\le\liminf_{n\to\infty}\frac{1}{a_n}\log\mathbb{Q}_n(A)\le\limsup_{n\to\infty}\frac{1}{a_n}\log\mathbb{Q}_n(A)\le-\inf_{x\in\bar{A}}J(x).$$

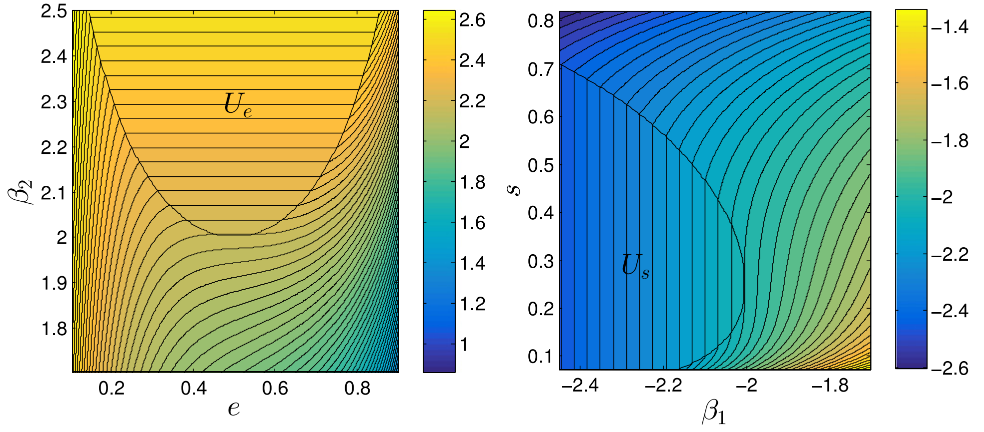

Fig. 2. Contour plots of $\partial\psi(e,\beta_2)/\partial e$ (left) and $\partial\psi(\beta_1,s)/\partial s$ (right) for $p = 2$. Contour lines are included to emphasize features. The boundaries of the U-shaped regions, ∂U_e and ∂U_s, are outlined. Here, $(e^c,\beta_2^c) = (1/2,2)$ and $(\beta_1^c,s^c) = (-2,1/4)$.

Here, A^o is the interior of A and $\bar{A}$ is its closure. See e.g. Dembo and Zeitouni [6] or Varadhan [31].

We will equip the set $\mathcal{G}$ of measurable functions $[0,1] \to [0,1]$ with the *cut norm*, written $\|\cdot\|_\Box$ and defined by

$$\|g\|_\Box = \sup \left| \int_A g(x)\,dx \right|, \tag{4.1}$$

where the supremum is taken over measurable subsets A of $[0,1]$.

Theorem 5. *The sequence of probability measures $\mathbb{P}_n\,(e(X) \in \cdot,\ s(X) \in \cdot)$ satisfies a large deviation principle on the space $[0,1]^2$ with speed n^2 and rate function*

$$J(e,s) = \inf_{g\in\mathcal{G}_{e,s}} \int_0^1 I(g(x))\,dx,$$

where $\mathcal{G}_{e,s}$ is the set of measurable functions $g : [0,1] \to [0,1]$ such that

$$\int_0^1 g(x)\,dx = e, \qquad \int_0^1 g(x)^p\,dx = s.$$

By convention, $I(x) = \infty$ if $x \notin [0,1]$, and the infimum is ∞ if $\mathcal{G}_{e,s}$ is empty.

Theorems 3 and 5 together imply that for large number of nodes, a typical sample X from $\mathbb{P}_n$ conditioned on the event $e(X) \approx e$ and $s(X) \approx s$ has the following behavior. Approximately $n(x_2 - e)/(x_2 - x_1)$ of the nodes of X each have on average nx_1 outward pointing edges, while the other approximately $n(e - x_1)(x_2 - x_1)$ nodes each have on average nx_2 outward pointing edges. When $x_1 \neq x_2$ we call this structure *bipodal*; otherwise we call it *uniform*.

The following is an immediate consequence of Theorem 5.

Theorem 6. *For any $e, s \in [0,1]$,*

$$\psi(e,s) = \sup_{\mathcal{G}_{e,s}} \left[-\int_0^1 I(g(x))\,dx \right].$$

The next theorem is an easy consequence of Theorem 5 and Varadhan's lemma (see e.g. [6]).

Theorem 7. *For any $e, s \in [0, 1]$ and $\beta_1, \beta_2 \in \mathbb{R}$,*

$$\psi(e, \beta_2) = \sup_{\mathcal{G}_{e,\cdot}} \left[\beta_2 \int_0^1 g(x)^p \, dx - \int_0^1 I(g(x)) \, dx \right]$$

$$\psi(\beta_1, s) = \sup_{\mathcal{G}_{\cdot,s}} \left[\beta_1 \int_0^1 g(x) \, dx - \int_0^1 I(g(x)) \, dx \right],$$

where $\mathcal{G}_{e,\cdot}$ (resp. $\mathcal{G}_{\cdot,s}$) is the set of measurable functions $g : [0, 1] \to [0, 1]$ satisfying

$$\int_0^1 g(x) \, dx = e \quad \left(\text{resp. } \int_0^1 g(x)^p \, dx = s \right).$$

From Theorem 5, the proofs of Theorems 6 and 7 are standard, so we omit them.

Chatterjee and Varadhan [4] established large deviations for undirected random graphs on the space of graphons (see also Lovász [19]). Szemerédi's lemma was needed in order to establish the compactness needed for large deviations. Since our model consists of directed graphs, the results in [4] do not apply directly. Theorem 5 avoids these technical difficulties: it is large deviations principle only on the space $[0, 1]^2$ of edge and star densities, instead of on the (quotient) function space of graphons. Our proof relies on the simplicity of our edge/directed p-star model; it cannot be easily extended to the case where edges and directed triangles (or other more complicated directed subgraphs) are constrained. We expect that these models can be handled by adapting the results of [4] to the directed case.

In Aristoff and Zhu [1], it was proved that

$$\lim_{n\to\infty} \frac{1}{n^2} \log \mathbb{E}\left[e^{n^2(\beta_1 e(X) + \beta_2 s(X))} \right] = \sup_{0 \le x \le 1} \left(\beta_1 x + \beta_2 x^p - I(x) \right). \tag{4.2}$$

Observe that the Gärtner–Ellis theorem cannot be used to obtain the large deviations principle in Theorem 5, due to the fact, first observed in Radin and Yin [26], that the right hand side of (4.2) is not differentiable. Instead, we used Mogulskii's theorem and the contraction principle in the proof of Theorem 5.

On the other hand, once we have established the large deviations principle in Theorem 5, we can use Varadhan's lemma to obtain an alternative expression for the limiting free energy,

$$\lim_{n\to\infty} \frac{1}{n^2} \log \mathbb{E}\left[e^{n^2(\beta_1 e(X) + \beta_2 s(X))} \right] = \sup_{(e,s)\in[0,1]^2} \left(\beta_1 e + \beta_2 s + \psi(e, s) \right).$$

The limiting free energy in the directed and undirected models differs by only a constant factor of $1/2$ (see Chatterjee and Diaconis [3] and Radin and Yin [26] for the undirected model, and Aristoff and Zhu [1] for the directed model). On the other hand, the limiting entropy density $\psi(e, s)$ we obtain here differs nontrivially from the one recently obtained in Kenyon et al. [14] for undirected graphs.

5. Proofs

We start with the proof of the large deviations principle of Section 4. Then we turn to the proofs of our main results in Section 3.

Proof of Theorem 5. By definition, $\mathbb{P}_n$ is the uniform probability measure on the set of directed graphs on n nodes. Recall that

$$e(X) = n^{-2} \sum_{1\leq i,j\leq n} X_{ij},$$

$$s(X) = n^{-p-1} \sum_{1\leq i\leq n} \left(\sum_{1\leq j\leq n} X_{ij} \right)^p .$$

Under $\mathbb{P}_n$, $(X_{ij})_{1\leq i,j\leq n}$ are i.i.d. Bernoulli random variables that take the value 1 with probability $\frac{1}{2}$ and 0 with probability $\frac{1}{2}$. Therefore, the logarithm of the moment generating function of X_{ij} is, for $\theta \in \mathbb{R}$,

$$\log \mathbb{E}[e^{\theta X_{ij}}] = \log \left(\frac{1}{2} e^{\theta} + \frac{1}{2} \right) = \log(e^{\theta} + 1) - \log 2.$$

Its Legendre transform is

$$\sup_{\theta\in\mathbb{R}} (\theta x - \log \mathbb{E}[e^{\theta X_{ij}}]) = x \log x + (1 - x) \log(1 - x) + \log 2 = I(x),$$

where by convention $I(x) = \infty$ for $x \notin [0, 1]$. Let Y_i be the ith entry of the vector

$$(X_{11}, X_{12}, \ldots, X_{1n}, X_{21}, X_{22}, \ldots, X_{2n}, \ldots, X_{n1}, X_{n2}, \ldots, X_{nn}) .$$

Then, Y_i are i.i.d. Bernoulli random variables. Mogulskii theorem (see e.g. Dembo and Zeitouni [6]) shows that

$$\mathbb{P} \left(\frac{1}{n^2} \sum_{i=1}^{\lfloor n^2 x \rfloor} Y_i \in \cdot, 0 \leq x \leq 1 \right)$$

satisfies a sample path large deviation principle on the space $L_\infty[0, 1]$ consisting of functions on $[0, 1]$ equipped with the supremum norm; the rate function is given by

$$\mathcal{I}(G) = \begin{cases} \int_0^1 I(G'(x)) dx & \text{if } G \in \mathcal{AC}_0[0, 1] \\ +\infty & \text{otherwise,} \end{cases}$$

where $\mathcal{AC}_0[0, 1]$ is the set of absolutely continuous functions defined on $[0, 1]$ such that $G(0) = 0$ and $0 \leq G'(x) \leq 1$. The restriction $0 \leq G'(x) \leq 1$ comes from the fact that $0 \leq Y_i \leq 1$. On the other hand, for any $\epsilon > 0$,

$$\begin{aligned} &\limsup_{n\to\infty} \frac{1}{n^2} \log \mathbb{P} \left(\sup_{0\leq x\leq 1} \left| \frac{1}{n^2} \sum_{i=1}^{\lfloor nx \rfloor} \sum_{j=1}^{n} X_{ij} - \frac{1}{n^2} \sum_{i=1}^{\lfloor n^2 x \rfloor} Y_i \right| \geq \epsilon \right) \\ &\quad \leq \limsup_{n\to\infty} \frac{1}{n^2} \log \mathbb{P} \left(\sup_{0\leq i\leq n^2-n} (Y_{i+1} + Y_{i+2} + \cdots Y_{i+n}) \geq n^2 \epsilon \right) \\ &\quad = -\infty, \end{aligned}$$

since

$$\sup_{0\leq i\leq n^2-n} (Y_{i+1} + Y_{i+2} + \cdots Y_{i+n}) \leq n.$$

Therefore,

$$\mathbb{P}\left(\frac{1}{n^2}\sum_{i=1}^{\lfloor nx\rfloor}\sum_{j=1}^{n} X_{ij} \in \cdot, 0 \le x \le 1\right)$$

satisfies a large deviation principle with the same space and rate function as

$$\mathbb{P}\left(\frac{1}{n^2}\sum_{i=1}^{\lfloor n^2 x\rfloor} Y_i \in \cdot, 0 \le x \le 1\right).$$

To complete the proof, we need to use the contraction principle, see e.g. Dembo and Zeitouni [6] or Varadhan [31]. Given $G \in \mathcal{AC}_0[0, 1]$, we may write $G(x) = \int_0^x g(y)dy$ for a measurable function $g : [0, 1] \to [0, 1]$. It is easy to see that if $G_n \in \mathcal{AC}_0[0, 1]$ and $G_n \to G$ in the supremum norm, then $g_n \to g$ in the cut norm. Hence, if $G_n \to G$ in the supremum norm, then

$$\int_0^1 g_n(x)dx \to \int_0^1 g(x)dx.$$

Moreover, for any $p \ge 2$,

$$\int_0^1 (g_n(x)^p - g(x)^p)dx = \left[\int_{g_n \ge g} (g_n(x)^p - g(x)^p)dx\right] - \left[\int_{g_n < g} (g(x)^p - g_n(x)^p)dx\right],$$

and since $g_n \to g$ in the cut norm and $0 \le g_n, g \le 1$, the mean value theorem shows that

$$\int_{g_n \ge g} (g_n^p(x) - g(x)^p)dx \le p\int_{g_n \ge g} (g_n(x) - g(x))dx \to 0$$

and

$$\int_{g_n < g} (g^p(x) - g_n(x)^p)dx \le p\int_{g_n < g} (g(x) - g_n(x))dx \to 0.$$

Therefore, the maps

$$G \mapsto \int_0^1 G'(x)dx, \qquad G \mapsto \int_0^1 (G'(x))^p dx,$$

are continuous from $L_\infty[0, 1] \cap \mathcal{AC}_0[0, 1]$ to $[0, 1]$ and thus the map

$$G \mapsto \left(\int_0^1 G'(x)dx, \int_0^1 (G'(x))^p dx\right)$$

is continuous from $L_\infty[0, 1] \cap \mathcal{AC}_0[0, 1]$ to $[0, 1]^2$. By the contraction principle, we conclude that $\mathbb{P}_n\,(e(X) \in \cdot,\ s(X) \in \cdot)$ satisfies a large deviation principle on the space $[0, 1]^2$ with speed n^2 and rate function $J(e, s) = \inf_{g \in \mathcal{G}_{e,s}} \int_0^1 I(g(x))\,dx$. □

The following proofs are for our main results in Section 3.

Proof of Theorem 3. By Theorem 6,

$$\psi(e, s) = \sup_{g \in \mathcal{G}_{e,s}}\left[-\int_0^1 I(g(x))\,dx\right] \tag{5.1}$$

where we recall $\mathcal{G}_{e,s}$ is the set of measurable functions $g : [0, 1] \to [0, 1]$ satisfying

$$\int_0^1 g(x)\,dx = e, \qquad \int_0^1 g(x)^p\,dx = s. \tag{5.2}$$

Domain of $\psi(e, s)$. By Jensen's inequality,

$$s = \int_0^1 g(x)^p\,dx \ge \left(\int_0^1 g(x)\,dx\right)^p = e^p.$$

On the other hand, since g takes values in $[0, 1]$ and $p > 1$,

$$s = \int_0^1 g(x)^p\,dx \le \int_0^1 g(x)\,dx = e.$$

Hence by convention $\psi(e, s)$ is infinite outside $\bar{D}$. We will show later in the proof that $\psi(e, s)$ is finite in $\bar{D}$.

Optimizing over a quotient space. So that we can use standard variational techniques, we show that it suffices to optimize over a compact quotient space of $\mathcal{G}_{e,s}$. Recall that $\mathcal{G}$ is the set of measurable functions $[0, 1] \to [0, 1]$ endowed with the cut norm (4.1). Let $\tilde{\mathcal{G}}$ be the quotient space obtained by the following equivalence relation: $g \sim h$ if and only if there is a measure-preserving bijection $\sigma : [0, 1] \to [0, 1]$ such that $g = h \circ \sigma$. Note that for any such σ we have $\|g\|_\square = \|g \circ \sigma\|_\square$, and so

$$\delta_\square(\tilde{g}, \tilde{h}) := \inf_\sigma \|g - h \circ \sigma\|_\square$$

defines a metric on $\tilde{\mathcal{G}}$. With this metric $\tilde{\mathcal{G}}$ is compact [13]. Observe that

$$-\int_0^1 I(g(x))\,dx, \qquad \int_0^1 g(x)\,dx, \qquad \int_0^1 g(x)^p\,dx$$

are all unchanged when g is replaced by $g \circ \sigma$. Thus,

$$\psi(e, s) = \sup_{g \in \mathcal{G}_{e,s}} \left[-\int_0^1 I(g(x))\,dx\right] = \sup_{\tilde{g} \in \tilde{\mathcal{G}}_{e,s}} \left[-\int_0^1 I(g(x))\,dx\right],$$

where $\tilde{\mathcal{G}}_{e,s}$ is the projection of $\mathcal{G}_{e,s}$ in the quotient space, and on the right hand side g is any element of the equivalence class of $\tilde{g}$. Next we check that the functionals

$$\tilde{g} \mapsto -\int_0^1 I(g(x))\,dx, \qquad \tilde{g} \mapsto \int_0^1 g(x)\,dx, \qquad \tilde{g} \mapsto \int_0^1 g(x)^p\,dx, \tag{5.3}$$

defined on $\tilde{\mathcal{G}}$, are all continuous. Let $\tilde{g}_n$ be a sequence in $\tilde{\mathcal{G}}$ converging to $\tilde{g}$, and let g_n, g be representatives for $\tilde{g}_n, \tilde{g}$, respectively. We will show that

$$\left|\int_0^1 F(g(x))\,dx - \int_0^1 F(g_n(x))\,dx\right| \to 0 \quad \text{as } n \to \infty \tag{5.4}$$

for any uniformly continuous function F defined on $[0, 1]$. For each n, choose σ_n such that

$$\|g - g_n \circ \sigma_n\|_\square \to 0 \quad \text{as } n \to \infty.$$

For any $\delta > 0$, define

$$A_{n,\delta} = \{x : |g(x) - g_n(\sigma_n(x))| \geq \delta\}.$$

As the cut norm is equivalent to the L^1 norm (this is true only in one dimension; see [13]), we have for any $\delta > 0$,

$$|A_{n,\delta}| \to 0 \quad \text{as } n \to \infty, \tag{5.5}$$

where $|A_{n,\delta}|$ is the Lebesgue measure of $A_{n,\delta}$. Now, observe that

$$\begin{aligned}
&\left|\int_0^1 F(g(x))\,dx - \int_0^1 F(g_n(x))\,dx\right| \\
&\quad = \left|\int_0^1 F(g(x))\,dx - \int_0^1 F(g_n \circ \sigma_n(x))\,dx\right| \\
&\quad \leq 2|A_{n,\delta}| \sup_{x\in[0,1]} |F(x)| + \left|\int_{[0,1]\setminus A_{n,\delta}} [F(g(x)) - F(g_n \circ \sigma_n(x))]\,dx\right|.
\end{aligned} \tag{5.6}$$

Let $\epsilon > 0$. Using uniform continuity of F, let $\delta > 0$ be such that $|x - y| < \delta$ implies $|F(x) - F(y)| < \epsilon$. Then (5.6) shows that

$$\left|\int_0^1 F(g(x))\,dx - \int_0^1 F(g_n(x))\,dx\right| \leq 2|A_{n,\delta}| \sup_{x\in[0,1]} |F(x)| + \epsilon. \tag{5.7}$$

Now (5.5) establishes (5.4), as desired. These arguments show that there exist (global) maximizers g of

$$\psi(e, s) = \sup_{g\in\mathcal{G}_{e,s}} \left[-\int_0^1 I(g(x))\,dx\right]. \tag{5.8}$$

Bipodal structure of the optimizers. If $(e, s) \in D^{up}$, then for any $g \in \mathcal{G}_{e,s}$,

$$|\{x : g(x) \in \{0, 1\}\}| = 1.$$

Suppose then that $(e, s) \in \bar{D} \setminus D^{up}$. Since $I'(x) \to -\infty$ as $x \to 0$ and $I'(x) \to \infty$ as $x \to 1$, it is not hard to see that, given $(s, e) \in \bar{D} \setminus D^{up}$, there exists $\epsilon > 0$ such that any optimizers of (5.8) must be in the following set:

$$\mathcal{G}^\epsilon := \{g \in \mathcal{G} : g(x) \in [\epsilon, 1 - \epsilon] \text{ for a.e. } x \in [0, 1]\}. \tag{5.9}$$

To see this, note that if $g \in \mathcal{G}_{e,s} \setminus \mathcal{G}^\epsilon$, then the values of g can be adjusted to be ϵ distance away from 0 and 1 so that still $g \in \mathcal{G}_{e,s}$ and g attains a larger value for the integral in (5.8). Then, optimizing over $\mathcal{G}^\epsilon$, standard results in variational calculus (see [5, Theorem 9.1, pg. 178]) show that optimizers g satisfy, for all $\delta g \in \mathcal{G}$,

$$-\eta \int_0^1 I'(g(x))\delta g(x)\,dx + \beta_1 \int_0^1 \delta g(x)\,dx + \beta_2 \int_0^1 p g(x)^{p-1} \delta g(x)\,dx = 0, \tag{5.10}$$

where $\eta \in \{0, 1\}$, $\beta_1, \beta_2 \in \mathbb{R}$ are Lagrange multipliers, and at least one of η, β_1, β_2 is nonzero. The integrals in (5.10) are the Fréchet derivatives of

$$g \mapsto \int_0^1 I(g(x))\,dx, \qquad g \mapsto \int_0^1 g(x)\,dx, \qquad g \mapsto \int_0^1 g(x)^p\,dx,$$

evaluated at $\delta g \in \mathcal{G}$; a simplified version of the arguments in (5.4)–(5.7) shows the Fréchet derivatives are continuous in $g \in \mathcal{G}^\epsilon$. (To apply the results mentioned in [5], we must switch to the uniform topology induced by the sup norm $\|\cdot\|_\infty$ on the Banach space of bounded measurable functions $[0,1] \to \mathbb{R}$. The statement about continuity of Fréchet derivatives then follows from the simple fact that $\|g-h\|_\infty < \epsilon$ implies $\|g-h\|_1 < \epsilon$.)

When $\eta = 1$ (called the *normal* case), we see that for some β_1, β_2,

$$\ell'(g(x)) = 0, \quad \text{for a.e. } x \in [0,1].$$

When $\eta = 0$ (called the *abnormal* case), we find that $\beta_1 + p\beta_2 g(x)^{p-1} = 0$ for a.e. $x \in [0,1]$, so that g is constant a.e. From (5.2) and Jensen's inequality, this occurs precisely when $s = e^p$.

Let g be a maximizer of (5.1). We have shown that:

(a) if $(e,s) \in D^{up}$, then $g(x) \in \{0,1\}$ for a.e. $x \in [0,1]$;

(b) if $(e,s) \in D$, then $\ell'(g(x)) = 0$ for a.e. $x \in [0,1]$; or

(c) if $s \in \partial D \setminus D^{up}$, then g is constant a.e.

Theorem 1 shows that, for each (β_1, β_2), either $\ell'(y) = 0$ at a unique $y \in (0,1)$ or $\ell'(y) = 0$ at exactly two points $0 < y_1 < y_2 < 1$. Thus, to maximize (5.1) it suffices to maximize

$$-\int_0^1 I(g(x))\,dx$$

over the set of functions $g \in \mathcal{G}_{e,s}$ of the form

$$g(x) = \begin{cases} y_1, & x \in A \\ y_2, & x \notin A \end{cases} \tag{5.11}$$

where $|A| = \lambda \in [0,1]$ and $0 \le y_1 \le y_2 \le 1$. Observe that for such g,

$$\begin{aligned} -\int_0^1 I(g(x))\,dx &= -\lambda I(y_1) - (1-\lambda) I(y_2), \\ \int_0^1 g(x)\,dx &= \lambda y_1 + (1-\lambda) y_2, \\ \int_0^1 g(x)^p\,dx &= \lambda y_1^p + (1-\lambda) y_2^p. \end{aligned}$$

It is therefore enough to maximize

$$-\lambda I(y_1) - (1-\lambda) I(y_2) \tag{5.12}$$

subject to the constraints

$$\begin{aligned} \lambda y_1 + (1-\lambda) y_2 &= e, \\ \lambda y_1^p + (1-\lambda) y_2^p &= s. \end{aligned} \tag{5.13}$$

In case (a), we must have $y_1 = 0$, $y_2 = 1$ and $1-\lambda = e = s$. In case (c), we must take $y_1 = y_2 = e$. This establishes formula (3.2) for $(e,s) \in \partial D$. It remains to consider (b). In this case $(e,s) \in D$ we have seen that $0 < y_1 \le y_2 < 1$. Moreover if $y_1 = y_2$ or $\lambda \in \{0,1\}$ then (5.13) cannot be satisfied, so $y_1 < y_2$ and $\lambda \in (0,1)$. Introducing Lagrange multipliers β_1, β_2, we see that

$$-\lambda I'(y_1) + \lambda\beta_1 + \lambda p\beta_2 y_1^{p-1} = 0,$$

$$-(1-\lambda)I'(y_2)+(1-\lambda)\beta_1+(1-\lambda)p\beta_2 y_2^{p-1}=0,$$
$$-[I(y_1)-I(y_2)]+\beta_1(y_1-y_2)+\beta_2(y_1^p-y_2^p)=0.$$

Thus,

$$\ell'(y_1)=\ell'(y_2)=0, \qquad \ell(y_1)=\ell(y_2).$$

This means that $y_1=x_1<x_2=y_2$ are global maximizers of ℓ on the phase transition curve, away from the critical point.

Uniqueness of the optimizer (geometric proof). We must prove uniqueness for $(e,s)\in D$. Consider the class of lines

$$\lambda(x_1,x_1^p)+(1-\lambda)(x_2,x_2^p), \quad 0<\lambda<1, \tag{5.14}$$

where $0<x_1<x_2<1$ are global maximizers of ℓ on the phase transition curve, away from the critical point. Since x_1 (resp. x_2) is strictly increasing (resp. strictly decreasing) in β_1 with $x_1\le x_2$, no two distinct lines of this class can intersect. Continuity of x_1, x_2 in β_1 along with (3.1) show that the union of all the lines equals D. Thus, given $(e,s)\in D$, there is a unique optimizer g of the form (5.11), obtained by locating the unique line from (5.14) which contains the point (e,s) and choosing λ satisfying the constraint (5.13). Solving for λ in the first constraint, we get $\lambda=(x_2-e)/(x_2-x_1)$. Putting this into the second constraint gives

$$\frac{x_2-e}{x_2-x_1}x_1^p+\frac{e-x_1}{x_2-x_1}x_2^p=s.$$

This establishes formula (3.2) for $(e,s)\in D$.

Uniqueness of the optimizer (algebraic proof). Fix $(e,s)\in D$. From (5.13),

$$\lambda=\frac{e-x_2}{x_2-x_1}=\frac{s-x_2^p}{x_2^p-x_1^p}$$

and so

$$e-x_2=-q'(\beta_1)(s-x_2^p). \tag{5.15}$$

We must show that (5.15) has a unique solution. Define

$$F(\beta_1):=e-x_2+q'(\beta_1)(s-x_2^p).$$

Note that

$$\lim_{\beta_1\to-\infty}F(\beta_1)=e-1-(s-1)=e-s\ge 0,$$

and

$$\begin{aligned}\lim_{\beta_1\to\beta_1^c}F(\beta_1)&=e-\frac{p-1}{p}-\frac{p^{p-2}}{(p-1)^{p-1}}\left(s-\left(\frac{p-1}{p}\right)^p\right)\\&=e-\left(\frac{p-1}{p}\right)^2-s\frac{p^{p-2}}{(p-1)^{p-1}}\\&\le e-\left(\frac{p-1}{p}\right)^2-e^p\frac{p^{p-2}}{(p-1)^{p-1}}.\end{aligned}$$

Also define

$$G(x) := x - \left(\frac{p-1}{p}\right)^2 - x^p \frac{p^{p-2}}{(p-1)^{p-1}}, \quad 0 \le x \le 1.$$

Then $G(0) < 0$ and $G(1) < 0$. Moreover,

$$G'(x) = 1 - x^{p-1}\left(\frac{p}{p-1}\right)^{p-1}$$

is positive if $0 < x < \frac{p-1}{p}$ and it is negative if $\frac{p-1}{p} < x < 1$. Finally,

$$G\left(\frac{p-1}{p}\right) = \frac{p-1}{p} - \frac{(p-1)^2}{p^2} - \frac{p-1}{p^2} = 0.$$

Therefore,

$$\lim_{\beta_1 \to \beta_1^c} F(\beta_1) \le e - \left(\frac{p-1}{p}\right)^2 - e^p \frac{p^{p-2}}{(p-1)^{p-1}} \le 0.$$

Now by the Intermediate Value Theorem, there exists $\beta_1^* \le \beta_1^c$ such that $F(\beta_1^*) = 0$.

Next we prove β_1^* is unique. Observe that

$$F'(\beta_1) = -\left[1 + q'(\beta_1) p x_2^{p-1}\right] \frac{\partial x_2}{\partial \beta_1} + q''(\beta_1)(s - x_2^p).$$

Since $0 \le \lambda \le 1$, we have $s - x_2^p \le 0$. We also know that $q''(\beta_1) \ge 0$ and $\partial x_2/\partial \beta_1 > 0$. Moreover, since

$$q'(\beta_1) = -\frac{x_1 - x_2}{x_1^p - x_2^p} < 0$$

and $0 < x_1 < \frac{p-1}{p} < x_2 < 1$, we have

$$1 + q'(\beta_1) p x_2^{p-1} = 1 - \frac{x_1 - x_2}{x_1^p - x_2^p} p x_2^{p-1} < 0.$$

Hence, we conclude that $\partial F/\partial \beta_1 < 0$ for $\beta_1 < \beta_1^c$ and therefore β_1^* is unique.

Regularity of $\psi(e, s)$. We turn now to the claimed regularity of $\psi(e, s)$, starting with analyticity.

Fix $e \in (0, 1)$. Then each $s \in (e^p, e)$ satisfies, for some $x_1 < x_2$,

$$\frac{x_2 - e}{x_2 - x_1} x_1^p + \frac{e - x_1}{x_2 - x_1} x_2^p = s. \tag{5.16}$$

We claim that (5.16) defines β_1 implicitly as an analytic function of s for $s \in (e^p, e)$. By differentiating the left hand side of (5.16) with respect to β_1, we find the expression

$$\left(\frac{x_1 - e}{x_2 - x_1}\frac{\partial x_2}{\partial \beta_1} + \frac{e - x_2}{x_2 - x_1}\frac{\partial x_1}{\partial \beta_1}\right)\frac{x_2^p - x_1^p}{x_2 - x_1} - \frac{x_1 - e}{x_2 - x_1}\frac{\partial x_2}{\partial \beta_1} p x_2^{p-1} - \frac{e - x_2}{x_2 - x_1}\frac{\partial x_1}{\partial \beta_1} p x_1^{p-1}.$$

By the mean value theorem, there is $x_1 < y < x_2$ such that this expression becomes

$$\left[\frac{x_1 - e}{x_2 - x_1}\frac{\partial x_2}{\partial \beta_1}(p y^{p-1} - p x_2^{p-1})\right] + \left[\frac{e - x_2}{x_2 - x_1}\frac{\partial x_1}{\partial \beta_1}(p y^{p-1} - p x_1^{p-1})\right].$$

Since $\partial x_1/\partial\beta_1 > 0$ and $\partial x_2/\partial\beta_1 < 0$, each of the terms in brackets is negative, so this expression is nonzero. By the analytic implicit function theorem [16], we conclude that β_1 is an analytic function of s inside D. Similar arguments show that β_1 is an analytic function of e inside D. (By Theorem 1, this means β_2 must also be an analytic function of e and s inside D.) Since x_1 and x_2 are analytic functions of β_1, we see that x_1 and x_2 are analytic functions of e and s inside D. Inspecting (3.2), we conclude that $\psi = \psi(e, s)$ is analytic in D.

Next we show that ψ is continuous on $\bar{D}$. We have already shown that ψ is analytic, hence continuous, in D. It is easy to check that ψ is continuous along each line

$$\lambda(x_1, x_1^p) + (1-\lambda)(x_2, x_2^p), \quad 0 \le \lambda \le 1,$$

and that the restriction of ψ to the lower boundary $\{(e, s) : s = e,\ 0 < e < 1\}$ is continuous. This is enough to conclude that ψ is continuous on $\bar{D} \setminus D^{up}$. Finally for $(e_0, e_0) \in D^{up}$, we have

$$\lim_{(e,s)\to(e_0,e_0)} \psi(e,s) = \lim_{x_1\to 0,\, x_2\to 1} -\frac{x_2-e}{x_2-x_1} I(x_1) - \frac{e-x_1}{x_2-x_1} I(x_2) = -\log 2 = \psi(e_0, e_0).$$

Next we prove continuity of the first order derivatives on $\bar{D} \setminus D^{up}$. It suffices to show that: (i) the first order partial derivatives of ψ are continuous in D; and (ii) the limits of the first order partial derivatives of ψ at the lower boundary $\{(e, s) : s = e^p,\ 0 < e < 1\}$ exist. With

$$\lambda = \frac{x_2 - e}{x_2 - x_1}$$

and using (3.2), we see that for $(e, s) \in D$,

$$\frac{\partial\psi}{\partial s} = \frac{\partial\lambda}{\partial s}(I(x_1) - I(x_2)) + \lambda I'(x_1)\frac{\partial x_1}{\partial s} + (1-\lambda)I'(x_2)\frac{\partial x_2}{\partial s}. \tag{5.17}$$

Using the fact that $I(x) = \beta_1 x + \beta_2 x^p - \ell(x) + \log 2$,

$$\begin{aligned}\frac{\partial\psi}{\partial s} &= \frac{\partial\lambda}{\partial s}\left(\beta_1 x_1 + \beta_2 x_1^p - \beta_1 x_2 - \beta_2 x_2^p\right)\\ &\quad + \lambda\frac{\partial x_1}{\partial s}\left(\beta_1 + p\beta_2 x_1^{p-1}\right) + (1-\lambda)\frac{\partial x_2}{\partial s}\left(\beta_1 + p\beta_2 x_2^{p-1}\right).\end{aligned}$$

It is straightforward to compute that

$$\frac{\partial\lambda}{\partial s} = \lambda\frac{\partial x_1}{\partial s}\frac{1}{x_2-x_1} + (1-\lambda)\frac{\partial x_2}{\partial s}\frac{1}{x_2-x_1}. \tag{5.18}$$

Thus,

$$\begin{aligned}\frac{\partial\psi}{\partial s} &= -\left(\lambda\frac{\partial x_1}{\partial s} + (1-\lambda)\frac{\partial x_2}{\partial s}\right)\left(\beta_1 + \beta_2\frac{x_2^p - x_1^p}{x_2 - x_1}\right)\\ &\quad + \lambda\frac{\partial x_1}{\partial s}\left(\beta_1 + p\beta_2 x_1^{p-1}\right) + (1-\lambda)\frac{\partial x_2}{\partial s}\left(\beta_1 + p\beta_2 x_2^{p-1}\right)\\ &= \beta_2\left[\lambda\frac{\partial x_1}{\partial s}\left(p x_1^{p-1} - \frac{x_2^p - x_1^p}{x_2 - x_1}\right) + (1-\lambda)\frac{\partial x_2}{\partial s}\left(p x_2^{p-1} - \frac{x_2^p - x_1^p}{x_2 - x_1}\right)\right].\end{aligned}$$

Differentiating

$$\lambda x_1^p + (1-\lambda)x_2^p = s$$

with respect to s and using (5.18), we find that

$$1 = \lambda \frac{\partial x_1}{\partial s}\left(p x_1^{p-1} - \frac{x_2^p - x_1^p}{x_2 - x_1}\right) + (1-\lambda)\frac{\partial x_2}{\partial s}\left(p x_2^{p-1} - \frac{x_2^p - x_1^p}{x_2 - x_1}\right),$$

and so

$$\frac{\partial \psi}{\partial s} = \beta_2.$$

We have already seen that β_2 is an analytic, hence continuous, function of (e, s) inside D. Since

$$q^{-1}(\beta_2) + p\beta_2 x_i^{p-1} - I'(x_i) = 0, \quad i = 1, 2,$$

we see that for $e_0 \in (0, 1)$,

$$\begin{aligned}\lim_{(e,s)\to(e_0,e_0^p)} \beta_2 &= \lim_{x_i \to e_0} \frac{I'(x_i) - q^{-1}(\beta_2)}{p x_i^{p-1}} \\ &= \frac{I'(e_0) - \beta_1^c}{p e_0^{p-1}}\end{aligned}$$

where $i = 1$ if $e_0 \in (0, (p-1)/p)$ and otherwise $i = 2$. Also, for any $e_0 \in [0, 1]$,

$$\lim_{(e,s)\to(e_0,e_0)} \beta_2 = \lim_{x_1 \to 0} \beta_2 = \infty.$$

Proofs of regularity for $\partial\psi/\partial e$ are similar so we omit them.

Explicit formula when $p = 2$. When $p = 2$, we can explicitly solve

$$\begin{aligned}&\lambda x_1 + (1-\lambda) x_2 = e \\ &\lambda x_1^2 + (1-\lambda) x_2^2 = s\end{aligned}$$

to obtain

$$x_1 = \frac{1 - \sqrt{1 - 4(e-s)}}{2} \qquad x_2 = \frac{1 + \sqrt{1 - 4(e-s)}}{2}$$

and

$$\lambda = \frac{x_2 - e}{x_2 - x_1} = \frac{\sqrt{1 - 4(e-s)} + 1 - 2e}{2\sqrt{1 - 4(e-s)}}.$$

This yields

$$\begin{aligned}\psi(e, s) &= -\lambda I(x_1) - (1-\lambda) I(x_2) \\ &= -\left(\frac{1}{2} + \frac{1 - 2e}{2\sqrt{1 - 4(e-s)}}\right) I\left(\frac{1}{2} - \frac{1}{2}\sqrt{1 - 4(e-s)}\right) \\ &\quad - \left(\frac{1}{2} - \frac{1 - 2e}{2\sqrt{1 - 4(e-s)}}\right) I\left(\frac{1}{2} + \frac{1}{2}\sqrt{1 - 4(e-s)}\right) \\ &= -I\left(\frac{1}{2} + \frac{1}{2}\sqrt{1 - 4(e-s)}\right),\end{aligned}$$

where the last line uses the fact that I is symmetric around $1/2$. □

Proof of Theorem 4. *Proof of part* (i). By Theorem 7,

$$\psi(e,\beta_2) = \sup_{\mathcal{G}_{e,\cdot}} \left[\beta_2 \int_0^1 g(x)^p\, dx - \int_0^1 I(g(x))\, dx \right], \tag{5.19}$$

where we recall $\mathcal{G}_{e,\cdot}$ is the set of measurable functions $g : [0,1] \to [0,1]$ satisfying

$$\int_0^1 g(x)\, dx = e. \tag{5.20}$$

Arguments similar to those in the proof of Theorem 3 show that optimizers of (5.19) exist and have the form

$$g(x) = \begin{cases} y_1, & x \in A \\ y_2, & x \notin A \end{cases} \tag{5.21}$$

where $|A| = \lambda \in [0,1]$ and $0 \le y_1 \le y_2 \le 1$. For such functions g, we have

$$\beta_2 \int_0^1 g(x)^p\, dx - \int_0^1 I(g(x))\, dx = \beta_2 \left[\lambda y_1^p + (1-\lambda) y_2^p\right] - \left[\lambda I(y_1) + (1-\lambda) I(y_2)\right],$$
$$\int_0^1 g(x)\, dx = \lambda y_1 + (1-\lambda) y_2.$$

It is therefore enough to maximize

$$\beta_2 \left[\lambda y_1^p + (1-\lambda) y_2^p\right] - \lambda I(y_1) - (1-\lambda) I(y_2) \tag{5.22}$$

subject to the constraint

$$\lambda y_1 + (1-\lambda) y_2 = e. \tag{5.23}$$

If $\lambda \in \{0,1\}$ or $e \in \{0,1\}$, then (5.23) shows that $g(x) \equiv e$. So assume that $e, \lambda \in (0,1)$. It is easy to see in this case that $0 < y_1 \le y_2 < 1$. Moreover if $y_1 = y_2$ then again $g(x) \equiv e$, so assume $y_1 < y_2$. Introducing the Lagrange multiplier β_1, we find that

$$\begin{aligned} &-\lambda I'(y_1) + \lambda\beta_1 + \lambda p \beta_2 y_1^{p-1} = 0, \\ &-(1-\lambda) I'(y_2) + (1-\lambda)\beta_1 + (1-\lambda) p \beta_2 y_2^{p-1} = 0, \\ &-[I(y_1) - I(y_2)] + \beta_1 (y_1 - y_2) + \beta_2 (y_1^p - y_2^p) = 0, \end{aligned}$$

and so

$$\ell'(y_1) = \ell'(y_2) = 0, \qquad \ell(y_1) = \ell(y_2). \tag{5.24}$$

If $\beta_2 \le \beta_2^c$, then there are no solutions to (5.24), while if $\beta_2 > \beta_2^c$, solutions occur precisely when $y_1 = x_1 < x_2 = y_2$ are the global maximizers of ℓ at the point $(q^{-1}(\beta_2), \beta_2)$ along q. In the latter case, (5.23) implies

$$x_1 \le e \le x_2.$$

If $e = x_1$ or $e = x_2$, then $g(x) \equiv e$. Since for $(e, \beta_2) \in U_e^c$ we have (by definition) $e \notin (x_1, x_2)$, we have established that (3.3) is valid in U_e^c.

Next, for $(e, \beta_2) \in U_e$ define

$$\lambda = \frac{x_2 - e}{x_2 - x_1}$$

and

$$H(e) = \beta_2 e^p - I(e) - \left(\beta_2 \left[\lambda x_1^p + (1-\lambda)x_2^p\right] - [\lambda I(x_1) + (1-\lambda)I(x_2)]\right).$$

To establish that (3.3) is valid in U_e, it suffices to show that $H(e) < 0$ for all $e \in (x_1, x_2)$. It is easy to check that $H(x_1) = H(x_2) = 0$ and

$$\begin{aligned} H'(e) &= p\beta_2 e^{p-1} - I'(e) - \frac{1}{x_2 - x_1}\left(\beta_2(x_2^p - x_1^p) - (I(x_2) - I(x_1))\right) \\ &= \ell'(e) - \beta_1 - \frac{1}{x_2 - x_1}\left(\ell(x_2) - \ell(x_1) + \beta_1(x_2 - x_1)\right) \\ &= \ell'(e). \end{aligned}$$

Thus, $H'(x_1) = H'(x_2) = 0$. Finally,

$$H''(e) = p(p-1)\beta_2 e^{p-2} - I''(e) = \ell''(e).$$

From the proof of Proposition 11 in [1], we know $\ell''(x_1) < 0$, $\ell''(x_2) < 0$, and moreover there exists $x_1 < u_1 < u_2 < x_2$ such that $\ell''(e) < 0$ for $e \in (x_1, u_1) \cup (u_2, x_2)$ while $\ell''(e) > 0$ for $e \in (u_1, u_2)$. The result follows.

We note that the star density can be computed easily as

$$s(e, \beta_2) := \frac{\partial}{\partial \beta_2}\psi(e, \beta_2) = \begin{cases} e^p, & (e, \beta_2) \in U_e^c, \\ \dfrac{e - x_2}{x_1 - x_2}x_1^p + \dfrac{x_1 - e}{x_1 - x_2}x_2^p, & (e, \beta_2) \in U_e. \end{cases}$$

It is clear that the star density $s(e, \beta_2)$ is continuous everywhere.

Regularity of $\psi(e, \beta_2)$. From the formula it is easy to see that $\psi(e, \beta_2)$ is analytic away from ∂U_e. Write $\psi = \psi(e, \beta_2)$ and fix $\beta_2 > \beta_2^c$. In the interior of U_e^c,

$$\frac{\partial \psi}{\partial e} = p\beta_2 e^{p-1} - I'(e),$$

while the interior of U_e,

$$\begin{aligned} \frac{\partial \psi}{\partial e} &= \frac{\beta_2 x_1^p}{x_1 - x_2} - \frac{\beta_2 x_2^p}{x_1 - x_2} - \frac{I(x_1)}{x_1 - x_2} + \frac{I(x_2)}{x_1 - x_2} \\ &= \frac{\ell(x_1) - \ell(x_2)}{x_1 - x_2} - \beta_1 = -\beta_1. \end{aligned}$$

Observe that

$$\lim_{(e,\beta_2)\in U_e^c,\, e\to x_1} \frac{\partial \psi}{\partial e} = p\beta_2 x_1^{p-1} - I'(x_1) = \ell'(x_1) - \beta_1 = -\beta_1.$$

So $\partial\psi/\partial e$ is continuous across ∂U_e when $e < e^c$. Note that

$$\frac{\partial^j \psi}{\partial e^j} = 0, \quad (e, \beta_2) \in U_e, \; j \geq 2. \tag{5.25}$$

Proposition 11 of [1] gives

$$\lim_{(e,\beta_2)\in U_e^c,\, e\to x_1} \partial^2\psi(e, \beta_2)/\partial e^2 = \ell''(x_1) < 0.$$

Thus, $\partial^2\psi(e,\beta_2)/\partial e^2$ has a jump discontinuity across ∂U_e when $e < e^c$. Analogous statements hold true for $e > e^c$. Now consider the situation at the point (e^c, β_2^c). Proposition 11 of [1] gives

$$\lim_{(e,\beta_2)\in U_e^c,\, e\to e^c} \frac{\partial^4\psi}{\partial e^4} = \ell^{(4)}(e^c) < 0.$$

From (5.25) we see that $\partial^4\phi/\partial e^4$ is discontinuous at (e^c, β_2^c).

Proof of part (ii). By Theorem 7,

$$\psi(\beta_1, s) = \sup_{\mathcal{G}_{e,\cdot}} \left[\beta_1 \int_0^1 g(x)\,dx - \int_0^1 I(g(x))\,dx\right] \tag{5.26}$$

where we recall $\mathcal{G}_{\cdot,s}$ is the set of measurable functions $g : [0, 1] \to [0, 1]$ satisfying

$$\int_0^1 g(x)^p\,dx = s. \tag{5.27}$$

Arguments analogous to those in the proof of part (i) show that $\psi(\beta_1, s)$ has the formula (3.4).

We note that the limiting edge density has the formula

$$e(\beta_1, s) := \frac{\partial}{\partial\beta_1}\psi(\beta_1, s) = \begin{cases} s^{\frac{1}{p}} & (\beta_1, s) \in U_s^c, \\ \dfrac{s - x_2^p}{x_1^p - x_2^p}x_1 + \dfrac{x_1^p - s}{x_1^p - x_2^p}x_2 & (\beta_1, s) \in U_s. \end{cases}$$

It is easy to see that $e(\beta_1, s)$ is continuous everywhere.

Regularity of $\psi(\beta_1, s)$. From the formula it is easy to see that $\psi(\beta_1, s)$ is analytic away from ∂U_s. Write $\psi = \psi(\beta_1, s)$ and fix $\beta_1 < \beta_1^c$. In the interior of U_s^c,

$$\frac{\partial\psi}{\partial s} = \frac{1}{p}\beta_1 s^{\frac{1}{p}-1} - \frac{1}{p}s^{\frac{1}{p}-1}I'\left(s^{\frac{1}{p}}\right),$$

and in the interior of U_s,

$$\begin{aligned}\frac{\partial\psi}{\partial s} &= \frac{\beta_1 x_1}{x_1^p - x_2^p} - \frac{\beta_1 x_2}{x_1^p - x_2^p} - \frac{I(x_1)}{x_1^p - x_2^p} + \frac{I(x_2)}{x_1^p - x_2^p} \\ &= \frac{\ell(x_1) - \ell(x_2)}{x_1^p - x_2^p} - \beta_2 = -\beta_2.\end{aligned}$$

Note that

$$\begin{aligned}\lim_{(\beta_1,s)\in U_s^c,\, s\to x_1^p} \frac{\partial\psi}{\partial s} &= \frac{1}{p}\beta_1 x_1^{1-p} - \frac{1}{p}x_1^{1-p}I'(x_1) \\ &= \frac{x_1^{1-p}}{p}\left(\ell'(x_1) - p\beta_2 x_1^{p-1}\right) = -\beta_2.\end{aligned}$$

So $\partial\psi/\partial s$ is continuous across ∂U_s when $s < s^c$. Note that

$$\frac{\partial^j\psi}{\partial s^j} = 0, \quad (\beta_1, s) \in U_s,\ j \geq 2. \tag{5.28}$$

In the computations below, let $(\beta_1, s) \in U_s^c$ and $t = s^{\frac{1}{p}}$. Note that

$$\frac{\partial\psi}{\partial s} = [\beta_1 - I'(t)]\frac{\partial t}{\partial s} = [\ell'(t) - \beta_2 p t^{p-1}]\frac{\partial t}{\partial s},$$

and

$$\frac{\partial^2\psi}{\partial s^2} = [\ell'(t) - \beta_2 p t^{p-1}]\frac{\partial^2 t}{\partial s^2} + \left[\ell''(t) - \beta_2 p(p-1)t^{p-2}\right]\left(\frac{\partial t}{\partial s}\right)^2$$
$$= \ell'(t)\frac{\partial^2 t}{\partial s^2} + \ell''(t)\left(\frac{\partial t}{\partial s}\right)^2. \tag{5.29}$$

Using (5.29) and Proposition 11 of [1], we get

$$\lim_{(\beta_1,s)\in U_s^c,\, s\to x_1^p} \frac{\partial^2\psi}{\partial s^2} = \ell''(x_1)\left(\frac{1}{p}x_1^{1-p}\right)^2 < 0.$$

Comparing with (5.28), we see that $\partial^2\psi/\partial s^2$ has a jump discontinuity across ∂U_s for $s < s^c$. Analogous results hold for $s > s^c$. Now consider the situation at the point (β_1^c, s^c). From (5.29),

$$\frac{\partial^3\psi}{\partial s^3} = \ell'(t)\frac{\partial^3 t}{\partial s^3} + \ell'''(t)\left(\frac{\partial t}{\partial s}\right)^3 + 3\ell''(t)\frac{\partial t}{\partial s}\frac{\partial^2 t}{\partial s^2},$$

and

$$\frac{\partial^4\psi}{\partial s^4} = 4\ell''(t)\frac{\partial t}{\partial s}\frac{\partial^3 t}{\partial s^3} + \ell'(t)\frac{\partial^4 t}{\partial s^4} + \ell^{(4)}(t)\left(\frac{\partial t}{\partial s}\right)^4$$
$$+ 6\ell'''(t)\left(\frac{\partial t}{\partial s}\right)^2\frac{\partial^2 t}{\partial s^2} + 3\ell''(t)\left(\frac{\partial^2 t}{\partial s^2}\right)^2.$$

As $s \to s^c$, $t \to e^c$ and since $\ell'(e^c) = \ell''(e^c) = \ell'''(e^c) = 0$, $\ell^{(4)}(e^c) < 0$, we have

$$\lim_{s\to s^c}\frac{\partial^2\psi}{\partial s^2} = \lim_{s\to s^c}\frac{\partial^3\psi}{\partial s^3} = 0,$$

while Proposition 11 of [1] gives

$$\lim_{s\to s^c}\frac{\partial^4\psi}{\partial s^4} = \lim_{s\to s^c}\ell^{(4)}(t)\left(\frac{\partial t}{\partial s}\right)^4 = \ell^{(4)}(e^c)\left(\frac{(e^c)^{1-p}}{p}\right)^4 < 0. \quad \square$$

The next result concerns the curve $\beta_2 = q(\beta_1)$ and the shapes of U_e and U_s.

Proposition 8. (i) *The curve $\beta_2 = q(\beta_1)$ is analytic in $\beta_1 < \beta_1^c$.*
(ii) *For any $\beta_1 < \beta_1^c$, $\frac{\partial x_1}{\partial \beta_1} > 0$ and $\frac{\partial x_2}{\partial \beta_1} < 0$. Moreover,*

$$\lim_{\beta_1\to\beta_1^c}\frac{\partial x_1}{\partial\beta_1} = +\infty, \qquad \lim_{\beta_1\to-\infty}\frac{\partial x_1}{\partial\beta_1} = 0,$$
$$\lim_{\beta_1\to\beta_1^c}\frac{\partial x_2}{\partial\beta_1} = -\infty, \qquad \lim_{\beta_1\to-\infty}\frac{\partial x_2}{\partial\beta_1} = 0.$$

For any $\beta_2 > \beta_2^c$, $\frac{\partial x_1}{\partial \beta_2} < 0$ and $\frac{\partial x_2}{\partial \beta_2} > 0$. Moreover,

$$\lim_{\beta_2\to\beta_2^c}\frac{\partial x_1}{\partial\beta_2} = -\infty, \qquad \lim_{\beta_2\to+\infty}\frac{\partial x_1}{\partial\beta_2} = 0,$$
$$\lim_{\beta_2\to\beta_2^c}\frac{\partial x_2}{\partial\beta_2} = +\infty, \qquad \lim_{\beta_2\to+\infty}\frac{\partial x_2}{\partial\beta_2} = 0.$$

Proof. First we show that q is analytic. There is an open V-shaped set containing the phase transition curve $\beta_2 = q(\beta_1)$ except the critical point (β_1^c, β_2^c), inside which ℓ has exactly two local maximizers, $y_1 < y_2$. (See [26,1].) It can be seen from the proof of Proposition 11 of [1] that $\ell''(y_1) < 0$ and $\ell''(y_2) < 0$ inside the V-shaped region. The analytic implicit function theorem [16] then shows that y_1 and y_2 are analytic functions of β_1 and β_2 inside this region. Note that q is defined implicitly by the equation

$$\beta_1 y_1 + \beta_2 y_1^p - I(y_1) - (\beta_1 y_2 + \beta_2 y_2^p - I(y_2)) = 0.$$

Differentiating the left hand side of this equation w.r.t. β_2 gives

$$\begin{aligned}&\beta_1 \frac{\partial y_1}{\partial \beta_2} + y_1^p + \left(p\beta_2 y_1^{p-1} - I'(y_1)\right)\frac{\partial y_1}{\partial \beta_2} - \left[\beta_1 \frac{\partial y_2}{\partial \beta_2} + y_2^p + \left(p\beta_2 y_2^{p-1} - I'(y_2)\right)\frac{\partial y_2}{\partial \beta_2}\right]\\ &\quad = y_1^p - y_2^p < 0.\end{aligned}$$

Another application of the analytic implicit function theorem implies $\beta_2 = q(\beta_1)$ is analytic for $\beta_1 < \beta^c$.

Now we turn to the statements involving x_1 and x_2. Along $\beta_2 = q(\beta_1)$, we have

$$\beta_1 + pq(\beta_1)x_1^{p-1} - \log\left(\frac{x_1}{1-x_1}\right) = 0.$$

Differentiating with respect to β_1, we get

$$1 + pq'(\beta_1)x_1^{p-1} + \left[p(p-1)q(\beta_1)x_1^{p-2} - \frac{1}{x_1(1-x_1)}\right]\frac{\partial x_1}{\partial \beta_1} = 0.$$

Therefore,

$$\frac{\partial x_1}{\partial \beta_1} = \frac{1 + pq'(\beta_1)x_1^{p-1}}{\frac{1}{x_1(1-x_1)} - p(p-1)q(\beta_1)x_1^{p-2}} = \frac{1 - \frac{x_1 - x_2}{x_1^p - x_2^p} p x_1^{p-1}}{-\ell''(x_1)} > 0.$$

Since $\lim_{\beta_1 \to \beta_1^c}[1 + pq'(\beta_1)x_1^{p-1}] = \lim_{\beta_1 \to \beta_1^c}[-\ell''(x_1)] = 0$, by L'Hôpital's rule,

$$\lim_{\beta_1 \to \beta_1^c} \frac{\partial x_1}{\partial \beta_1} = \lim_{\beta_1 \to \beta_1^c} \frac{pq''(\beta_1)x_1^{p-1} + p(p-1)q'(\beta_1)x_1^{p-2}\frac{\partial x_1}{\partial \beta_1}}{-p(p-1)q'(\beta_1)x_1^{p-2} - \ell'''(x_1)\frac{\partial x_1}{\partial \beta_1}},$$

which implies that

$$\lim_{\beta_1 \to \beta_1^c} \frac{\partial x_1}{\partial \beta_1} = +\infty.$$

Since $x_1 \to 0$ as $\beta_1 \to \beta_1^c$, it is easy to see that

$$\lim_{\beta_1 \to -\infty} \frac{\partial x_1}{\partial \beta_1} = 0.$$

The results for x_2 can be proved using the similar methods. Finally, notice that

$$\frac{\partial x_i}{\partial \beta_2} = \frac{\partial x_i}{\partial \beta_1}\frac{\partial q^{-1}(\beta_2)}{\partial \beta_2} = \frac{\partial x_i}{\partial \beta_1}\frac{1}{q'(\beta_1)}, \qquad i = 1, 2.$$

Therefore the results involving β_2 also hold. $\square$

Acknowledgments

David Aristoff was supported in part by AFOSR Award FA9550-12-1-0187. Lingjiong Zhu is grateful to his colleague Tzu-Wei Yang for helpful discussions. The authors are grateful for useful comments from the anonymous referees and associate editor.

References

[1] D. Aristoff, L. Zhu, On the phase transition curve in a directed exponential random graph model, 2014. Preprint Available at: arXiv:1404.6514.
[2] J. Besag, Statistical analysis of non-lattice data, J. R. Stat. Soc. Ser. D Stat. 24 (1975) 179–195.
[3] S. Chatterjee, P. Diaconis, Estimating and understanding exponential random graph models, Ann. Statist. 41 (2013) 2428–2461.
[4] S. Chatterjee, S.R.S. Varadhan, The large deviation principle for the Erdős–Rényi random graph, European J. Combin. 32 (2011) 1000–1017.
[5] F. Clarke, Functional Analysis, Calculus of Variations and Control, Springer, London, 2013.
[6] A. Dembo, O. Zeitouni, Large Deviations Techniques and Applications, second ed., Springer, New York, 1998.
[7] S.E. Fienberg, Introduction to papers on the modeling and analysis of network data, Ann. Appl. Stat. 4 (2010) 1–4.
[8] S.E. Fienberg, Introduction to papers on the modeling and analysis of network data—II, Ann. Appl. Stat. 4 (2010) 533–534.
[9] O. Frank, D. Strauss, Markov graphs, J. Amer. Statist. Assoc. 81 (1986) 832–842.
[10] G. Gallavotti, Statistical Mechanics: A Short Treatise, Springer-Verlag, Berlin, 1999.
[11] O. Häggström, J. Jonasson, Phase transition in the random triangle model, J. Appl. Probab. 36 (1999) 1101–1115.
[12] P. Holland, S. Leinhardt, An exponential family of probability distributions for directed graphs, J. Amer. Statist. Assoc. 76 (1981) 33–50.
[13] S. Janson, Graphons, cut norm and distance, couplings and rearrangements, NYJM Monogr. 4 (2013) 1–76.
[14] R. Kenyon, C. Radin, K. Ren, L. Sadun, Multipodal structure and phase transitions in large constrained graphs, Preprint Available at: arXiv:1405.0599.
[15] R. Kenyon, M. Yin, On the asymptotics of constrained exponential random graphs, Preprint Available at: arXiv:1406.3662.
[16] S.G. Krantz, H.R. Parks, The Implicit Function Theorem: History, Theory, and Applications, Birkhäuser, Boston, MA, 2002.
[17] J.L. Lebowitz, A.E. Mazel, E. Presutti, Rigorous proof of a liquid–vapor phase transition in a continuum particle system, Phys. Rev. Lett. 80 (1998) 4701.
[18] L. Lovász, Very large graphs, Curr. Dev. Math. 2008 (2009) 67–128.
[19] L. Lovász, B. Szegedy, Limits of dense graph seqeunces, J. Combin. Theory Ser. B 96 (2006) 933–957.
[20] M.E.J. Newman, Networks: An Introduction, Oxford University Press, Oxford, 2010.
[21] J. Park, M.E.J. Newman, Solution of the two-star model of a network, Phys. Rev. E 70 (2004) 006146.
[22] J. Park, M.E.J. Newman, Solution for the properties of a clustered network, Phys. Rev. E 72 (2005) 026136.
[23] C. Radin, K. Ren, L. Sadun, The asymptotics of large constrained graphs, J. Phys. A 47 (2014) 175001.
[24] C. Radin, L. Sadun, Phase transitions in a complex network, J. Phys. A 46 (2013) 305002.
[25] C. Radin, L. Sadun, Singularities in the entropy of asymptotically large simple graphs, J. Stat. Phys. 158 (2015) 853–865.
[26] C. Radin, M. Yin, Phase transitions in exponential random graphs, Ann. Appl. Probab. 23 (2013) 2458–2471.
[27] A. Rinaldo, S. Fienberg, Y. Zhou, On the geometry of discrete exponential families with application to exponential random graph models, Electron. J. Stat. 3 (2009) 446–484.
[28] G. Robins, T. Snijders, P. Wang, M. Handcock, P. Pattison, Recent developments in exponential random graph (p^*) models for social networks, Social Networks 29 (2007) 192–215.
[29] T.A.B. Snijders, P. Pattison, G.L. Robins, M. Handcock, New specifications for exponential random graph models, Sociol. Methodol. 36 (2006) 99–153.
[30] D. Strauss, On a general class of models for interaction, SIAM Rev. 28 (1986) 513–527.
[31] S.R.S. Varadhan, Large Deviations and Applications, SIAM, Philadelphia, 1984.
[32] S. Wasserman, K. Faust, Social Network Analysis: Methods and Applications, second ed., in: Structural Analysis in the Social Sciences, Cambridge Univ. Press, Cambridge, 2010.
[33] M. Yin, Critical phenomena in exponential random graphs, J. Stat. Phys. 153 (2013) 1008–1021.