R. Kenyon et al. (2016) "Bipodal Structure in Oversaturated Random Graphs,"
International Mathematics Research Notices, Vol. 2018, No. 4, pp. 1009–1044
Advance Access Publication December 24, 2016
doi: 10.1093/imrn/rnw261

Bipodal Structure in Oversaturated Random Graphs

Richard Kenyon[1,*], Charles Radin[2], Kui Ren[3], and Lorenzo Sadun[4]

[1]Department of Mathematics, Brown University, Providence, RI 02912, USA; [2]Department of Mathematics, University of Texas, Austin, TX 78712, USA; [3]Department of Mathematics and ICES, University of Texas, Austin, TX 78712, USA; and [4]Department of Mathematics, University of Texas, Austin, TX 78712, USA

[*]*Correspondence to be sent to: e-mail: rkenyon@math.brown.edu*

We study the asymptotics of large simple graphs directly constrained by the limiting subgraph densities of edges and of an arbitrary fixed graph H. We prove that, for all but finitely many values of the edge density, if the density of H is constrained to be slightly higher than that for the corresponding Erdős–Rényi graph, the typical large graph is bipodal with parameters varying analytically with the densities. Asymptotically, the parameters depend only on the degree sequence of H.

1 Introduction

We study the asymptotics of large, simple, labeled graphs directly constrained to have subgraph densities ϵ of edges, and τ of some fixed subgraph H with $\ell \geq 2$ edges. To study the asymptotics we use the graphon formalism of Borgs *et al.* [2, 3], Lovász *et al.* [7–9] and the large deviations theorem of Chatterjee and Varadhan [5], from which one can reduce the analysis to the study of the graphons which maximize the entropy subject to the density constraints [6, 14–16]. See definitions in Section 2.

The *phase space* (parameter space) is the subset of $[0, 1]^2$ consisting of accumulation points of all pairs of densities $\bar{\tau} = (\epsilon, \tau)$ achievable by finite graphs. (See Figure 1

Received September 30, 2015; Revised April 22, 2016; Accepted October 21, 2016

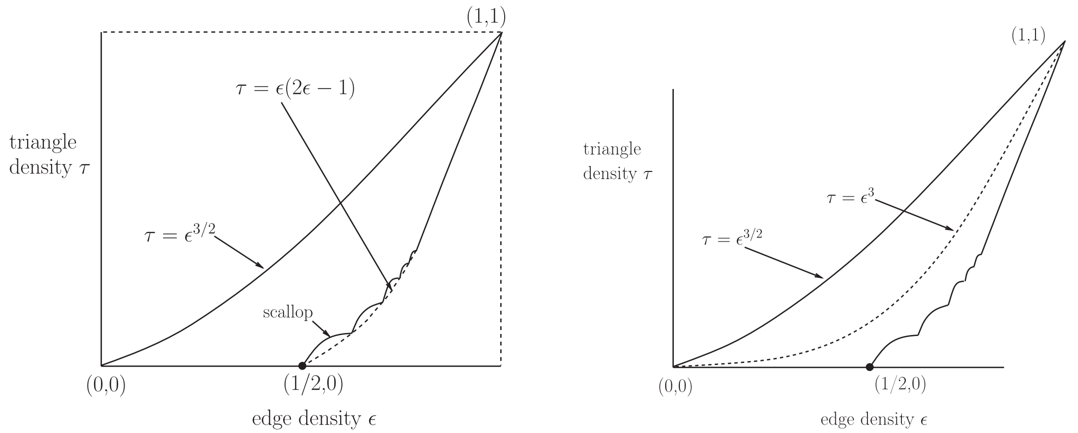

Fig. 1. Boundary of the phase space for the edge/triangle model in solid lines, see [10]. On the right, the ER curve is shown with dashes.

for the model where H is a triangle.) Within the phase space is the "Erdős–Rényi curve" (ER curve) $\{(\epsilon, \tau) \mid \tau = \epsilon^\ell\}$, attained when edges are chosen independently. In this paper, we study the typical behavior of large graphs for $\bar{\tau}$ just above the ER curve. We will show that the qualitative behavior of such graphs is the same for all choices of H and for all but finitely many choices of ϵ depending on H.

To be precise, we show that for fixed H, for ϵ outside a finite set, and for τ close enough to ϵ^ℓ, there is a *unique* entropy-maximizing graphon (up to measure-preserving transformations of the unit interval); furthermore it is bipodal and depends analytically on (ϵ, τ), implying that the entropy is an analytic function of (ϵ, τ). In particular we prove the existence of one or more well-defined phases just above the ER curve.

This is the first proof, as far as we know, of the existence of a phase in any constrained-density graphon model, where by *phase* we mean a (maximal) open set in the phase space at each point of which the entropy has a unique graphon maximizer, which varies analytically with the constraint parameters. (Conjecturally, the phase space is made up of a union of phases and a subset of lower dimension, the latter providing boundaries for the phases [14].) The unique maximizers provide an embedding of each phase into the metric space of reduced graphons. Variation of constraint values in the phase space is therefore mirrored by this embedding into variation in the space of graphons. This has the consequence that smoothness or singularity under variation can be interpreted among the graphons, which are thought of as the emergent states of the large graphs. In contrast, in exponential random graph models (see, e.g., [1, 4, 11, 13]) the parameters, which are associated with graphons by optimization of free energy rather than entropy, play a fundamentally different role; different parameters values can be associated with the *same* optimal graphon. For an extreme example, the whole two-dimensional parameter space for edge/2-star constraints is mapped in this way into the one-dimensional set of Erdös–Rényi graphons [4]. Clearly, smoothness or singularity under variation of parameter values in such models is more naturally interpreted as a feature of the model, as in [4], rather than as a feature of states of large constrained graphs. For further analysis of this see the discussion in the Conclusion in [6].

The study of constrained graphs in the sense we are considering, was initiated by Turán in 1941 [18], addressing in particular the case of edge and triangle constraints. The extremal graph theory of these constraints was recently completed by Razborov *et al.*, in [12, 17], which also contain a good history of this problem. Partial results describing the entropy maximizing graphons in the interior of that phase space were then obtained in [14–16].

For the edge-k-star model, we proved multipodality of all entropy optimizers in [6]. This will be an important tool in this paper. It is also important for the heuristics as it provides a simple interpretation of the emergence of the large scale state of the constrained graphs, through partitioning of nodes. We should also mention that the region *below* the ER curve in the edge/triangle model seems to be more mysterious; no proof of multipodality is known, for example, except on a line segment [16], though there is good simulation evidence of it [14].

A *bipodal graphon* is a function $g : [0,1]^2 \to [0,1]$ of the form:

$$g(x,y) = \begin{cases} p_{11} & x,y < c, \\ p_{12} & x < c < y, \\ p_{12} & y < c < x, \\ p_{22} & x,y > c. \end{cases} \tag{1}$$

Bipodal graphons are generalizations of *bipartite graphons*, in which $p_{11} = p_{22} = 0$. Here c, p_{11}, p_{12}, and p_{22} are constants taking values between 0 and 1. We prove that as $\tau \searrow \epsilon^\ell$, the parameters $c \to 0$, $p_{22} \to \epsilon$, and p_{11} and p_{12} approach the solutions of a problem in single-variable calculus. The inputs to that calculus problem depend only on the degrees of the vertices of H.

We say that a finite graph H is *k-starlike* if all the vertices of H have degree k or 1, where $k > 1$ is a fixed integer. k-starlike graphs include k-stars (where one vertex has degree k and k vertices have degree 1), and the complete graph on $k+1$ vertices. For fixed k, all k-starlike graphs behave essentially the same for our asymptotics. We prove our results first for k-stars, and then apply perturbation theory to show that the differences between different k-starlike graphs are irrelevant, and then prove the general case.

To state our results more precisely, we need some notation. Let

$$S_0(w) = -\frac{1}{2}[w \log w + (1-w)\log(1-w)], \tag{2}$$

and define the *graphon entropy* (or *entropy* for short) of a graphon g to be

$$s(g) = \int_0^1 \int_0^1 S_0(g(x,y))\mathrm{d}x\,\mathrm{d}y. \tag{3}$$

Let

$$\psi_k(\epsilon, \tilde{\epsilon}) = \frac{2[S_0(\tilde{\epsilon}) - S_0(\epsilon) - S_0'(\epsilon)(\tilde{\epsilon}-\epsilon)]}{\tilde{\epsilon}^k - \epsilon^k - k\epsilon^{k-1}(\tilde{\epsilon}-\epsilon)}. \tag{4}$$

$\psi_k(\epsilon, \tilde{\epsilon})$, viewed as a function of $\tilde{\epsilon}$, has a removable singularity at $\tilde{\epsilon} = \epsilon$, which we fill by defining

$$\psi_k(\epsilon, \epsilon) = \frac{2S_0''(\epsilon)}{k(k-1)\epsilon^{k-2}}. \tag{5}$$

For fixed ϵ, let $\zeta_k(\epsilon)$ be the value of $\tilde{\epsilon}$ that maximizes $\psi_k(\epsilon, \tilde{\epsilon})$. (We will prove in Theorem 3.3 below that this maximizer is unique and depends continuously on ϵ.)

Theorem 1.1. Let H be a k-starlike graph with $\ell \geq 2$ edges. Let $\epsilon \in (0,1)$ be any point other than $(k-1)/k$. Then there is a number $\tau_0 > \epsilon^\ell$ (depending on ϵ) such that for all $\tau \in (\epsilon^\ell, \tau_0)$, the entropy-maximizing graphon at (ϵ, τ) is unique (up to measure-preserving transformations of $[0,1]$) and bipodal. Its parameters $(c, p_{11}, p_{12}, p_{22})$ are analytic functions of ϵ and τ on the region $\epsilon \neq (k-1)/k$, $\tau \in (\epsilon^\ell, \tau_0(\epsilon))$. Furthermore, as $\tau \searrow \epsilon^\ell$ we have that $p_{22} \to \epsilon$, $p_{12} \to \zeta_k(\epsilon)$, p_{11} satisfies $S_0'(p_{11}) = 2S_0'(p_{12}) - S_0'(p_{22})$, and $c = O(\tau - \epsilon^\ell)$. □

Theorem 1.1 proves that there is part of a phase just above the ER curve for $\epsilon < (k-1)/k$ and also for $\epsilon > (k-1)/k$; numerical evidence suggests these are in fact parts of a single phase; the only "singular" behavior is the manner in which the graphon approaches the constant graphon associated with the ER curve. We will see in Theorem 1.2 that this behavior is only slightly more complicated for general H than it is for k-starlike H.

When H has vertices with different degrees > 1, the problem resembles that of a formal positive linear combination of k-stars. As in the k-starlike case, we first solve the problem for the linear combination of k-stars and then use perturbation theory to extend the results to arbitrary H.

Theorem 1.2. Let H be an arbitrary graph with ℓ edges with at least one vertex of degree 2 or greater. Then there exists a finite set $B_H \subset (0,1)$ such that if $\epsilon \neq B_H$, then there is a number $\tau_0 > \epsilon^\ell$ (depending on ϵ) such that for all $\tau \in (\epsilon^\ell, \tau_0)$, the entropy-maximizing graphon at (ϵ, τ) is unique (up to measure-preserving transformations of $[0,1]$) and bipodal. Its parameters $(c, p_{11}, p_{12}, p_{22})$ are analytic functions of ϵ and τ on the region $\epsilon \notin B_H$, $\tau \in (\epsilon^\ell, \tau_0(\epsilon))$. Furthermore, as $\tau \searrow \epsilon^\ell$ we have that $p_{22} \to \epsilon$, p_{12} approaches the maximizer of an explicit function whose data depends on ϵ, p_{11} satisfies $S_0'(p_{11}) = 2S_0'(p_{12}) - S_0'(p_{22})$, and $c = O(\tau - \epsilon^\ell)$. □

The key differences between the Theorems 1.1 and 1.2 are:

- For k-starlike graphs, the set B_H of bad values of ϵ consists of a single point, and this point is explicitly known: $\epsilon = (k-1)/k$.
- For k-starlike graphs, the behavior of ζ_k is explicit. It is a continuous and strictly decreasing function of ϵ, and gives an involution of $(0,1)$. (That is, $\zeta_k(\zeta_k(\epsilon)) = \epsilon$.) For $k = 2$ it is given by $\zeta_2(\epsilon) = 1 - \epsilon$. In the general case, the limiting value of p_{12}, and its dependence on ϵ, appear to be much more complicated. We do not know whether this limiting value is always continuous across the bad set B_H.

The organization of this paper is as follows. In Section 2 we review the formalism of graphons and establish basic notation. In Section 3 we establish a number of technical results for k-star models. Using these results, in Section 4 we prove Theorem 1.1 for the case that H is a k-star. In Section 5 we show that just above the ER curve a model with an arbitrary k-starlike H can be approximated by a k-star model. By bounding the error terms, we prove Theorem 1.1 in full generality. In Section 6 we consider formal positive linear combinations of k-stars, and prove a theorem much like Theorem 1.2 for those models. Finally, in Section 7 we show that the model for an arbitrary H can be approximated by a formal linear combination of k-stars, thus completing the proof of Theorem 1.2.

2 Notation and background

We consider a simple graph G (undirected, with no multiple edges or loops) with a vertex set $V(G)$ of labeled vertices. For a subgraph H of G, let $T_H(G)$ be the number of maps from $V(H)$ into $V(G)$ which sends edges to edges. The *density* $\tau_H(G)$ of H in G is then defined to be

$$\tau_H(G) := \frac{|T_H(G)|}{n^{|V(H)|}}, \tag{6}$$

where $n = |V(G)|$. An important special case is where H is a "k-star," a graph with k edges, all with a common vertex, for which we use the notation $\tau_k(G)$. In particular $\tau_1(G)$, which we also denote by $\epsilon(G)$, is the edge density of G.

For $\alpha > 0$ and $\bar{\tau} = (\epsilon, \tau_H)$ define $Z_{\bar{\tau}}^{n,\alpha}$ to be the number of graphs G on n vertices with densities satisfying

$$\epsilon(G) \in (\epsilon - \alpha, \epsilon + \alpha),\ \tau_H(G) \in (\tau_H - \alpha, \tau_H + \alpha). \tag{7}$$

Define the *(constrained) entropy* $s_{\bar{\tau}}$ to be the exponential rate of growth of $Z_{\bar{\tau}}^{n,\alpha}$ as a function of n:

$$s_{\bar{\tau}} = \lim_{\alpha \searrow 0} \lim_{n \to \infty} \frac{\ln(Z_{\bar{\tau}}^{n,\alpha})}{n^2}. \tag{8}$$

The double limit defining the entropy $s_{\bar{\tau}}$ is known to exist [15]. To analyze it we make use of a variational characterization of $s_{\bar{\tau}}$, and for this we need further notation to analyze limits of graphs as $n \to \infty$. (This work was recently developed in [2, 3, 7–9]; see also the recent book [10].) The (symmetric) adjacency matrices of graphs on n vertices are replaced, in this formalism, by symmetric, measurable functions $g : [0,1]^2 \to [0,1]$; the former are recovered by using a partition of $[0,1]$ into n consecutive subintervals. The functions g are called *graphons*.

For a graphon g define the *degree function* $d(x)$ to be $d(x) = \int_0^1 g(x,y)\mathrm{d}y$. The k-star density of g, $\tau_k(g)$, then takes the simple form

$$\tau_k(g) = \int_0^1 d(x)^k \,\mathrm{d}x. \tag{9}$$

For any fixed graph H, the H-density τ_H of g can be similarly expressed as an integral of a product of factors $g(x_i, x_j)$.

The following is Theorem 4.1 in [16]:

Theorem 2.1 (The Variational Principle). For any values $\bar{\tau} = \bar{\tau}(g) := (\epsilon, \tau_H)$ in the phase space we have $s_{\bar{\tau}} = \max[s(g)]$, where the entropy is maximized over all graphons g with $\bar{\tau}(g) = \bar{\tau}$. □

(Instead of using $s(g)$, some authors use the *rate function* $I(g) := -s(g)$, and then minimize I.) The existence of a maximizing graphon $g = g_{\bar\tau}$ for any constraint $\bar\tau(g) = \bar\tau$ was proven in [15], again adapting a proof in [5]. If the densities are that of edges and k-star subgraphs we refer to this maximization problem as a *star model*, though we emphasize that the result applies much more generally [15, 16].

We consider two graphs *equivalent* if they are obtained from one another by relabeling the vertices. For graphons, the analogous operation is applying a measure-preserving map ψ of $[0,1]$ into itself, replacing $g(x,y)$ with $g(\psi(x),\psi(y))$, see [10]. The equivalence classes of graphons under relabeling are called *reduced graphons*, and graphons are equivalent if and only if they have the same subgraph densities for all possible finite subgraphs [10]. In the remaining sections of the paper, whenever we claim that a graphon has a property (e.g., monotonicity in x and y, or uniqueness as an entropy maximizer), the caveat "up to relabeling" is implied.

The graphons which maximize the constrained entropy can tell us what "most" or "typical" large constrained graphs are like: if $g_{\bar\tau}$ is the only reduced graphon maximizing $s(g)$ with $\bar\tau(g) = \bar\tau$, then as the number n of vertices diverges and $\alpha_n \to 0$, exponentially most graphs with densities $\bar\tau_i(G) \in (\tau_i - \alpha_n, \tau_i + \alpha_n)$ will have reduced graphon close to $g_{\bar\tau}$ [15]. This is based on large deviations from [5]. We emphasize that this interpretation requires that the maximizer be unique; this has been difficult to prove in most cases of interest and is an important focus of this work.

A graphon g is called M-podal if there is decomposition of $[0,1]$ into M intervals ("vertex clusters") C_j, $j = 1, 2, \ldots, M$, and $M(M+1)/2$ constants p_{ij} such that $g(x,y) = p_{ij}$ if $(x,y) \in C_i \times C_j$ (and $p_{ji} = p_{ij}$). We denote the length of C_j by c_j.

3 Technical properties of star models

For each star model, all entropy-maximizing graphons are multipodal with a fixed upper bound on the number of clusters, also called the *podality* [6]. (The term multi/bipartite is sometimes used instead of multipodal in the literature.) For any fixed podality M, an M-podal graphon is described by $N = M(M+3)/2$ parameters, namely the values p_{ij} $(1 \le i \le j \le M)$ and the widths c_i $(1 \le i \le M)$ of the clusters. When it does not cause confusion, we will use g to denote the vector

$$(p_{11}, \ldots, p_{1M}, p_{22}, \ldots, p_{2M}, \ldots, \ldots, p_{M-1M-1}, p_{M-1M}, p_{MM}, c_1, \ldots, c_M), \tag{10}$$

which contains all these parameters. The problem of optimizing the graphon then reduces to a finite-dimensional calculus problem. To be precise, let us recall that for an M-podal graphon, we have

$$\epsilon(g) = \sum_{1\le i,j\le M} c_i c_j p_{ij}, \quad \tau_k(g) = \sum_{1\le i\le M} c_i d_i^k, \quad s(g) = \sum_{1\le i,j\le M} c_i c_j S_0(p_{ij}), \tag{11}$$

where $d_i = \sum_{1\le j\le M} c_j p_{ij}$ is the value of the degree function on the ith cluster. The problem of searching for entropy-maximizing graphons with fixed edge density ϵ and k-star density τ_k can now be formulated as

$$\max_{g\in[0,1]^N} s(g), \quad \text{subject to:} \quad \epsilon(g) - \epsilon = 0, \quad \tau_k(g) - \tau = 0, \quad C(g) = 1, \tag{12}$$

where $C(g) = \sum_{1\le j\le M} c_j$.

The following result says that the maximization problem (12) can be solved using the method of Lagrange multipliers. The existence of finite Lagrange multipliers was previously established in [6], treating the space of graphons as a linear space of functions $[0,1]^2 \to [0,1]$, intuitively considering perturbations of graphons localized about points in $[0,1]^2$. For star models we may restrict to M-podal graphons, as noted above, and thus consider perturbations in the relevant parameters p_{ij} and c_j.

Lemma 3.1. Let g be a local maximizer in (12). Then for constraints ϵ, τ off the ER curve, there exist unique $\alpha, \beta, \gamma \in \mathbb{R}$ such that

$$\nabla s(g) - \alpha\nabla\epsilon(g) - \beta\nabla\tau_k(g) - \gamma\nabla C(g) = \mathbf{0}. \tag{13}$$

□

We do not include the proof, which follows easily from that of Lemma 3.5 in [6]. We also note that one can remove the variable c_M and the constraint $C(g) = 1$, eliminating the multiplier γ.

For convenience later, we now write down the exact form of the Euler–Lagrange equation (13). We first verify that

$$\frac{\partial\epsilon}{\partial p_{ij}} = A_{ij}, \qquad \frac{\partial\epsilon}{\partial c_i} = 2\sum_{j=1}^{M} c_j p_{ij} = 2d_i, \tag{14}$$

$$\frac{\partial\tau_k}{\partial p_{ij}} = \frac{k}{2}(d_i^{k-1} + d_j^{k-1})A_{ij}, \qquad \frac{\partial\tau_k}{\partial c_i} = d_i^k + k\sum_{j=1}^{M} c_j d_j^{k-1} p_{ij}, \tag{15}$$

$$\frac{\partial C}{\partial p_{ij}} = 0, \qquad \frac{\partial C}{\partial c_i} = 1, \tag{16}$$

$$\frac{\partial s}{\partial p_{ij}} = S_0'(p_{ij})A_{ij}, \qquad \frac{\partial s}{\partial c_i} = 2\sum_{j=1}^{M} c_j S_0(p_{ij}), \tag{17}$$

where $A_{ij} = 2c_ic_j$ if $i \neq j$ and $A_{ij} = c_i^2$ if $i = j$. We can then write down (13) explicitly as

$$S_0'(p_{ij}) = \alpha + \beta \frac{k}{2}(d_i^{k-1} + d_j^{k-1}), \quad 1 \le i \le j \le M \tag{18}$$

$$2\sum_{j=1} c_j S_0(p_{ij}) = 2\alpha d_i + \beta\big(d_i^k + k\sum_{j=1}^{M} c_j d_j^{k-1} p_{ij}\big) + \gamma, \quad 1 \le i \le M. \tag{19}$$

These Euler–Lagrange equations, together with the constraints,

$$\epsilon(g) - \epsilon = 0, \qquad \tau_k(g) - \tau = 0, \qquad C(g) - 1 = 0, \tag{20}$$

are the optimality conditions for the maximization problem (12). In principle, we can solve this system to find the maximizer g.

Next we consider the significance of the Lagrange multipliers α and β. Suppose that g_0 is the unique entropy maximizer for $\epsilon = \epsilon_0$ and $\tau = \tau_0$. Then any sequence of graphons that maximize entropy for (ϵ, τ) approaching (ϵ_0, τ_0) must approach g_0: this follows from continuity of the entropy on the space of M-podal graphons and the fact that we can perturb g_0 to any nearby (ϵ, τ) by changing some p_{ij}'s (as follows easily from (11)). But if $g = g_0 + \delta g$, then

$$s(g) = s(g_0) + \nabla s(g_0) \cdot \delta g + O(\|\delta g\|^2), \tag{21}$$

where $\|\delta g\|$ denotes the norm of δg as a vector in $\mathbb{R}^N$. However, from (13) we have

$$\begin{aligned} s(g) &= s(g_0) + \alpha \nabla \epsilon(g_0) \cdot \delta g + \beta \nabla \tau(g_0) \cdot \delta g + O(\|\delta g\|^2) \\ &= s(g_0) + \alpha(\epsilon - \epsilon_0) + \beta(\tau - \tau_0) + O(\|\delta g\|^2). \end{aligned} \tag{22}$$

Thus $\partial s_{(\epsilon,\tau)}/\partial \epsilon = \alpha$ and $\partial s_{(\epsilon,\tau)}/\partial \tau = \beta$. If g_0 is not a unique entropy maximizer, then a similar argument shows that we have 1-sided (directional) derivatives:

Lemma 3.2. The function $s_{(\epsilon,\tau)}$ admits directional derivatives in all directions at all points (ϵ, τ) in the interior of the phase space. □

Proof. Suppose that there are multiple entropy-maximizing graphons at a particular (ϵ_0, τ_0). Given a vector $v = (v_\epsilon, v_\tau) \in \mathbb{R}^2$, we wish to compute $s_{(\epsilon_0 + tv_E, \tau_0 + tv_\tau)} - s_{(\epsilon_0, \tau_0)}$ to first order in t for t small and positive. As $t \to 0$, the optimizing graphon g must approach an entropy-maximizing graphon g_0 with $\epsilon(g_0) = \epsilon_0$ and $\tau(g_0) = \tau_0$. But then, by (21), $s(g) - s(g_0) = t(\alpha v_\epsilon + \beta v_\tau) + O(t^2)$, where α and β depend on the choice g_0. Among the choices for g_0, there is one (or more) that maximizes $\alpha v_E + \beta v_\tau$, and our directional derivative is that maximal value of $\alpha v_E + \beta v_\tau$. ■

Existence of directional derivatives implies the fundamental theorem of calculus, so for fixed ϵ we can write

$$s_{(\epsilon,\tau)} = s_{(\epsilon,\epsilon^k)} + \int_{\epsilon^k}^{\tau} \beta(g_{\max}(\epsilon,\tau))\mathrm{d}\tau, \tag{23}$$

where $g_{\max}(\epsilon,\tau)$ is the entropy-maximizing graphon at (ϵ,τ) that maximizes its right derivative (with respect to τ).

Before proving Theorem 1.1 for k-stars, we record some properties of the function $\psi_k(\epsilon,\tilde{\epsilon})$ of (4) and its critical points.

Theorem 3.3. For fixed k and ϵ, there is a unique solution to $\partial\psi_k'(\epsilon,\tilde{\epsilon})/\partial\tilde{\epsilon} = 0$, which we denote $\tilde{\epsilon} = \zeta_k(\epsilon)$. The function ζ_k is strictly decreasing, with nowhere-vanishing derivative and with fixed point at $\epsilon = (k-1)/k$. Furthermore, ζ_k is an involution: $\tilde{\epsilon} = \zeta_k(\epsilon)$ if and only if $\epsilon = \zeta_k(\tilde{\epsilon})$. Moreover, if $\zeta_k(\epsilon) \neq \epsilon$, then $\psi_k(\epsilon,\epsilon) < \psi_k(\epsilon,\zeta_k(\epsilon))$ and $\psi_k(\zeta_k(\epsilon),\zeta_k(\epsilon)) < \psi_k(E,\zeta_k(\epsilon))$. □

Even though the proof is elementary we will need some parts of it later, so we give it in the Appendix.

4 Theorem 1.1 for k-stars

Theorem 4.1. Let H be the k-star graph and suppose that $\epsilon \neq (k-1)/k$. Then there exists a number $\tau_0 > \epsilon^k$ such that for all $\tau \in (\epsilon^k, \tau_0)$, the entropy-optimizing graphon at (ϵ,τ) is unique and bipodal. The parameters $(p_{11}, p_{12}, p_{22}, c)$ are analytic functions of ϵ and τ. As τ approaches ϵ^k from above, $p_{22} \to \epsilon$, $p_{12} \to \zeta_k(\epsilon)$, p_{11} satisfies $S_0'(p_{11}) = 2S_0'(p_{12}) - S_0'(p_{22})$, and $c = O(\tau - \epsilon^k)$. □

Proof. The entropy-maximizing graphon for each (ϵ,τ) is multipodal [6], and the parameters $\{c_j\}$ and $\{p_{ij}\}$ must satisfy the optimality conditions (18) and (19). The first step of the proof is to estimate the terms in the optimality equations to within $o(1)$. This will determine the solutions to within $o(1)$ and demonstrate that our optimizing graphon is close to bipodal of the desired form. The second step, based on a separate argument, will show that the optimizer is exactly bipodal. The third step shows that the optimizer is in fact unique.

In doing our asymptotic analysis, our small parameter is $\Delta\tau := \tau - \epsilon^k$. However, we claim that

$$\Delta\tau \asymp \|\Delta g\|^2 \asymp |\Delta s|, \tag{24}$$

(the notation $A \asymp B$ means $A = O(B)$ and $B = O(A)$) where $\Delta s := s(g) - S_0(\epsilon)$ and $\|\Delta g\|^2$ is the squared L^2 norm of $\Delta g := g - g_0$, where $g_0(x, y) = \epsilon$ (here g denotes the graphon as a function $[0, 1]^2 \to [0, 1]$, not a vector of multipodal parameters). However, $\Delta\tau = O(\|\Delta g\|^2)$ (adapting the argument of [16], Theorem 3.1 to arbitrary graphs), and $\|\Delta g\|^2 = O(|\Delta s|)$ (by equation (16) of [16]). By considering a bipodal graphon with $p_{11} = p_{12} = \zeta_k(\epsilon)$ and p_{22} close to ϵ, we see that $|\Delta s| = O(\Delta\tau)$. This shows (24). In the rest of the proof, unless otherwise specified, by terms such as "close to" and "small" we mean within $o(1)$ as $\Delta\tau \to 0$.

Order the M vertex clusters so that the largest cluster is the last cluster (of length c_M). By subtracting the equation (19) for c_M from the equations for c_j, we eliminate γ from our equations:

$$\begin{aligned} S_0'(p_{ij}) &= \alpha + \frac{k}{2}\beta(d_i^{k-1} + d_j^{k-1}) \\ 2\sum_{j=1}^{M} c_j\left(S_0(p_{ij}) - S_0(p_{Mj})\right) &= 2\alpha(d_i - d_M) + \beta\left(d_i^k - d_M^k + k\sum_{j=1}^{M} c_j d_j^{k-1}(p_{ij} - p_{Mj})\right). \end{aligned} \tag{25}$$

Step 1.

Since

$$\|\Delta g\|^2 = \iint (g(x, y) - \epsilon)^2 \mathrm{d}x\,\mathrm{d}y = \sum_{i,j} c_i c_j (p_{ij} - \epsilon)^2,$$

the ith cluster must either have $d_i = \sum_j c_j p_{ij}$ close to ϵ (i.e., within $o(1)$), or c_i close to zero, or both. We call a cluster Type I if c_i is close to 0 and Type II if d_i is close to ϵ. (If a cluster meets both conditions, we arbitrarily throw it into one camp or the other.) The first equation in (25) implies that, for fixed i, the values of p_{ij} are nearly constant for all j of Type II. Since the c_j's are small for j of Type I, this common value must be close to d_i. Our equations then simplify to

$$\begin{aligned} S_0'(d_i) &= \alpha + \frac{k}{2}\beta(d_i^{k-1} + \epsilon^{k-1}) + o(1), \\ S_0(d_i) - S_0(\epsilon) &= \alpha(d_i - \epsilon) + \beta[d_i^k - \epsilon^k + k\epsilon^{k-1}(d_i - \epsilon)] + o(1). \end{aligned} \tag{26}$$

Since $d_M = \epsilon + o(1)$, the first of those equations applied to d_M implies that

$$\alpha + k\epsilon^{k-1}\beta = S_0'(\epsilon) + o(1). \tag{27}$$

We can thus replace α with $S_0'(\epsilon) - k\epsilon^{k-1}\beta + o(1)$ throughout. This gives the equations:

$$\begin{aligned} 2(S_0'(d_i) - S_0'(\epsilon)) &= k\beta(d_i^{k-1} - \epsilon^{k-1}) + o(1), \\ 2[S_0(d_i) - S_0(\epsilon) - S_0'(\epsilon)(d_i - \epsilon)] &= \beta[d_i^k - \epsilon^k - k\epsilon^{k-1}(d_i - \epsilon)] + o(1). \end{aligned} \tag{28}$$

There are two solutions to these equations. One is simply to have $d_i = \epsilon + o(1)$, in which case both sides of both equations are $o(1)$. Indeed, we already know that there must be clusters with d_i close to ϵ. In looking for solutions with d_i not close to ϵ, the second equation says that $\beta = \psi_k(\epsilon, d_i) + o(1)$.

In this case we can divide the first equation by the second to eliminate β. This gives an equation that is algebraically equivalent to $\partial\psi_k(\epsilon, d_i)/\partial d_i = o(1)$. In other words, d_i must be tending to the unique critical point $\zeta_k(\epsilon)$ of ψ_k, and β must be tending to the critical value. In fact, the critical point is a maximum of ψ_k. Remember that $s_{(\epsilon,\tau)} = s_{(\epsilon,\epsilon^k)} + \int_{\epsilon^k}^{\tau} \beta$ from (23). Since the computation of β is independent of $\Delta\tau$ (to lowest order), we have $s_{(\epsilon,\tau)} - s_{(\epsilon,\epsilon^k)} = \beta\Delta\tau + o(\Delta\tau)$, so maximizing β is tantamount to maximizing s.

Step 2.

We have shown so far that the optimizing graphon is multipodal, with all of the clusters either having d_i close to $\zeta_k(\epsilon)$ or close to ϵ. Furthermore, the clusters with d_i close to $\zeta_k(\epsilon)$ have total size $\sum c_i = o(1)$. We refine our definitions of Types I and II so that all the clusters with d_i close to $\zeta_k(\epsilon)$ are Type I and all the clusters with d_i close to ϵ are Type II. We order the clusters so that the Type I clusters come before Type II, thereby dividing $[0, 1]^2$ into $I \times I$, $I \times II$, $II \times I$, and $II \times II$ quadrants. Since the value of $g(x, y)$ is determined by $d(x)$ and $d(y)$ (and α and β), this means that the optimizing graphon is nearly constant (i.e., with pointwise small fluctuations) on each quadrant.

Let g_b be the bipodal graphon obtained by averaging g over each quadrant. That is, c is the total size of all the Type I clusters, and the parameters p_{11}, p_{12}, and p_{22} are chosen such that $0 = \iint_{I\times I}(g(x, y)-p_{11})\mathrm{d}x\,\mathrm{d}y = \iint_{I\times II}(g(x, y)-p_{12})\mathrm{d}x\,\mathrm{d}y = \iint_{II\times II}(g(x, y)-p_{22})\mathrm{d}x\,\mathrm{d}y$. Let $\Delta g_f = g - g_b$. (The f stands for "further.") We will show that having Δg_f non-zero is an inefficient way to increase τ, that is, $(s(g) - s(g_b))/(\tau(g) - \tau(g_b))$ is less than β.

By the first equation in (25), $S_0'(g(x, y))$ is the sum of a function of x and the same function of y. This means that there is a function $F(x)$ on $[0, 1]$, with $\int_I F(x)\mathrm{d}x =$

$\int_{II} F(x)\mathrm{d}x = 0$, such that on each quadrant

$$S_0'(g(x,y)) = \text{constant} + F(x) + F(y). \tag{29}$$

Furthermore, $F(x)$ is pointwise small (meaning it approaches 0 pointwise at $\tau \to \epsilon^k$), so we can write the Taylor series

$$S_0'(g(x,y)) = S_0'(g_b(x,y) + \Delta g_f(x,y)) = S_0'(g_b(x,y)) + S_0''(g_b(x,y))\Delta g_f(x,y) + O(\Delta g_f(x,y)^2). \tag{30}$$

Since $S_0'(g(x,y))$ is not a linear function of $g(x,y)$, the constant in (29) is not exactly $S_0'(g_b(x,y))$. The correction to $S_0'(g_b(x,y))$ is obtained by integrating higher-order terms in the Taylor series (30) over the quadrant, and so is controlled by the squared L^2 norm of F.

Using (29), on each quadrant we can solve (30) for $\Delta g_f(x,y)$ as

$$\Delta g_f(x,y) = \begin{cases} \frac{F(x)+F(y)}{S_0''(p_{11})} + O(F^2) & \text{On } I \times I \\ \frac{F(x)+F(y)}{S_0''(p_{12})} + O(F^2) & \text{On } I \times II \text{ and } II \times I \\ \frac{F(x)+F(y)}{S_0''(p_{22})} + O(F^2) & \text{On } II \times II, \end{cases} \tag{31}$$

where $O(F^2)$ is shorthand for terms that are bounded by quadratic functions of $F(x)$ and $F(y)$ and a quadratic function of the L^2 norm of F. Corrections involving $F(x)$ and $F(y)$ come from higher terms in the Taylor series of $S_0'(g(x,y))$, while corrections involving the L^2 norm come from the average value of $S_0'(g(x,y))$ on a quadrant being slightly different from $S_0'(p_{ij})$.

The resulting changes Δd_f in the degree function $d(x)$ from g_b to $g_b + \Delta g_f$ are then:

$$\Delta d_f(x) = \begin{cases} F(x)\left(\frac{c}{S''(p_{11})} + \frac{1-c}{S''(p_{12})}\right) + O(F^2) & x \in I \\ F(x)\left(\frac{c}{S''(p_{12})} + \frac{1-c}{S''(p_{22})}\right) + O(F^2) & x \in II. \end{cases} \tag{32}$$

Next we compute $\Delta\tau_f := \tau(g) - \tau(g_b)$ and $\Delta s_f := s(g) - s(g_b)$ to lowest order in F. If we expand $\Delta\tau_f$ and Δs_f in powers of Δg_f, the linear terms vanish exactly, because $\iint \Delta g_f$ is exactly zero on each quadrant. For the quadratic term, we approximate Δg_f using (31). The resulting errors in the quadratic term, and all of the neglected higher-order terms, are then bounded by the sup norm of F times the squared L^2 norm, which we denote $O(F^3)$:

$$\Delta s_f = \frac{1}{2}\iint S_0''(g_b(x,y))\Delta g_f(x,y)^2 + O(\Delta g_f^3)\mathrm{d}x\,\mathrm{d}y$$

$$= \iint_{I\times I} \frac{F(x)^2+F(y)^2}{2S''(p_{11})} + 2\iint_{I\times II} \frac{F(x)^2+F(y)^2}{2S''(p_{12})} + \iint_{II\times II} \frac{F(x)^2+F(y)^2}{2S''(p_{22})} + O(F^3)$$
$$= \left(\frac{c}{S''(p_{11})} + \frac{1-c}{S''(p_{12})}\right)\int_I F(x)^2 \mathrm{d}x + \left(\frac{c}{S''(p_{12})} + \frac{1-c}{S''(p_{22})}\right)\int_{II} F(x)^2 \mathrm{d}x + O(F^3)$$

$$\Delta\tau_f = \int_0^1 \frac{k(k-1)d(x)^{k-2}}{2}(\Delta d(x))^2 + O(\Delta d(x)^3)\mathrm{d}x$$
$$= \frac{k(k-1)d_1^{k-2}}{2}\left(\frac{c}{S''(p_{11})} + \frac{1-c}{S''(p_{12})}\right)^2 \int_I F(x)^2 \mathrm{d}x$$
$$+ \frac{k(k-1)d_2^{k-2}}{2}\left(\frac{c}{S''(p_{12})} + \frac{1-c}{S''(p_{22})}\right)^2 \int_{II} F(x)^2 \mathrm{d}x + O(F^3), \tag{33}$$

where $d_1 = cp_{11}+(1-c)p_{12}$ and $d_2 = cp_{12}+(1-c)p_{22}$ are the values of the degree function for the bipodal graphon g_b. The ratio $\Delta s_f/\Delta\tau_f$ is then a weighted average of

$$\frac{2}{k(k-1)d_1^{k-2}}\left(\frac{c}{S''(p_{11})} + \frac{1-c}{S''(p_{12})}\right)^{-1} \tag{34}$$

and

$$\frac{2}{k(k-1)d_2^{k-2}}\left(\frac{c}{S''(p_{12})} + \frac{1-c}{S''(p_{22})}\right)^{-1} \tag{35}$$

with relative weights

$$d_1^{k-2}\left(\frac{c}{S''(p_{11})} + \frac{1-c}{S''(p_{12})}\right)^2 \int_I F(x)^2 \mathrm{d}x \text{ and } d_2^{k-2}\left(\frac{c}{S''(p_{12})} + \frac{1-c}{S''(p_{22})}\right)^2 \int_{II} F(x)^2 \mathrm{d}x. \tag{36}$$

As $\tau \to \epsilon^k$ (and $c \to 0$ and $F \to 0$), the first ratio being averaged approaches $\psi_k(\zeta_k(\epsilon), \zeta_k(\epsilon))$ and the second approaches $\psi_k(\epsilon, \epsilon)$.

However, both of these numbers are smaller than $\beta = \psi_k(\epsilon, \zeta_k(\epsilon))$. We have already established that $\mathrm{d}s/\mathrm{d}\tau = \beta + o(1)$ for changes in c that preserve the bipodal structure. This means that, for sufficiently small c, if we perturb a bipodal graphon to maximize s for fixed additional change $\Delta\tau_f$, it is better to perturb c than to make F non-zero. Thus $F(x)$ is identically zero, implying that the optimizing graphon is exactly bipodal.

Step 3.

We have established that the minimizing graphon is bipodal, with $p_{22} = \epsilon + o(1)$ and $p_{12} = \zeta_k(\epsilon) + o(1)$. We now show that the form of this graphon is unique. Since the graphon is bipodal, we consider the exact optimality equations for bipodal graphons. The argument then reduces to showing that a certain four-dimensional Jacobian determinant is non-zero. After eliminating γ, we have

$$\begin{aligned} S_0'(p_{11}) &= \alpha + k\beta d_1^{k-1}, \\ S_0'(p_{12}) &= \alpha + \frac{k}{2}\beta(d_1^{k-1} + d_2^{k-1}), \\ S_0'(p_{22}) &= \alpha + k\beta d_2^{k-1}, \\ \frac{\partial S}{\partial c} &= \alpha\frac{\partial\epsilon}{\partial c} + \beta\frac{\partial\tau}{\partial c}, \\ \epsilon &= \epsilon_0, \\ \tau &= \tau_0. \end{aligned} \tag{37}$$

We use the second and third equations to solve for α and β:

$$\begin{aligned} \alpha &= \frac{-S_0'(p_{22})(d_2^{k-1} + d_1^{k-1}) + 2d_2^{k-1}S_0'(p_{12})}{d_2^{k-1} - d_1^{k-1}}, \\ \beta &= \frac{2}{k}\frac{S_0'(p_{22}) - S_0'(p_{12})}{d_2^{k-1} - d_1^{k-1}}. \end{aligned} \tag{38}$$

Plugging this into the first equation then gives

$$S_0'(p_{11}) - 2S_0'(p_{12}) + S_0'(p_{22}) = 0. \tag{39}$$

This leaves four equations in four unknowns, which we write as

$$(f_1, f_2, f_3, f_4) = (0, 0, \epsilon_0, \tau_0), \tag{40}$$

where

$$\begin{aligned} f_1(p_{11}, p_{12}, p_{22}, c) &= S_0'(p_{11}) - 2S_0'(p_{12}) + S_0'(p_{22}), \\ f_2(p_{11}, p_{12}, p_{22}, c) &= \frac{\partial s}{\partial c} - \alpha\frac{\partial\epsilon}{\partial c} - \beta\frac{\partial\tau}{\partial c}, \\ f_3(p_{11}, p_{12}, p_{22}, c) &= c^2p_{11} + 2c(1-c)p_{12} + (1-c)^2p_{22}, \\ f_4(p_{11}, p_{12}, p_{22}, c) &= cd_1^k + (1-c)d_2^k, \end{aligned} \tag{41}$$

and where α and β are given by (38).

We know a solution when $\tau_0 = \epsilon_0^k$, namely $p_{22} = \epsilon_0$, $p_{12} = \zeta_k(\epsilon_0)$, $c = 0$, and $p_{11} = {S_0'}^{-1}(2S_0'[\zeta_k(\epsilon_0)] - S_0'(\epsilon_0))$. We will show that df has non-zero determinant at this point. By the inverse function theorem, this implies that, when τ_0 is close to ϵ_0^k, there is only one value of $(p_{11}, p_{12}, p_{22}, c)$ close to this point for which $f(p_{11}, p_{12}, p_{22}, c) = (0, 0, \epsilon_0, \tau_0)$. Moreover, the parameters $(p_{11}, p_{12}, p_{22}, c)$ depend analytically on ϵ_0 and τ_0. This will complete the proof.

The derivatives of f_1, f_3, and f_4 are:

$$\begin{aligned}
df_1(p_{11}, p_{12}, p_{22}, c) &= (S_0''(p_{11}), -2S_0''(p_{12}), S_0''(p_{22}), 0),\\
df_3(p_{11}, p_{12}, p_{22}, c) &= (c^2, 2c(1-c), (1-c)^2, 2cp_{11} + 2(1-2c)p_{12} - 2(1-c)p_{22}),\\
df_4(p_{11}, p_{12}, p_{22}, c) &= (kc^2 d_1^{k-1}, kc(1-c)(d_1^{k-1} + d_2^{k-1}), k(1-c)^2 d_2^{k-1},\\
&\qquad d_1^k - d_2^k + kcd_1^{k-1}(p_{11} - p_{12}) + k(1-c)d_2^{k-1}(p_{12} - p_{22})). \qquad (42)
\end{aligned}$$

Evaluating at $c = 0$ gives

$$\begin{aligned}
df_1(p_{11}, p_{12}, p_{22}, 0) &= (S_0''(p_{11}), -2S_0''(p_{12}), S_0''(p_{22}), 0),\\
df_3(p_{11}, p_{12}, p_{22}, 0) &= (0, 0, 1, 2p_{12} - 2p_{22}),\\
df_4(p_{11}, p_{12}, p_{22}, 0) &= (0, 0, kp_{22}^{k-1}, p_{12}^k - p_{22}^k + kp_{22}^{k-1}(p_{12} - p_{22})). \qquad (43)
\end{aligned}$$

df is block triangular, with 2×2 blocks. The lower right block has determinant $p_{12}^k - p_{22}^k - kp_{22}^{k-1}(p_{12} - p_{22}) = D(p_{22}, p_{12})$, which is non-zero when $p_{12} \neq p_{22}$, that is, when $\epsilon_0 \neq (k-1)/k$.

When $c = 0$, d_1 and d_2 are independent of p_{11}, as are $\frac{\partial \epsilon}{\partial c}$ and $\frac{\partial \tau}{\partial c}$, so $\frac{\partial f_2}{\partial p_{11}} = 0$. As a result,

$$\det(df) = S_0''(p_{11}) D(p_{22}, p_{12}) \frac{\partial f_2}{\partial p_{12}}. \qquad (44)$$

Since $S_0''(p_{11})$ is never zero, and since $D(p_{22}, p_{12})$ only vanishes when $p_{12} = p_{22}$ (i.e., at $\epsilon_0 = (k-1)/k$), we need only show that $\frac{\partial f_2}{\partial p_{12}} \neq 0$. We compute

$$\begin{aligned}
\frac{\partial \beta}{\partial p_{12}} &= \frac{2}{k} \frac{(p_{22}^{k-1} - p_{12}^{k-1})(-S_0''(p_{12})) - (S_0'(p_{22}) - S_0'(p_{12}))(-(k-1)p_{12}^{k-2})}{(p_{22}^{k-1} - p_{12}^{k-1})^2}\\
&= \frac{2}{k} \frac{(k-1)p_{12}^{k-2}(S_0'(p_{22}) - S_0'(p_{12})) - (p_{22}^{k-1} - p_{12}^{k-1})S_0''(p_{12})}{(p_{22}^{k-1} - p_{12}^{k-1})^2} \qquad (45)
\end{aligned}$$

at $c = 0$.

Since $\alpha = S_0'(p_{22}) - k\beta d_2^{k-1}$,

$$\frac{\partial \alpha}{\partial p_{12}} = -kd_2^{k-1} \frac{\partial \beta}{\partial p_{12}} - k(k-1)\beta d_2^{k-2} \frac{\partial d_2}{\partial p_{12}}$$

$$= -kd_2^{k-1}\frac{\partial \beta}{\partial p_{12}} - k(k-1)d_2^{k-2}c\beta$$

$$\Rightarrow -kp_{22}^{k-1}\frac{\partial \beta}{\partial p_{12}}, \tag{46}$$

where $\Rightarrow$ denotes a limit as $c \to 0$. We also compute

$$\frac{\partial^2 S}{\partial c \partial p_{12}} = 2(1-2c)S_0'(p_{12}) \Rightarrow 2S_0'(p_{12})$$

$$\frac{\partial^2 \epsilon}{\partial c \partial p_{12}} = 2(1-2c) \Rightarrow 2$$

$$\frac{\partial^2 \tau}{\partial c \partial p_{12}} = k(1-2c)(d_1^{k-1} + d_2^{k-1}) \Rightarrow k(p_{12}^{k-1} + p_{22}^{k-1}). \tag{47}$$

Finally we combine everything:

$$\frac{\partial f_2}{\partial p_{12}}\Big|_{c=0} = \frac{\partial^2 S}{\partial c \partial p_{12}} - \frac{\partial \alpha}{\partial p_{12}}\frac{\partial \epsilon}{\partial c} - \alpha\frac{\partial^2 \epsilon}{\partial c \partial p_{12}} - \frac{\partial \beta}{\partial p_{12}}\frac{\partial \tau}{\partial c} - \beta\frac{\partial^2 \tau}{\partial c \partial p_{12}}$$

$$= 2S_0'(p_{12}) - 2\alpha - \beta k(p_{12}^{k-1} + p_{22}^{k-1})$$

$$+ \left(kp_{22}^{k-1}(2p_{12} - 2p_{22}) - (p_{12}^k - p_{22}^k + kp_{22}^{k-1}(p_{12} - p_{22}))\right)\frac{\partial \beta}{\partial p_{12}}. \tag{48}$$

The terms not involving $\partial\beta/\partial p_{12}$ all cancel, by the second equation of (37), and we are left with

$$\frac{\partial f_2}{\partial p_{12}} = -D(p_{12}, p_{22})\frac{\partial \beta}{\partial p_{12}}. \tag{49}$$

Finally, we need to show that $\partial\beta/\partial p_{12} \neq 0$. Since p_{12} maximizes $\psi_k(p_{22}, p_{12})$ for fixed p_{22}, we must have (referring to the notation of the proof of Theorem 3.3) $(N/D)' = 0$, or equivalently $N'/D' = N/D$, where we write $\psi_k = N/D$, as above. But $\beta = N'/D'$. If $\partial\beta/\partial p_{12}$ were equal to zero, then we would have $N''/D'' = N'/D'$. But we have previously shown that it is impossible to simultaneously have $N/D = N'/D' = N''/D''$, except at $p_{12} = p_{22} = (k-1)/k$, so $\partial\beta/\partial p_{12}$ must be non-zero whenever $\epsilon_0 \neq (k-1)/k$. This makes $\det(df)$ non-zero at $(p_{11}, \zeta_k(\epsilon_0), \epsilon_0, 0)$, so the solutions near this point are unique and analytic in (ϵ, τ). ■

5 Theorem 1.1 for k-starlike graphs

Now suppose that H is a k-starlike graph with ℓ edges, and with n_k vertices of degree k, and let τ be the density of H and τ_k be the density of k-stars. Our first result relates $\Delta\tau := \tau - \epsilon^\ell$ to $\Delta\tau_k := \tau_k - \epsilon^k$.

Lemma 5.1. If g is an entropy-maximizing graphon for (ϵ, τ) with $\tau > \epsilon^\ell$, then $\Delta\tau = n_k \epsilon^{\ell-k} \Delta\tau_k + O(\Delta\tau_k^{3/2})$. □

Proof. Writing $g(x,y) = \epsilon + \Delta g(x,y)$, we have

$$\tau = \int \mathrm{d}\mathbf{x} \prod g(x_i, x_j) = \int \mathrm{d}\mathbf{x} \prod (\epsilon + \Delta g(x_i, x_j)), \tag{50}$$

where there is a variable x_i for each vertex of H and the product is over all edges in H.

Expanding the product in the integrand, we get a sum of terms:

- The leading order term ϵ^ℓ.
- Terms with one factor of Δg. These integrate to zero, since $\iint \Delta g(x,y) \mathrm{d}x\, \mathrm{d}y = \Delta\epsilon = 0$.
- Terms with two or more factors of Δg, all coming from edges that share a fixed vertex of degree k. Up to an overall power of $\epsilon^{\ell-k}$, these are identical to the terms of order 2 and higher in Δg in the expansion of $\Delta\tau_k$. As such, they add up to $\epsilon^{\ell-k}\Delta\tau_k$. Summing over the vertices of H then gives $n_k\epsilon^{\ell-k}\Delta\tau_k$.
- Terms with two or more factors of Δg, corresponding to edges that do not all share a vertex. For each such term, let $\{e_i\}$ denote the edges corresponding to factors of Δg. We classify these further into three sub-cases:
    - If one of the e_i's is disconnected from the rest, then the term is identically zero, since $\iint \Delta g(x,y)\mathrm{d}x\, \mathrm{d}y = 0$.
    - If $\{e_i\}$ consists of two or more connected components (each with at least two edges), then the term is a power of ϵ times the product of integrals, one for each connected component. However, each such integral is $O(\|\Delta g\|^2)$, so the term is $O(\|\Delta g\|^4)$.
    - If there is a single connected component whose edges do not all share a vertex, then $\{e_i\}$ must contain three edges that either form a chain or a triangle. We bound such a term by taking absolute values of the Δg's for the three edges and replacing all other factors of Δg by 1. The resulting bound is a power of ϵ times either $\iiiint |\Delta g(w,x)||\Delta g(x,y)||\Delta g(y,z)|\mathrm{d}w\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z$ for a chain or $\iiint |\Delta g(x,y)||\Delta g(y,z)||\Delta g(z,x)|\mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z$ for a triangle, either of which is bounded by that power of ϵ times $\|\Delta g\|^3$.

Thus $\Delta\tau = n_k\epsilon^{\ell-k}\Delta\tau_k + O(\|\Delta g\|^3)$. Since g is entropy maximizing, $\Delta\tau_k$ goes as $\|\Delta g\|^2$, so the error is $O(\Delta\tau_k^{3/2})$. ■

5.1 Proof of Theorem 1.1

Since $\Delta\tau$ is proportional to $\Delta\tau_k$ (plus small errors), the problem of optimizing $\Delta s/\Delta\tau$ is a small perturbation of the problem of optimizing $\Delta s/\Delta\tau_k$, or equivalently optimizing Δs for fixed $\Delta\tau_k$, which we solved in Theorem 4.1. Since that problem has a unique optimizer, any optimizer for $\Delta s/\Delta\tau$ must come close to optimizing $\Delta s/\Delta\tau_k$, and so must be close to the bipodal graphon derived in Theorem 4.1.

We can thus write $g = g_b + \Delta g_f$, as in the last steps of the proof of Theorem 4.1, where $g_b = \epsilon + \Delta g_b$ is a bipodal graphon with $p_{22} = \epsilon + o(1)$ and $p_{12} = \zeta_k(\epsilon) + o(1)$ and where Δg_f is a function that averages to zero on each quadrant of g_b. We again use the convention that words like “small" and “close to" and “negligible” refer to quantities which tend to zero as $\Delta\tau := \tau - \varepsilon^\ell$ tends to zero. A quantity is “nearly constant” if it is constant up to an $o(1)$ correction.

Lemma 5.2. The function Δg_f is pointwise small. That is, as $\tau \to \epsilon^\ell$, Δg_f goes to zero in sup-norm. □

Proof of Lemma. Since we are no longer in the setting where the entropy maximizer is proven to be multipodal, we cannot use the equations (25) directly. However, we can still apply the method of Lagrange multipliers to pointwise variations of the graphon. (See [6] for a rigorous justification.) These variational equations are

$$\frac{1}{2}\ln\left(\frac{1}{g(x,y)} - 1\right) = \frac{\delta s}{\delta g(x,y)} = \alpha + \beta\frac{\delta\tau}{\delta g(x,y)}. \tag{51}$$

We need to compute $\delta\tau/\delta g$ and show that it is nearly constant on each quadrant. Since α and β are constants, (51) would then imply that $g(x,y)$ is nearly constant on each quadrant, and hence that Δg_f is pointwise small.

Let $g_0(x,y) \equiv \epsilon$. Since $\|\Delta g\|$ is small (where $\Delta g = g - g_0 = \Delta g_b + \Delta g_f$), we can find a small constant $a = o(1)$ such that, for all x outside a set $U \subset [0,1]$ of measure a, $\int_0^1 |\Delta g(x,y)|\mathrm{d}y < a$. (This set U is essentially what we previously called the Type I clusters, but at this stage of the argument we are not assuming a multipodal structure. Rather, we are just using the fact that $\tau - e^\ell = O(\|\Delta g\|^2)$.)

The functional derivative $\delta\tau/\delta g(x,y)$ has a diagrammatic expansion similar to the expansion of τ in (50). For each edge of H, we get a contribution by deleting the edge, assigning the values x and y to the endpoints of the edge, and integrating over the values of all other vertices. Since U is small, we can estimate $\delta\tau/\delta g$ to within $o(1)$ by restricting the integral to $(U^c)^{v-2}$, where v is the number of vertices in H and U^c is the complement of U. This implies that terms involving Δg can only contribute non-negligibly on edges connected to x or to y. Furthermore, they can only contribute non-negligibly when attached to x if $x \in U$, and can only contribute non-negligibly when attached to y if $y \in U$.

We now begin a bootstrap argument. We will show that $\delta\tau/\delta g$ is nearly constant on each quadrant $U^c \times U^c, U \times U^c, U \times U$ in turn. This will show that g is nearly constant on that quadrant, which will help us prove that $\delta\tau/\delta g$ is nearly constant on the next quadrant.

The simplest case is when x and y are both in U^c. Then the contributions of the terms involving Δg are negligible, so $\delta\tau/\delta g(x, y)$ can be computed, to within a small error, using the approximation $g \approx g_0$. But when Δg is negligible, $\delta\tau/\delta g(x, y)$ is nearly independent of x and y. Since $\delta\tau/\delta g(x, y)$ is nearly constant on $U^c \times U^c$, equation (51) implies that g is nearly constant on $U^c \times U^c$.

Next suppose that $y \in U^c$ and $x \in U$. Then all contributions from factors of $\Delta g(z, y)$ are negligible, so $\delta\tau/\delta g(x, y)$ is nearly independent of y. But then $g(x, y)$ is nearly independent of y, and is nearly equal to $d(x)$. The integrals involved in computing $\delta\tau/\delta g(x, y)$ are then easily approximated to within $o(1)$, using $g_0 + \Delta g$ on the edges connected to x, g_0 on all other edges, and only integrating over $(U^c)^{v-2}$. If the degree of x is k, then the edges connected to x contribute $d(x)^{k-1}e^{\ell-k}$. Summing over edges, and symmetrizing over the assignment of x and y to the two endpoints, we obtain the approximation

$$\frac{\delta\tau}{\delta g(x, y)} = \frac{kn_k\epsilon^{\ell-k}}{2}\left(d(x)^{k-1} + d(y)^{k-1}\right) + o(1). \tag{52}$$

Up to an overall factor of $n_k\epsilon^{\ell-k}$, this is the same functional derivative as for a k-star. This also applies if $x \in U^c$, except that in the latter case $d(x) \approx \epsilon$, and also applies if $x \in U^c$ and $y \in U$.

In other words, we can use the approximation (52) in (51) whenever *either* x or y (or both) is in U^c. This implies that the integrated equations (26) apply for all x (with d_i replaced by $d(x)$, and with β scaled up by $n_k\epsilon^{\ell-k}$). Following the exact same argument as in the proof of Theorem 4.1, we obtain that $d(x)$ only takes on two possible values (up to $o(1)$ errors), namely ϵ and $\zeta_k(\epsilon)$. We then define Types I and II points, depending on whether the degree function is close to $\zeta_k(\epsilon)$ or ϵ, respectively, and can take U to be precisely the set of Type I points. Our graphon is then nearly constant on $U \times U^c$ and $U^c \times U$, as well as on $U^c \times U^c$.

We still need to show that the graphon is nearly constant on $U \times U$. Suppose that x and y are in U. Since $g(x,z)$ is nearly independent of x for z in U^c, and since $\delta\tau/\delta g(x,y)$ is computed to within $o(1)$ by integrating over $(U^c)^{v-2}$, $\delta\tau/\delta g(x,y)$ is nearly independent of $x \in U$, and likewise nearly independent of $y \in U$. But then $g(x,y)$ is nearly constant on $U \times U$.

Note, by the way, that the approximation (52) does not apply on $U \times U$; in that case $\delta\tau/\delta g$ contains terms with powers of both $d(x)$ and $d(y)$. However, that approximation is not needed for our proof, since $U \times U$ (aka the *I*-*I* quadrant) only contributes $O(c)$ to the integrated equations (26). ■

Returning to the proof of Theorem 1.1, we need to compare $s(g_b + \Delta g_f) - s(g_b)$ to $\tau(g_b + \Delta g_f) - \tau(g_b)$.

As before, we expand $\tau(g)$ as the integral of a polynomial in g, obtained by assigning $g_0 + \Delta g_b + \Delta g_f$ to each edge of H and integrating. The difference between $\tau(g_b + \Delta g_f)$ and $\tau(g_b)$ consists of terms with at least one Δg_f. However, the terms with *exactly* one Δg_f are identically zero, since g_b is constant on quadrants, and Δg_f averages to zero on each quadrant. Furthermore, terms for which all of the Δg_b's and Δg_f's share a vertex are exactly what we would get from the approximation $\Delta\tau \approx n_k \epsilon^{\ell-k} \tau_k$. Any term that distinguishes between $\Delta\tau$ and $n_k \epsilon^{\ell-k} \Delta\tau_k$ must have at least two Δg_f's and either a third Δg_f or a Δg_b, forming either a 3-chain, a triangle, or two connected Δg_f's and a disconnected Δg_b.

Let $\Delta g_f'(x,y) = |\Delta g_f(x,y)|$, and let

$$\Delta g_b'(x,y) = \begin{cases} 2c & x, y \in II, \\ 1 & \text{otherwise.} \end{cases} \tag{53}$$

This is conveniently expressed in terms of outer products. Let $|1\rangle \in L^2([0,1])$ be the constant function 1, and let $|\omega\rangle$ be the function

$$\omega(x) = \begin{cases} 0 & x < c, \\ 1 & x > c. \end{cases} \tag{54}$$

Then

$$\begin{aligned} \Delta g_b' &= |1\rangle\langle 1| - |\omega\rangle\langle\omega| + 2c|\omega\rangle\langle\omega| \\ &= |1\rangle\langle 1-\omega| + |1-\omega\rangle\langle\omega| + 2c|\omega\rangle\langle\omega|. \end{aligned} \tag{55}$$

Note that $|\Delta g_b(x,y)| \leq \Delta g'_b(x,y)$ for all $x, y \in (0,1)$. To see this, the only issue is what happens when (x,y) is in the *II*-*II* quadrant, since otherwise we trivially have $|\Delta g_b| \leq 1$. Since $e(g)$ is fixed, $(1-c)^2$ times $\Delta g_b(x,y)$ for $x, y > c$ equals minus the integral of Δg_b over the other three quadrants. But the area of those three quadrants is $2c - c^2 < 2c$, and the biggest possible value of $|\Delta g_b|$ is $\max(e, 1-e) < 1$, so $\frac{1}{(1-c)^2}\int |\Delta g_b|$ (integrated over the *I*-*I*, *I*-*II*, and *II*-*I* quadrants) is strictly less than $2c + O(c^2)$, and so is bounded by $2c$ for small c (note that $O(c^2)$ errors are negligible).

We obtain upper bounds on the contributions of the relevant terms in the expansion of τ by replacing three $\Delta g_f(x,y)$'s and $\Delta g_b(x,y)$'s with $\Delta g'_f(x,y)$ and $\Delta g'_b(x,y)$, respectively, and replacing all other terms with 1.

Since all graphons are symmetric, hence Hermitian, their operator norms are bounded by their L^2 norms, so for any 3-chain

$$\langle 1|\Delta g'_1 \Delta g'_2 \Delta g'_3|1\rangle \leq \|\Delta g'_1\|\|\Delta g'_2\|\|\Delta g'_3\|. \tag{56}$$

Since $\|\Delta g'_b\|$ and $\|\Delta g'_f\|$ are both $o(1)$ (more precisely, $O(\sqrt{\tau - \epsilon^\ell})$), the contribution of any 3-chain is bounded by an $o(1)$ constant times $\|\Delta g_f\|^2$.

As for triangles, $\mathrm{Tr}(\Delta g'^3_f) \leq \|\Delta g'_f\|^3 = \|\Delta g_f\|^3$. Finally, we must estimate the trace of $\Delta g'_f \Delta g'_f \Delta g'_b$. But this trace is

$$\langle 1-\omega|\Delta g'_f \Delta g'_f|1\rangle + \langle \omega|\Delta g'_f \Delta g'_f|1-\omega\rangle + 2c\langle \omega|\Delta g'_f \Delta g'_f|\omega\rangle. \tag{57}$$

Since $\|1-\omega\| = \sqrt{c}$, the total is bounded by $(2\sqrt{c} + 2c^2)\|\Delta g_f\|^2$.

The upshot is that the ratio of $s(g_b + \Delta g_f) - s(g_b)$ and $\tau(g_b + \Delta g_f) - \tau(g_b)$ is the same as that computed for k-stars (up to an overall factor of $n_k \epsilon^{\ell-k}$), plus an $o(1)$ correction. But that ratio was bounded by a constant $\beta_0 < \beta$. Restricting attention to values of τ for which the correction is smaller than $(\beta - \beta_0)/2$, we still obtain the result that having a non-zero Δg_f is a less efficient way of generating additional τ than simply changing c. Thus the optimizing graphon is exactly bipodal.

Once bipodality is established, uniqueness follows exactly as in the proof of Theorem 4.1. The difference between $\Delta\tau$ and $n_k \epsilon^{\ell-k} \Delta\tau_k$ is of order $c^{3/2}$, and so does not affect the linearization of the optimality equations at $c = 0$.

6 Linear combinations of k-stars

We proved Theorem 1.1 by first showing that k-star models have the desired behavior, and then showing that, for an arbitrary k-starlike graph H, $\Delta\tau$ is well-approximated by a multiple of $\Delta\tau_k$, so the model with densities of edges and H behaves essentially the same as a model with densities of edges and k-stars.

To prove Theorem 1.2, we consider in this section a family of models in which we can prove bipodality and uniqueness of entropy maximizers directly, as we did for k-stars. In the next section, we will show how to approximate a model with an arbitrary H with a model in this family.

Let $h(x) = \sum_{k\geq 1} a_k x^k$ be a polynomial with non-negative coefficients and degree ≥ 2. Let $\tau = \sum a_k \tau_k$, and consider graphs with fixed edge density ϵ and fixed τ. In [6] it was proved that the entropy-maximizing graphons in such models are always multipodal.

Most of the analysis of k-star models carries over to positive linear combinations, and so will only be sketched briefly. We will provide complete details where the arguments differ.

In analogy to the notation of the proof of Theorem 3.3, let $\psi(\epsilon,\tilde{\epsilon}) = N/D$, where

$$\begin{aligned} N(\epsilon,\tilde{\epsilon}) &= 2[S_0(\tilde{\epsilon}) - S_0(\epsilon) - (\tilde{\epsilon}-\epsilon)S_0'(\epsilon)], \\ D(\epsilon,\tilde{\epsilon}) &= h(\tilde{\epsilon}) - h(\epsilon) - (\tilde{\epsilon}-\epsilon)h'(\epsilon). \end{aligned} \tag{58}$$

Since $h''(x)$ is positive for $x > 0$, D is only zero when $\tilde{\epsilon} = \epsilon$, and we fill in that removable singularity in ψ by defining $\psi(\epsilon,\epsilon) = 2S_0''(\epsilon)/h''(\epsilon)$.

Theorem 6.1. For all but finitely many values of ϵ, there is a $\tau_0 > h(\epsilon)$ such that, for $\tau \in (h(\epsilon), \tau_0)$, the entropy-optimizing graphon is bipodal and unique, with data varying analytically with ϵ and τ. As τ approaches $h(\epsilon)$ from above, $p_{22} \to \epsilon$, p_{12} approaches a point $\tilde{\epsilon}$ where $\psi'(\epsilon,\tilde{\epsilon}) = 0$, p_{11} satisfies $S_0'(p_{11}) = 2S_0'(p_{12}) - S_0'(p_{22})$, and $c \to 0$ as $O(\Delta\tau)$. □

Proof. For a multipodal graphon, $\tau(g) = \sum c_i h(d_i)$. After eliminating γ, the optimality equations become

$$S_0'(p_{ij}) = \alpha + \beta(h'(d_i) + h'(d_j))/2, \tag{59}$$

$$2\sum_{j=1} c_j(S_0(p_{ij}) - S_0(p_{Mj})) = 2\alpha(d_i - d_M) + \beta\Big[h(d_i) - h(d_M) + \sum_{j=1}^{M} c_j h'(d_j)(p_{ij} - p_{Mj})\Big]. \tag{60}$$

As before, we distinguish between Type I clusters that are small and Type II clusters that have $d_i \approx \epsilon$. Summing the optimality equations over j of Type II, and approximating d_j by ϵ, we obtain the equations

$$S_0'(d_i) = \alpha + \beta\left(h'(d_i) + h'(\epsilon)\right)/2 + o(1), \tag{61}$$

$$S_0(d_i) - S_0(\epsilon) = \alpha(d_i - \epsilon) + \beta\left[h(d_i) - h(\epsilon) + h'(\epsilon)(d_i - \epsilon)\right] + o(1). \tag{62}$$

We use the first equation, with $i = M$ (a type II cluster), to solve for α, and plug it into the equations for $i < M$ to get

$$2(S_0'(d_i) - S_0'(\epsilon)) = \beta(h'(d_i) - h'(\epsilon)) + o(1), \tag{63}$$

$$2[S_0(d_i) - S_0(\epsilon) - S_0'(\epsilon)(d_i - \epsilon)] = \beta[h(d_i) - h(\epsilon) - h'(\epsilon)(d_i - \epsilon)] + o(1). \tag{64}$$

As before in the proof of Theorem 4.1, this implies that either $d_i \approx \epsilon$ or that $\psi(\epsilon, d_i)$ is maximized with respect to d_i.

Unlike in the k-star case, it is not true that $\psi'(\epsilon, \tilde{\epsilon})$ has a unique solution for each ϵ. However, it remains true that $\psi(\epsilon, \tilde{\epsilon})$ has a unique global maximizer (w.r.t. $\tilde{\epsilon}$) for all but finitely many values of ϵ. Since the equations defining multiple maxima are analytic, they must be satisfied either for all ϵ or for only finitely many ϵ. But it is straightforward to check that there is only one maximizer when ϵ is sufficiently small, since then $h(\epsilon)$ and $h'(\epsilon)$ are dominated by the lowest order term in the polynomial.

Thus, for all but finitely many values of ϵ, the values of d_i must all either approximate ϵ or the unique value of $\tilde{\epsilon}$ that maximizes $\psi(\epsilon, \tilde{\epsilon})$. This allows for a re-segregation of the clusters into Type I (with d_i close to $\tilde{\epsilon}$) and Type II (with d_i close to ϵ) and yields a graphon that is approximately bipodal. Step 2 of the proof of Theorem 4.1, proving that the optimizing graphon is exactly bipodal with data of the desired form, then proceeds exactly as before.

What remains is showing that the optimizing graphon is unique by linearizing the exact optimality equations for bipodal graphons near $c = 0$. These equations are:

$$\begin{aligned}
S_0'(p_{11}) &= \alpha + \beta h'(d_1), \\
S_0'(p_{12}) &= \alpha + \beta(h'(d_1) + h'(d_2))/2, \\
S_0'(p_{22}) &= \alpha + \beta h'(d_2), \\
\frac{\partial S}{\partial c} &= \alpha \frac{\partial \epsilon}{\partial c} + \beta \frac{\partial \epsilon}{\partial c}, \\
\epsilon &= \epsilon_0, \\
\tau &= \tau_0.
\end{aligned} \tag{65}$$

Using the second and third equations to eliminate α and β gives:

$$\begin{aligned}
\alpha &= \frac{2h'(d_2)S_0'(p_{12}) - S_0'(p_{22})(h'(d_2) + h'(d_1))}{h'(d_2) - h'(d_1)}, \\
\beta &= \frac{2(S_0'(p_{22}) - S_0'(p_{12}))}{h'(d_2) - h'(d_1)}.
\end{aligned} \tag{66}$$

We also have $\alpha = S_0'(p_{22}) - \beta h'(d_2)$ and $S_0'(p_{11}) = 2S_0'(p_{12}) - S_0'(p_{22})$. Note that

$$\frac{\partial\alpha}{\partial p_{12}} = -\beta c h''(d_2) - h'(d_2)\frac{\partial\beta}{\partial p_{12}} \Rightarrow -h'(p_{22})\frac{\partial\beta}{\partial p_{12}} \tag{67}$$

as $c \searrow 0$.

We define $f(p_{11}, p_{12}, p_{22}, c) = (f_1, f_2, f_3, f_4)$ as before, with $f_3 = \epsilon$ and $f_4 = \tau$, and compute

$$\begin{aligned} df_3 &= (c^2, 2c(1-c), (1-c)^2, 2cp_{11} + 2(1-2c)p_{12} - 2(1-c)p_{22}) \\ &\Rightarrow (0, 0, 1, 2(p_{12} - p_{22})), \\ df_4 &= (c^2 h'(d_1), c(1-c)(h'(d_1) + h'(d_2)), (1-c)^2 h'(d_2), \\ &\qquad h(d_1) - h(d_2) + ch'(d_1)(p_{11} - p_{12}) + h'(d_2)(p_{12} - p_{22})) \\ &\Rightarrow (0, 0, h'(p_{22}), h(p_{12}) - h(p_{22}) + h'(p_{22})(p_{12} - p_{22})). \end{aligned} \tag{68}$$

The lower right block of df then gives a contribution of $h(p_{12}) - h(p_{22}) + h'(p_{22})(p_{12} - p_{22}) - 2h'(p_{22})(p_{12} - p_{22}) = h(p_{12}) - h(p_{22}) - h'(p_{22})(p_{12} - p_{22}) = D(p_{22}, p_{12})$.

As before, $\frac{\partial f_2}{\partial p_{11}} = 0$ when $c = 0$, so $\det(df) = S_0''(p_{11})(h(p_{11}) - h(p_{22}) - h'(p_{22})(p_{12} - p_{22}))\frac{\partial f_2}{\partial p_{11}}$. Now

$$\frac{\partial f_2}{\partial p_{12}} = \frac{\partial^2 S}{\partial c \partial p_{12}} - \alpha\frac{\partial^2 \epsilon}{\partial c \partial p_{12}} - \beta\frac{\partial^2 \tau}{\partial c \partial p_{12}} - \frac{\partial\alpha}{\partial p_{12}}\frac{\partial\epsilon}{\partial c} - \frac{\partial\beta}{\partial p_{12}}\frac{\partial\tau}{\partial c}. \tag{69}$$

Since α and β are independent of c, the first three terms are

$$\frac{\partial}{\partial c}\left(\frac{\partial S}{\partial p_{12}} - \alpha\frac{\partial\epsilon}{\partial p_{12}} - \beta\frac{\partial\tau}{\partial p_{12}}\right) = \frac{\partial}{\partial c}(0) = 0, \tag{70}$$

by the second equation of (65). This leaves

$$\partial f_2/\partial p_{12} = (h'(p_{22})(2p_{12} - 2p_{22}) - (h(p_{12}) - h(p_{22}) + h'(p_{22})(p_{12} - p_{22})))\partial\beta/\partial p_{12}. \tag{71}$$

Combining with our earlier results, we have:

$$\det(df) = -S_0''(p_{11})D(p_{22}, p_{12})^2\frac{\partial\beta}{\partial p_{12}}. \tag{72}$$

The expression $D(p_{22}, p_{12}) = h(p_{12}) - h(p_{22}) - h'(p_{22})(p_{12} - p_{22})$ has a double root at $p_{12} = p_{22}$ and is non-zero elsewhere, thanks to the monotonicity of h'.

As a last step, we consider when $\frac{\partial\beta}{\partial p_{12}}$ can be zero. Since $\beta = N'/D'$, we are interested in when $(N'/D')' = 0$. But that is equivalent to having $N''/D'' = N'/D'$. Since we already have $N/D = N'/D'$, this means that $\psi'' = (N/D)'' = 0$. Since we are looking at the value of $\tilde{\epsilon}$ that maximizes ψ, having $\psi' = \psi'' = 0$ would imply $\psi''' = 0$ (or else $\tilde{\epsilon}$ would only be a point of inflection, and not a local maximum). But if $(N/D)' = (N/D)'' = (N/D)''' = 0$, then $N/D = N'/D' = N''/D'' = N'''/D'''$. Note that N'', N''', D'', and D''' are functions of $\tilde{e}$ only, and are rational functions:

$$
\begin{aligned}
N'' &= 2S_0''(\tilde{e}) = \frac{-1}{\tilde{e}} - \frac{1}{1-\tilde{e}},\\
N''' &= 2S_0'''(\tilde{e}) = \frac{1}{\tilde{e}^2} - \frac{1}{(1-\tilde{e})^2},\\
D'' &= h''(\tilde{e}),\\
D''' &= h'''(\tilde{e}).
\end{aligned}
\tag{73}
$$

Setting $D''N''' = D'''N''$ gives a polynomial equation for $\tilde{\epsilon}$, which has only finitely many roots. Since the equation $\psi' = 0$ is symmetric is ϵ and $\tilde{\epsilon}$, $\tilde{\epsilon}$ determines ϵ, so there are only finitely many values of ϵ for which $\frac{\partial \beta}{\partial p_{12}}$ is zero.

In summary, we exclude the finitely many values of ϵ for which ψ achieves its maximum more than once, and the finitely many values of ϵ for which $\frac{\partial \beta}{\partial p_{12}} = 0$. For all other values of ϵ, the optimizing graphon is bipodal of the prescribed form and unique. ■

7 Proof of Theorem 1.2

The proof has three steps.

- Step 1. Showing that, for fixed ϵ, $\Delta\tau$ can be approximated by the change in a positive linear combination of τ_k's.
- Step 2. Defining a set $B_H \subset (0,1)$ of "bad values," determined by analytic equations, such that for all $\epsilon \notin B_H$ and for τ close enough to ϵ^ℓ, the optimizing graphon is unique and bipodal and of the desired form.
- Step 3. Showing that B_H is finite.

Step 1.

This is a repetition of the proof of Lemma 5.1. In the expansion of $\Delta\tau$, we get a contribution $n_k \epsilon^{\ell-k} \Delta\tau_k$ from diagrams where all the edges associated with Δg are connected to a vertex of degree k, where n_k is the number of vertices of H of degree k. Summing over k, and bounding the remaining terms by $O(\|\Delta g\|^3)$, as before, we have

$$\Delta\tau = \sum_k n_k \epsilon^{\ell-k} \Delta\tau_k + O(\Delta\tau^{3/2}). \tag{74}$$

Step 2.

For fixed ϵ, we consider a model whose density is $\sum_k n_k \epsilon^{\ell-k} \tau_k$. As long as $\psi(\epsilon, \tilde{\epsilon})$ for this model achieves its maximum at a unique value of $\tilde{\epsilon}$, and as long as $\partial\beta/\partial p_{12} \neq 0$ when p_{12} equals this value of $\tilde{\epsilon}$, the proofs of Theorems 1.1 and 6.1 carry over almost verbatim.

That is, the model problem has a unique bipodal maximizer by the reasoning of Theorem 6.1. The entropy maximizer for the actual problem involving H must approximate the entropy maximizer for the model problem, and in particular must be approximately bipodal, and so can be written as $g_b + \Delta g_f$, where Δg_f averages to zero on each quadrant. The same arguments as in the proof of Theorem 1.1 show that Δg_f is pointwise small. By a power series expansion, $\frac{s(g_b+\Delta g_f)-s(g_b)}{\tau(g_b+\Delta g_f)-\tau(g_b)} < \beta$, so for small c we can increase the entropy by setting Δg_f to zero and varying the bipodal data to achieve the correct value of τ.

Step 3.

For any fixed ϵ, the model problem has only a finite number of bad values of ϵ, but this is not enough to prove that B_H is finite. Rather

$$B_H = \{\epsilon | \epsilon \text{ is one of the bad points for the model with } a_k = n_k \epsilon^{\ell-k}\}, \tag{75}$$

where a value of ϵ is bad for a model if either ψ has multiple maxima or if $\partial\beta/\partial p_{12} = 0$. Since the bad points for any linear combination of k-stars depends analytically on the coefficients of that linear combination, and since these coefficients are powers of ϵ, the set B_H is cut out by analytic equations in ϵ.

As such, B_H is either the entire interval $(0, 1)$, or a finite set, or a countable set with limit points only at 0 and/or 1. We will show that neither 0 nor 1 is a limit point of B_H, implying that B_H is finite.

Let $k_{\max}$ be the largest degree of any vertex in H, and consider the model problem with $h(x) = \sum_{k=2}^{k_{\max}} a_k x^k$, where $a_k = n_k \epsilon^{\ell-k}$. We begin with some constraints on the values of $\tilde{\epsilon}$ for which $\psi' = 0$.

Lemma 7.1. Suppose that $\psi'(\epsilon, \tilde{\epsilon}) = 0$. If $\tilde{\epsilon} = \epsilon$, or if $\partial\beta/\partial p_{12} = 0$ when $p_{22} = \epsilon$ and $p_{12} = \tilde{\epsilon}$, then $(1/2) \leq \tilde{\epsilon} \leq (k_{\max} - 1)/k_{\max}$. □

Proof of Lemma. In both cases we are looking for solutions to $N''D''' = N'''D''$. Since $N'' = 2S_0''(\tilde\epsilon)$, $N''' = 2S_0'''(\tilde\epsilon)$, $D'' = h''(\tilde\epsilon)$, and $D''' = h'''(\epsilon)$, this equation does not involve ϵ (except insofar as the coefficients of h depend on ϵ). We have

$$\begin{aligned}
\frac{2S_0'''(\tilde\epsilon)}{2S_0''(\tilde\epsilon)} &= \frac{h'''(\tilde\epsilon)}{h''(\tilde\epsilon)},\\
\frac{1}{1-\tilde\epsilon} - \frac{1}{\tilde\epsilon} &= \frac{h'''(\tilde\epsilon)}{h''(\tilde\epsilon)},\\
\frac{2\tilde\epsilon - 1}{1-\tilde\epsilon} &= \frac{\tilde\epsilon h'''(\tilde\epsilon)}{h''(\tilde\epsilon)},\\
\frac{1}{1-\tilde\epsilon} - 2 &= \frac{\sum k(k-1)(k-2)a_k\tilde\epsilon^{k-2}}{\sum k(k-1)a_k\tilde\epsilon^{k-2}}.
\end{aligned} \tag{76}$$

The right-hand side of the last line is a weighted average of $k-2$ with weights $k(k-1)$ $a_k\tilde\epsilon^{k-2}$, and so is at least zero and at most $k_{\max} - 2$. Thus $(1-\tilde\epsilon)^{-1}$ is between 2 and $k_{\max}$ and $\tilde\epsilon$ is between $1/2$ and $(k_{\max}-1)/k_{\max}$. ■

Lemma 7.2. If $\psi'(\epsilon, \tilde\epsilon) = 0$, and if ϵ is sufficiently close to 1, then $\tilde\epsilon$ is uniquely defined and approaches 0 as $\epsilon \to 1$. Likewise, if ϵ is sufficiently close to 0, then $\tilde\epsilon$ is uniquely defined and approaches 1 as $\epsilon \to 0$. □

Proof. When $\epsilon < 1/2$, or when $\epsilon > (k_{\max}-1)/k_{\max}$, we cannot have $\tilde\epsilon = \epsilon$, so the equation $\psi' = 0$ is equivalent to $ND' = DN'$ and $\tilde\epsilon \neq \epsilon$. Writing $DN' - ND' = 0$ explicitly, and doing some simple algebra, yields the equation

$$S_0'(\epsilon)[h(\tilde\epsilon) - h(\epsilon) - (\tilde\epsilon - \epsilon)h'(\epsilon)] - S'(\tilde\epsilon)[[h(\tilde\epsilon) - h(\epsilon) - (\tilde\epsilon - \epsilon)h'(\tilde\epsilon)] + (S_0(\tilde\epsilon) - S_0(\epsilon))(h'(\tilde\epsilon) - h'(\epsilon)) = 0. \tag{77}$$

If ϵ approaches 0 or 1 and $\tilde\epsilon$ does not, then the first term diverges, while the other terms do not, insofar as S_0' has singularities at 0 and 1 but S_0, h, and h' do not. Thus $\tilde\epsilon$ must go to 0 or 1 as ϵ goes to 0 or 1.

We next rule out the possibility that both ϵ and $\tilde\epsilon$ approach 1. Suppose that ϵ is close to 1. We expand both N and D in powers of $(\tilde\epsilon - \epsilon)$:

$$\begin{aligned}
N &= \sum_{m=2}^{\infty} \frac{2S_0^{(m)}(\epsilon)}{m!}(\tilde\epsilon - \epsilon)^m\\
&= -\sum_{m=2}^{\infty}\left(\frac{1}{(1-\epsilon)^{m-1}} + \frac{(-1)^m}{\epsilon^{m-1}}\right)\frac{(\tilde\epsilon - \epsilon)}{m(m-1)},\\
D &= \sum_{m=2}^{k_{\max}} \frac{h^{(m)}(\epsilon)}{m!}(\tilde\epsilon - \epsilon)^m,
\end{aligned} \tag{78}$$

where $S_0^{(m)}$ and $h^{(m)}$ denote mth derivatives. The coefficients of the numerator grow rapidly with m, while the growth of the coefficients of the denominator depend only on the degree of h. For $\tilde{\epsilon} > \epsilon > (k_{\max}-1)/k_{\max}$, $\psi = N/D$ is a decreasing function of $\tilde{\epsilon}$ (i.e., negative and increasing in magnitude), so we cannot have $\psi' = 0$. Since the equation $\psi' = 0$ is symmetric in ϵ and $\tilde{\epsilon}$ (apart from the dependence of the coefficients of h on ϵ), we also cannot have $\epsilon > \tilde{\epsilon} > (k_{\max} - 1)/k_{\max}$.

When ϵ is close to 1, we must thus have $\tilde{\epsilon}$ close to 0. But then $N \approx 2S_0'(\epsilon)$, $D \approx h'(\epsilon) - h(\epsilon)$, $D' \approx -h'(E)$, and the equation

$$2S_0'(\tilde{\epsilon}) = N' + 2S_0'(\epsilon) = 2S_0'(\epsilon) + ND'/D \tag{79}$$

determines $S_0'(\tilde{\epsilon})$, and therefore $\tilde{\epsilon}$, uniquely as a function of ϵ.

Next we consider $\epsilon \to 0$. If H is 2-starlike, then ψ is a multiple of ψ_2, and the result is already known. Otherwise, it is convenient to define a new polynomial $\bar{h}(z) = \sum n_k z^k$, so that $h(x) = \epsilon^\ell \bar{h}(x/\epsilon)$. Then

$$\begin{aligned} D &= h(\tilde{\epsilon}) - h(\epsilon) - h'(\epsilon)(\tilde{\epsilon} - \epsilon) \\ &= \epsilon^\ell[\bar{h}(r) - \bar{h}(1) - \bar{h}'(1)(r-1)], \end{aligned} \tag{80}$$

where $r := \tilde{\epsilon}/\epsilon$. Likewise,

$$N = -\left[\tilde{\epsilon}\ln(\tilde{\epsilon}) - \epsilon\ln(\epsilon) + (1-\tilde{\epsilon})\ln(1-\tilde{\epsilon}) - (1-\epsilon)(1-\tilde{\epsilon}) - (\tilde{\epsilon}-\epsilon)(\ln(\epsilon) - \ln(1-\epsilon))\right]. \tag{81}$$

Since ϵ and $\tilde{\epsilon}$ are small, we can approximate $\ln(1-\epsilon)$ and $\ln(1-\tilde{\epsilon})$ as $-\epsilon$ and $-\tilde{\epsilon}$, respectively, giving

$$N \approx -\epsilon[r\ln r - r + 1] + \epsilon^2(r - r^2). \tag{82}$$

The ratio $\psi = N/D$ is negative. Since $\bar{h}$ is a polynomial of degree at least 3, D grows faster than N as $r \to \infty$, so we can always increase ψ by taking larger and larger values of $r = \tilde{\epsilon}/\epsilon$. This argument only breaks down when the approximation $\ln(1-\tilde{E}) \approx -\tilde{\epsilon}$ breaks down, that is, at values of $\tilde{\epsilon}$ that are no longer close to 0. Thus we cannot have $\tilde{\epsilon}$ and ϵ both close to zero.

Finally, if ϵ is close to 0 and $\tilde{\epsilon}$ is close to 1, then $h(\epsilon)$ and $h'(\epsilon)$ are close to zero, while $h(\tilde{\epsilon})$ is close to a multiple of $x^{k_{\max}}$, since the coefficient of $x^{k_{\max}}$ is $O(1/\epsilon)$ larger than any other coefficient. Thus ψ behaves like $\psi_{k_{\max}}$, and has a unique maximizer. ■

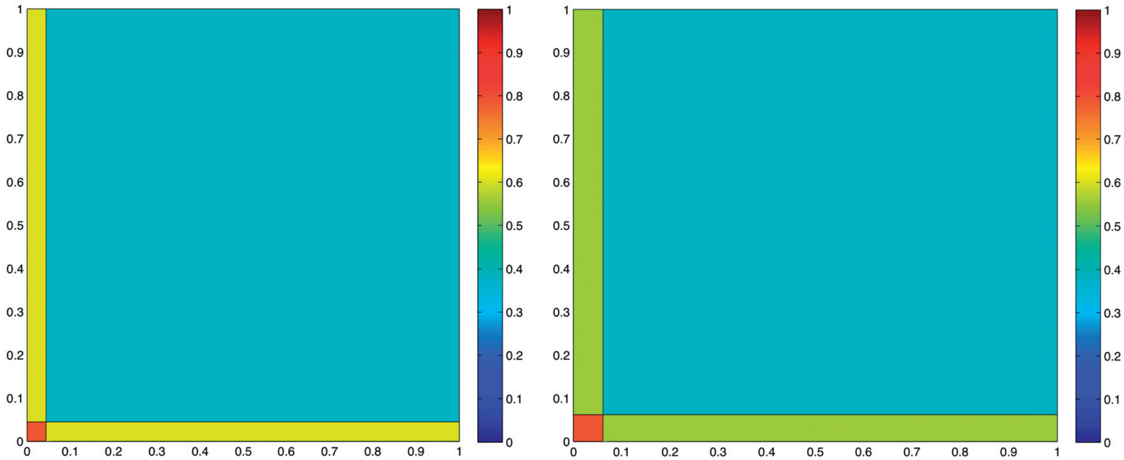

Fig. 2. Numerical estimates of the optimizing graphon for the 2-star model with $\epsilon = 0.4$ and $\tau_2 = 0.1620$ (left) and the optimizing graphon for the triangle model with $\epsilon = 0.4$ and $\tau_{\text{triangle}} = 0.0664$ (right). (Although theoretically we have not tried to prove that $\Delta\tau_2 = 0.002$ is small enough to fit into the interval provided by Theorem 1.1, numerically it appears to be the case.)

We have shown that when ϵ is close to 0 or 1, ψ has a unique maximizer. Furthermore, $\tilde{\epsilon}$ is not between 1/2 and $(k_{\max}-1)/k_{\max}$, so $\partial\beta/\partial p_{12} \neq 0$. So $\epsilon \notin B_H$, completing Step 3 and the proof of Theorem 1.2.

8 Conclusions

We have shown that just above the ER curve, entropy maximizing graphons, constrained by the densities of edges and any one other subgraph H, exhibit the same qualitative behavior for all H and for (almost) all values of ϵ. The optimizing graphon is unique and bipodal.

These results were proven by perturbation theory, using the fact that the optimizing graphon has to be L^2-close to a constant (Erdős–Rényi) graphon. Surprisingly, the optimizing graphon is *not* pointwise close to constant. Rather, it is bipodal, with a small cluster of size $O(\Delta\tau)$. As $\Delta\tau$ approaches 0, the size of the small cluster shrinks, but the values of the graphon on each quadrant do not approach one another. Rather, p_{22} approaches ϵ, p_{12} approaches the value of $\tilde{\epsilon}$ that maximizes a specific function $\psi(\epsilon, \tilde{\epsilon})$, and p_{11} satisfies $S_0'(p_{11}) - 2S_0'(p_{12}) + S_0'(p_{22}) = 0$.

Finally, the asymptotic behavior of these graphons as $\tau \to \epsilon^\ell$ depends only on the degree sequence of H. In particular, the cases where H is a triangle and when H is a 2-star are asymptotically the same. This is illustrated in Figure 2. Since $\Delta\tau_{\text{triangle}} \approx 3\epsilon\Delta\tau_2$, the optimizing graphon for the 2-star model with $\epsilon = 0.4$ and $\Delta\tau_2 = 0.002$ should resemble the optimizing graphon for the triangle model with $\epsilon = 0.4$ and $\Delta\tau_{\text{triangle}} = 0.0024$. These optimizing graphons are obtained using the algorithms we developed in [14] *without assuming bipodality*. Numerical estimates indicate that the optimizing graphons are not exactly the same, thanks to $O(\Delta\tau_2^{3/2})$ corrections to $\Delta\tau_{\text{triangle}}$, but are still qualitatively similar.

Funding

This work was supported by the Simons Foundation [grant 327929 to R.K.] and National Science Foundation (NSF) [grants DMS-1208191, DMS-1509088, DMS-1321018, and DMS-1101326].

Acknowledgments

The computational results shown in this work were obtained on the computational facilities in the Texas Super Computing Center (TACC). We gratefully acknowledge this computational support.

Conflict of Interest

I (Richard Kenyon) am an associate editor of IMRN.

Appendix: Proof of Theorem 3.3

Proof. Fix $k \geq 2$ and let

$$\begin{aligned} N(\epsilon,\tilde{\epsilon}) &= 2[S_0(\tilde{\epsilon}) - S_0(\epsilon) - S_0'(\epsilon)(\tilde{\epsilon}-\epsilon)] \\ D(\epsilon,\tilde{\epsilon}) &= \tilde{\epsilon}^k - \epsilon^k - k\epsilon^{k-1}(\tilde{\epsilon}-\epsilon) \end{aligned} \tag{A1}$$

be the numerator and denominator of the function $\psi_k(\epsilon,\tilde{\epsilon}) = N/D$. These definitions make sense for all real values of k, not just for integers. When taking derivatives of N, D, and ψ, we will denote a derivative with respect to the first variable by a dot, and a derivative with respect to the second variable by $'$. That is, $D'(\epsilon,\tilde{\epsilon}) = \partial D/\partial\tilde{\epsilon}$ and $\dot{D}(\epsilon,\tilde{\epsilon}) = \partial D/\partial\epsilon$. As noted earlier, this definition of ψ_k has a removable singularity at $\tilde{\epsilon} = \epsilon$, which we fill in by defining

$$\psi_k(\epsilon,\epsilon) = N''(\epsilon,\epsilon)/D''(\epsilon,\epsilon) = 2S_0''(\epsilon)/[k(k-1)\epsilon^{k-2}]. \tag{A2}$$

The denominator D vanishes only at $\tilde{\epsilon} = \epsilon$.

Some useful explicit derivatives are:

$$N' = 2[S_0'(\tilde{\epsilon}) - S_0'(\epsilon)], \qquad N'' = 2S_0''(\tilde{\epsilon}) = \frac{-1}{\tilde{\epsilon}(1-\tilde{\epsilon})},$$

$$\dot{N} = -2S_0''(\epsilon)(\tilde{\epsilon} - \epsilon), \qquad \dot{N}' = -2S_0''(\epsilon),$$
$$D' = k[\tilde{\epsilon}^{k-1} - \epsilon^{k-1}], \qquad D'' = k(k-1)\tilde{\epsilon}^{k-2},$$
$$\dot{D} = -k(k-1)\epsilon^{k-2}(\tilde{\epsilon} - \epsilon), \qquad \dot{D}' = -k(k-1)\epsilon^{k-2}. \tag{A3}$$

Note that D and N both vanish when $\tilde{\epsilon} = \epsilon$, so we can write

$$N(\epsilon, \tilde{\epsilon}) = \int_{\epsilon}^{\tilde{\epsilon}} N'(\epsilon, x)\mathrm{d}x = \int_{\tilde{\epsilon}}^{\epsilon} \dot{N}(x, \tilde{\epsilon})\mathrm{d}x, \tag{A4}$$

and similarly for $D(\epsilon, \tilde{\epsilon})$.

We proceed in steps:

- Step 1. Analyzing ψ near $\tilde{\epsilon} = \epsilon$ to see that $\psi_k'(\epsilon, \epsilon) = 0$ only when $\epsilon = (k-1)/k$.
- Step 2. Showing that we can never have $\psi_k' = \psi_k'' = 0$.
- Step 3. Showing that the equation $\psi_k'(\epsilon, \tilde{\epsilon})$ is symmetric in ϵ and $\tilde{\epsilon}$, implying that ζ_k is an involution.
- Step 4. Showing that ψ_k has a unique critical point.
- Step 5. Showing that $\mathrm{d}\zeta_k/\mathrm{d}\epsilon$ is never zero.
- Step 6. Showing that $\psi_k(\epsilon, \zeta_k(\epsilon)) > \max(\psi_k(\epsilon, \epsilon), \psi_k(\zeta(\epsilon), \zeta(\epsilon)))$.

The following calculus fact will be used repeatedly. When $D \neq 0$, $\psi_k' = 0$ is equivalent to $N/D = N'/D'$, and $\psi_k' = \psi_k'' = 0$ is equivalent to $N/D = N'/D' = N''/D''$. This follows from the quotient rule:

$$\psi' = \frac{DN' - ND'}{D^2},$$
$$\psi'' = \frac{DN'' - ND''}{D^2} - 2\frac{D'(DN' - ND')}{D^3}. \tag{A5}$$

Step 1.

Since N and D have double roots at $\tilde{\epsilon} = \epsilon$, we can do a Taylor series for both of them near $\tilde{\epsilon} = \epsilon$:

$$\psi_k(\epsilon, \tilde{\epsilon}) = \frac{N''(\epsilon,\epsilon)(\tilde{\epsilon} - \epsilon)^2/2 + N'''(\epsilon,\epsilon)(\tilde{\epsilon} - \epsilon)^3/6 + \cdots}{D''(\epsilon,\epsilon)(\tilde{\epsilon} - \epsilon)^2/2 + D'''(\epsilon,\epsilon)(\tilde{\epsilon} - \epsilon)^3/6 + \cdots}$$
$$= \frac{N''(\epsilon,\epsilon) + N'''(\epsilon,\epsilon)(\tilde{\epsilon} - \epsilon)/3 + \cdots}{D''(\epsilon,\epsilon) + D'''(\epsilon,\epsilon)(\tilde{\epsilon} - \epsilon)/3 + \cdots}. \tag{A6}$$

$\psi_k'(\epsilon, \epsilon) = 0$ is then equivalent to

$$N''(\epsilon,\epsilon)D'''(\epsilon,\epsilon) = N'''(\epsilon,\epsilon)D''(\epsilon,\epsilon)$$

$$\frac{-k(k-1)(k-2)\epsilon^{k-3}}{\epsilon(1-\epsilon)} = \frac{-k(k-1)\epsilon^{k-2}(1-2\epsilon)}{\epsilon^2(1-\epsilon)^2}$$
$$(k-2)(1-\epsilon) = 1-2\epsilon$$
$$k\epsilon = k-1. \tag{A7}$$

Step 2.

If $\psi'_k = \psi''_k = 0$, then we must have $N'D'' = D'N''$ and $ND'' = DN''$. We will explore these in turn. We write

$$0 = N'D'' - D'N'' = \int_{\tilde\epsilon}^{\epsilon} D''(\epsilon,\tilde\epsilon)\dot{N}'(x,\tilde\epsilon) - N''(\epsilon,\tilde\epsilon)\dot{D}'(x,\tilde\epsilon)\mathrm{d}x. \tag{A8}$$

Explicitly, this becomes

$$0 = \int_{\tilde\epsilon}^{\epsilon} \frac{k(k-1)}{\tilde\epsilon(1-\tilde\epsilon)x(1-x)}\left[\tilde\epsilon^{k-1}(1-\tilde\epsilon) - x^{k-1}(1-x)\right]\mathrm{d}x. \tag{A9}$$

The function $x^{k-1}(1-x)$ has a single maximum at $x=(k-1)/k$. If both ϵ and $\tilde\epsilon$ are on the same side of this maximum, then the integrand will have the same sign for all x between $\tilde\epsilon$ and ϵ, and the integral will not be zero. Thus we must have $\epsilon < (k-1)/k < \tilde\epsilon$, or vice-versa, and we must have $\epsilon^{k-1}(1-\epsilon) < \tilde\epsilon^{k-1}(1-\tilde\epsilon)$. In this case the integrand changes sign exactly once.

Now we apply the same sort of analysis to the other equation:

$$0 = ND'' - DN'' = \int_{\tilde\epsilon}^{\epsilon} D''(x,\tilde\epsilon)\dot{N}(x,\tilde\epsilon) - N''(x,\tilde\epsilon)\dot{D}(x,\tilde\epsilon)\mathrm{d}x. \tag{A10}$$

Explicitly, this becomes

$$0 = \int_{\tilde\epsilon}^{\epsilon} \frac{k(k-1)}{\tilde\epsilon(1-\tilde\epsilon)x(1-x)}\left[\tilde\epsilon^{k-1}(1-\tilde\epsilon) - x^{k-1}(1-x)\right](\tilde\epsilon - x)\mathrm{d}x. \tag{A11}$$

This is the same integral as before, only with an extra factor of $(\tilde\epsilon - x)$. If we view the first integral (A9) as a mass distribution (with total mass zero), then the second integral is (minus) the first moment of this mass distribution relative to the endpoint $\tilde\epsilon$. But we have already seen that the distribution changes sign exactly once, and so must have a non-zero first moment. This is a contradiction.

Step 3.

If $ND' = DN'$, then $N/D = N'/D'$. Call this common ration r. Then

$$N = rD \quad \text{and} \quad N' = rD'. \tag{A12}$$

Note that N' and D' are odd under interchange of ϵ and $\tilde{\epsilon}$, so the second equation is invariant under this interchange. Furthermore, we have $(\tilde{\epsilon}-\epsilon)N'-N = r[(\tilde{\epsilon}-\epsilon)D'-D]$. However, $(\tilde{\epsilon}-\epsilon)N'-N$ is the same as N with the roles of ϵ and $\tilde{\epsilon}$ reversed, while $(\tilde{\epsilon}-\epsilon)D'-D$ is the same as D with the roles of ϵ and $\tilde{\epsilon}$ reversed. Thus the two equations are satisfied for $(\epsilon,\tilde{\epsilon})$ if and only if they are satisfied for $(\tilde{\epsilon},\epsilon)$.

Step 4.

For $k=2$ we explicitly compute that $\psi_2'=0$ only at $\tilde{\epsilon}=1-\epsilon$. If $k_{\min}$ is the infimum of all values of k for which ψ_k has multiple critical points, then at a critical point of $\psi_{k_{\min}}$ we must have $\psi_k'=\psi_k''=0$, which is a contradiction. Thus $k_{\min}$ does not exist, and ψ_k has a unique critical point for all $k\geq 2$. In particular, ζ_k is a well-defined function.

Step 5.

The function ζ_k is defined by the condition that $DN'-ND'=0$ (and $\tilde{\epsilon}\neq\epsilon$, except when $\epsilon=(k-1)/k$). Let $f(\tilde{e},e)=DN'-ND'=D^2\psi'$. Moving along the curve $\tilde{\epsilon}=\zeta_k(\epsilon)$ (i.e., $f=0$), we differentiate implicitly:

$$0=df=\dot{f}\,d\epsilon+f'd\tilde{\epsilon}, \tag{A13}$$

so

$$\frac{d\tilde{\epsilon}}{d\epsilon}=\frac{-\dot{f}}{f'}. \tag{A14}$$

We compute $f'=DN''-ND''$. This is non-zero by Step 2. We also have

$$\begin{aligned}\dot{f}&=D\dot{N}'-\dot{N}D'+\dot{D}N'-N\dot{D}'\\&=-2S_0''(\epsilon)(D-(\tilde{\epsilon}-\epsilon)D)'+k(k-1)\epsilon^{k-2}(N-(\tilde{\epsilon}-\epsilon)N')\\&=2S_0''(\epsilon)[\epsilon^k-\tilde{\epsilon}^k+k(\tilde{\epsilon}-\epsilon)\tilde{\epsilon}^{k-1}]-2k(k-1)\epsilon^{k-2}[S_0(\epsilon)-S_0(\tilde{\epsilon})+(\tilde{\epsilon}-\epsilon)S_0'(\tilde{\epsilon})]\\&=D(\tilde{\epsilon},\epsilon)N''(\tilde{\epsilon},\epsilon)-N(\tilde{\epsilon},\epsilon)D''(\tilde{\epsilon},\epsilon).\end{aligned} \tag{A15}$$

That is, $\dot{f}$ is the same as f', only with the roles of ϵ and $\tilde{\epsilon}$ reversed. Since the equation $f=0$ is symmetric in ϵ and $\tilde{\epsilon}$, the argument of Step 2 can be repeated to show that $\dot{f}\neq 0$.

Since $d\tilde{\epsilon}/d\epsilon$ is never zero, and since $d\tilde{\epsilon}/d\epsilon=-1$ at the fixed point (by symmetry), $\zeta_k'(\epsilon)=d\tilde{\epsilon}/d\epsilon$ must always be negative.

Step 6.

Since $\psi_k(\epsilon, \tilde{\epsilon})$ has a single critical point (with respect to $\tilde{\epsilon}$, for fixed k and ϵ), this critical point must either always be a local maximum or a local minimum, and hence a global maximum or minimum, and the answer must be the same for all k and all ϵ. By checking a single case (e.g., $k = 2$ and ϵ approaching 0) it is easy to see that it is a maximum. Thus $\psi_k(\epsilon, \zeta_k(\epsilon)) > \psi_k(\epsilon, \epsilon)$ for all $\epsilon \neq (k-1)/k$. Since the equations for a critical point are symmetric with respect to interchange of ϵ and $\tilde{\epsilon}$, $\epsilon = \zeta_k(\tilde{\epsilon})$ also gives the unique critical point of $\psi_k(\epsilon, \tilde{\epsilon})$ with respect to ϵ. By considering the limit of $\psi_k(\epsilon, \tilde{\epsilon})$ as $\epsilon \to 0$ or $\epsilon \to 1$, it is clear that this critical point is a maximum. Since $\zeta_k(\zeta_k(\epsilon)) = \epsilon$, this implies that $\psi_k(\epsilon, \zeta_k(\epsilon)) > \psi_k(\zeta_k(\epsilon), \zeta_k(\epsilon))$. ■

References

[1] Aristoff, D. and C. Radin. "Emergent structures in large networks." *Journal of Applied Probability* 50 (2013): 883–8.

[2] Borgs, C., J. Chayes, and L. Lovász. "Moments of two-variable functions and the uniqueness of graph limits." *Geometry and Functional Analysis* 19 (2010): 1597–619.

[3] Borgs, C., J. Chayes, L. Lovász, V.T. Sós, and K. Vesztergombi. "Convergent graph sequences I: subgraph frequencies, metric properties, and testing." *Advances in Mathematics* 219 (2008): 1801–51.

[4] Chatterjee, S., and P. Diaconis. "Estimating and understanding exponential random graph models." *Annals of Statistics* 41 (2013): 2428–61.

[5] Chatterjee, S., and S. R. S. Varadhan. "The large deviation principle for the Erdős-Rényi random graph." *European Journal of Combinatorics* 32 (2011): 1000–17.

[6] Kenyon, R., C. Radin, K. Ren, and L. Sadun. "Multipodal structures and phase transitions in large constrained graphs." arXiv:1405.0599, (2014).

[7] Lovász, L., and B. Szegedy. "Limits of dense graph sequences." *Journal of Combinatorial Theory Series B* 98 (2006): 933–57.

[8] Lovász, L., and B. Szegedy. "Szemerédi's lemma for the analyst." *Geometry and Functional Analysis* 17 (2007): 252–70.

[9] Lovász, L. and B. Szegedy. "Finitely forcible graphons." *Journal of Combinatorial Theory Series B* 101 (2011): 269–301.

[10] Lovász, L. *Large Networks and Graph Limits.* Providence: American Mathematical Society, 2012.

[11] Lubetzky, E., and Y. Zhao. "On replica symmetry of large deviations in random graphs." *Random Structures and Algorithms* 47 (2015): 109–46.

[12] Pikhurko, O., and A. Razborov. "Asymptotic structure of graphs with the minimum number of triangles." *Combinatorics, Probability and Computing* (2016): 1–23. ISSN 0963-5483 (In Press).

[13] Radin, C. and M. Yin. “Phase transitions in exponential random graphs.” *Annals of Applied Probability* 23 (2013): 2458–71.

[14] Radin, C., K. Ren, and L. Sadun. “The asymptotics of large constrained graphs.” *Journal of Physics A: Mathematical and Theoretical* 47 (2014): 175001.

[15] Radin, C., and L. Sadun. “Phase transitions in a complex network.” *Journal of Physics A: Mathematical and Theoretical* 46 (2013): 305002.

[16] Radin, C., and L. Sadun. “Singularities in the entropy of asymptotically large simple graphs.” *Journal of Statistical Physics* 158 (2015): 853–65.

[17] Razborov, A. “On the minimal density of triangles in graphs.” *Combinatorics, Probability and Computing* 17 (2008): 603–18.

[18] Turán, P. “On an extremal problem in graph theory, (in Hungarian).” *Matematikai é s Fizikai Lapok* 48 (1941): 436–52.