Journal of Physics A:
Mathematical and Theoretical

PAPER • OPEN ACCESS

Existence of a symmetric bipodal phase in the edge-triangle model

To cite this article: Joe Neeman *et al* 2024 *J. Phys. A: Math. Theor.* **57** 095003

OPEN ACCESS
IOP Publishing
Journal of Physics A: Mathematical and Theoretical

J. Phys. A: Math. Theor. **57** (2024) 095003 (30pp) https://doi.org/10.1088/1751-8121/ad259d

Existence of a symmetric bipodal phase in the edge-triangle model

Joe Neeman, Charles Radin and Lorenzo Sadun*

Department of Mathematics, The University of Texas at Austin, Austin, TX 78712, United States of America

E-mail: sadun@math.utexas.edu

Received 24 October 2023; revised 4 January 2024
Accepted for publication 1 February 2024
Published 16 February 2024

Abstract

In the edge-triangle model with edge density close to 1/2 and triangle density below 1/8 we prove that the unique entropy-maximizing graphon is symmetric bipodal. We also prove that, for any edge density e less than $e_0 = (3-\sqrt{3})/6 \approx 0.2113$ and triangle density slightly less than e^3, the entropy-maximizing graphon is not symmetric bipodal. We also discuss the implications for an old idea of Landau for using symmetry to give an intrinsic difference between solid and fluid phases of matter.

Keywords: BIPODAL, phase transition, random graph, graphon, symmetry, microcanonical, edge-triangle

1. Introduction and results

In this paper, we analyze emergent *smoothness with respect to change of competing constraints* in asymptotically large dense random graphs. More specifically, we determine and study smooth phases separated by sharp transitions. Our definitions of phases and their transitions are based on those of equilibrium statistical mechanics and the relation of our results to transitions in statistical mechanics is given in detail in sections 7 and 8. We derive a new phase in the model with sharp constraints on edge and triangle densities. The phase is called 'symmetric bipodal' and we show how to use its symmetry to distinguish the phase *intrinsically* from other phases. Unlike in previous work, graphs in this phase are not small perturbations of Erdős-Rényi graphs; this requires new techniques, which we develop. (The combinatorial terminology will be clarified in section 1.2.)

* Author to whom any correspondence should be addressed.

1.1. Physics motivation

There is an important open problem [35, p 11] in the physics of condensed matter based on the experimental fact that for **all** known materials if one varies thermodynamical variables quasistatically one cannot move between the solid and fluid phases without encountering a sharp phase transition. (This is in contrast to variation between the gas and liquid states where there is a critical point that one can go around.) An argument to explain this universal phenomenon was given by Landau [2, p 19] based on the ordered nature of the solid phase; he argued that one should not be able to move smoothly between an ordered and a disordered phase. This argument is controversial [23, p 122]; it has neither been given a firm foundation nor been refuted, nor has a substitute argument been given. One obstacle in producing an argument is that no statistical mechanics model, using short range forces between particles in Euclidean space $\mathbb{R}^d$, has ever been proven to exhibit both fluid and solid phases [5, 35]. (This has been simulated repeatedly, for instance for Lennard-Jones and hard-sphere forces, but without a proof it is hard to determine an obstructing mechanism.)

In this paper we work in a nonphysical, combinatorial framework of graphons (discussed below), a limit of large dense random graphs. While this framework is simpler than statistical mechanics, we can reproduce much of the rich structure of statistical mechanics to prove the existence of a variety of phases. Within this framework we produce a version of Landau's argument that is effective; it enables us to distinguish portions of phases that cannot be connected smoothly because of an intrinsic difference of symmetry. Of course this does not solve the original statistical mechanical problem, but it does allow for a more careful analysis of Landau's argument.

1.2. The combinatorial setting

The remainder of this section 1 is devoted to the statement and the motivation of our combinatorial results, and the relation of our setting to statistical mechanics. Terms such as 'graphon', 'bipodal' and 'symmetric bipodal' are given careful definitions in section 2.

In simple models in statistical mechanics one can use a *microcanonical ensemble*, a two-parameter family of probability distributions on the set of possible states of the particles, in which the total number of particles and total energy are fixed and all constrained states are equally likely. Alternatively one can use a *grand canonical ensemble*, where the energy and particle number are both allowed to float, with parameters/weights being simple combinations of the temperature and chemical potential. (There are also other ensembles in use, sometimes selected by experimental conditions.) If one is concerned with bulk properties of macroscopic systems, and therefore the limit of an infinite statistical mechanics system, all ensembles are equivalent: by varying the parameters in any ensemble one can obtain the desired state in another ensemble, except for states which correspond to coexisting phases. In theoretical work most research focuses on grand canonical ensembles, which are relatively easy to handle technically.

The situation is very different with random graphs, where in this Paper our key variables are the number of edges in a graph (analogous to the number of particles) and the number of triangles (analogous to a total energy). Early studies on graphs with these constraints on a fixed finite number of vertices go back at least to Strauss [33] from 1986, and have been widely used to model data; see [22] and references therein. They focused on analogues of grand canonical ensembles called exponential random graph models, or ERGMs. Taking a limit to infinite graphs with such models goes back at least to Park and Newman [24] in 2005 in the physics literature. The *mathematical* understanding of the infinite size limit for constrained random graphs changed with the development of graphons by Lovász *et al* around 2006–2010 [3, 4, 14–17] (following the work of Aldous [1] and Hoover [9]), the large deviation principle for Erdős-Rényi graphs [7] by Chatterjee and Varadhan (2010) and the followup paper [6] by Chatterjee and Diaconis (2011). At this point, it was discovered [6, 7, 29, 30] that *most* achievable simultaneous values (e,t) of edge and triangle densities cannot be accessed with *any* values of the conjugate variables. In the limit of infinite size graphs, the only graphs that appear in ERGMs are either Erdős-Rényi graphs for which $t = e^3$ or graphs with t much smaller than e^3. If one's object is to study asymptotic constrained graphs with $t > e^3$, or graphs with t moderately less than e^3, then ERGMs simply do not work. Instead, one should use the analogue of the microcanonical ensemble, which is by definition all-inclusive, a project which started in 2012 [29, 30].

This, then, is what we do here, using the machinery of graphons to study the *global* structure of large constrained graphs. A graphon $g(x,y)$ is a measurable symmetric function on the unit square $[0,1]^2$ taking values between 0 and 1, and can be viewed as an infinite-vertex limit of adjacency matrices. (Instead of attaching integer labels 1, 2, ... to vertices, we label them with a continuum of real numbers from 0 to 1.) If the function g is continuous at (x,y), then $g(x,y)$ is the density of edges between vertices with label near x and vertices with label near y. Put another way, if we sample two random vertices, with labels x and y, then $g(x,y)$ is the probability that they are connected by an edge. Given three vertices with indices x, y, and z, the probability that they form a triangle is $g(x,y)g(y,z)g(z,x)$.

In fact, graphons can be viewed as a process for generating random graphs as well as a limit of such graphs. To get a graph with n vertices from a graphon $g(x,y)$, pick n labels $x_1,\ldots,x_n$ independently and uniformly from the interval $[0,1]$. Then flip $\binom{n}{2}$ independent biased coins, placing an edge between vertices i and j with probability $g(x_i,x_j)$. The expected density of edges and triangles in the resulting graph is then given by the quantities

$$\varepsilon(g) = \iint g(x,y)\,\mathrm{d}x\,\mathrm{d}y, \text{and } \tau(g) = \iiint g(x,y)\,g(y,z)\,g(z,x)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z. \tag{1}$$

We call $\varepsilon(g)$ and $\tau(g)$ the edge and triangle densities of the graphon g.

There are $\binom{n}{2} \approx n^2/2$ independent coin flips, whose total expected entropy is $n^2 + O(n)$ times

$$S(g) = \iint H[g(x,y)]\,\mathrm{d}x\,\mathrm{d}y, \text{where } H(u) = -\frac{1}{2}\left[u\ln(u) + (1-u)\ln(1-u)\right]. \tag{2}$$

We refer to $S(g)$ as the '(Shannon) entropy of the graphon g,' distinct but related to the Boltzmann entropy defined in the next paragraph. The quantities $\varepsilon(g)$, $\tau(g)$ and $S(g)$ are all invariant under measure-preserving transformations of $[0,1]$. All statements about uniqueness of graphons should be understood to mean 'unique up to measure-preserving transformations of $[0,1]$.'

Given an achievable ordered pair (e,t) of edge and triangle densities and a tolerance δ, we define the Boltzmann entropy in terms of the partition function $Z^{n,\delta}_{e,t}$, which is the number of labeled graphs on n nodes with edge density within δ of e and triangle density within δ of t. From this partition function we define the Boltzmann entropy [29] as

$$B(e,t) = \lim_{\delta\searrow 0}\lim_{n\to\infty}\frac{1}{n^2}\ln\left[Z^{n,\delta}_{e,t}\right]. \tag{3}$$

Using [7] we established [29] the variational formula

$$B(e,t) = \sup_{g \in \mathcal{G}_{e,t}} S(g), \tag{4}$$

where

$$\mathcal{G}_{e,t} = \{g \mid \varepsilon(g) = e, \tau(g) = t\}. \tag{5}$$

If the supremum of $S(g)$ is attained by a unique graphon $g_{e,t} \in \mathcal{G}(e,t)$, then all but exponentially few large graphs with the constrained edge and triangle densities are described by $g_{e,t}$ [31]. This follows from theorem 3.1 in [29], generalized to more general models in later works; its significance for finite graphs will be discussed in section 2.

That is, the complicated problem of counting and characterizing large random *graphs* with constraints on the edge and triangle densities boils down to characterizing S-maximizing *graphons* with constraints on $\varepsilon(g)$ and $\tau(g)$.

1.3. Results

Our first major result is:

Theorem 1. *There is an open subset $\mathcal{O}$ of the (e, t) plane, containing the interval $e = \frac{1}{2}$, $0 < t < \frac{1}{8}$, on which the unique S-maximizing graphon is 'bipodal':*

$$g_{e,t}(x,y) = \begin{cases} e - \left(e^3 - t\right)^{1/3}, & 0 < x,y < \frac{1}{2} \text{ or } \frac{1}{2} < x,y < 1, \\ e + \left(e^3 - t\right)^{1/3}, & x < \frac{1}{2} < y \text{ or } y < \frac{1}{2} < x. \end{cases} \tag{6}$$

(The terms bipodal and symmetric bipodal are fully defined with other notation in section 2, and are not necessary for a full understanding of the results in this section, which only requires equation (6)).

See figures 1 and 2. The following is immediate by inspection.

Corollary 2. *$B(e,t)$ and the densities of all subgraphs are real analytic functions of (e, t) in $\mathcal{O}$.*

Our second major result describes a region where the optimizing graphon is not symmetric bipodal.

Theorem 3. *Let $e_0 = (3 - \sqrt{3})/6 \approx 0.2113$. For any fixed edge density $e < e_0$ and any sufficiently small positive σ, the symmetric bipodal graphon (6) does not maximize S among graphons with triangle density $t = e^3 - \sigma^3$.*

1.4. Technical background

The study of graphs with competing constraints is an old topic in extremal combinatorics. For graphs, the range of achievable values of the pair (e,t), and the graphs that achieve them, was completed in 2012 by Purkurko and Razborov [25]: see figure 3 for a distorted view of the 'Razborov triangle'.

We define a *phase* in our model as a maximal connected open subset of the Razborov triangle on which $B(e,t)$, and the density of every fixed subgraph (e.g. the density of squares, pentagons, tetrahedra, ...) of a typical graph, is an analytic function of (e,t). The system is said to have a *phase transition* wherever such quantities are not analytic. Usually this occurs

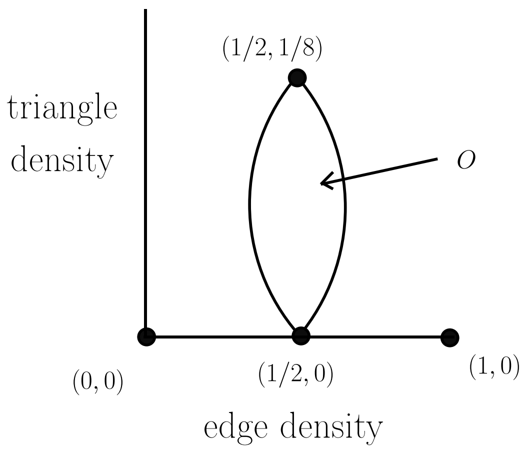

Figure 1. The open set $\mathcal{O}$ of theorem 1.

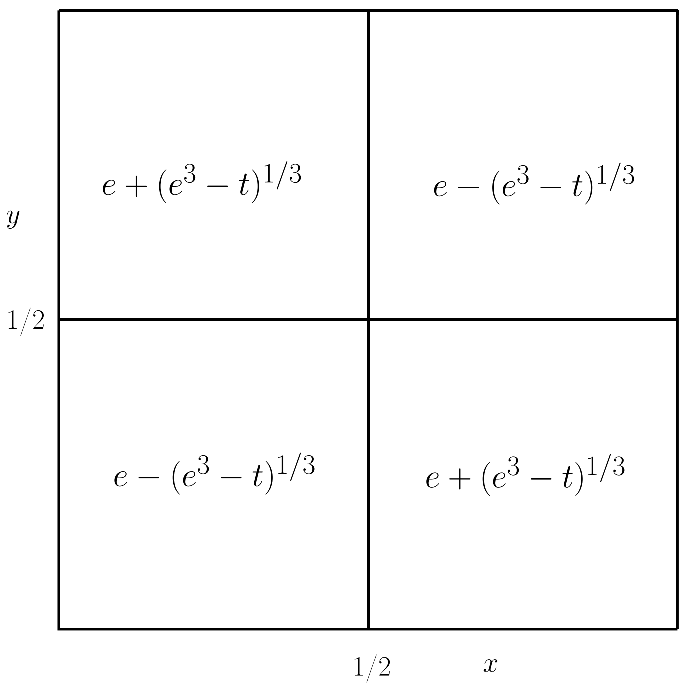

Figure 2. The piecewise-constant optimal graphon of theorem 1.

at the boundary of two or more phases, but sometimes there is an analytic path from one 'side' of a phase transition to another, as discussed in section 7. Note that if the optimal graphon in the variational formula (4) is not unique, the densities of some subgraphs are not even well-defined, much less analytic; by definition, such points of non-uniqueness can never lie within a phase. Even when uniqueness does hold, it can be difficult to prove. On the other hand, where there is a unique entropy-optimizing graphon $g_{e,t}$, all but exponentially few large graphs are

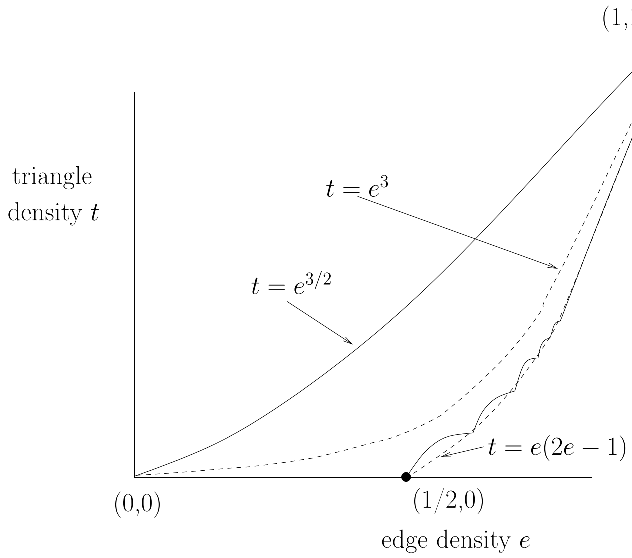

Figure 3. The Razborov triangle, outlined in solid curves.

close (in the so-called *cut metric*) to $g_{e,t}$; this facilitates the analysis of emergent features. (Such uniqueness is known to fail in similar models; see section 5 in [11].)

Let G be a fixed subgraph, such as a square or pentagon or tetrahedron. When $S(g)$ achieves the value $B(e,t)$ at a unique graphon, all but exponentially few graphs have the same density t_G of G, so we can speak of t_G being a function of (e,t) and ask whether that function is analytic. (Our method was based on large deviations of $G(n,p)$ [7]; for an approach based on $G(n,m)$ see [8].)

In 2015 we established [11] the existence of two open subsets of the Razborov triangle, both with $t > e^3$, in which $B(e,t)$ and all subgraph densities were real-analytic functions of (e,t). To prove this we proved that the constrained S-optimizing graphons were unique, and determined that they have a 2-block ('bipodal') [11] structure whose parameters were analytic in (e,t). More recently, we proved [21] a complementary result in the more difficult case of undersaturated triangles ($t < e^3$) when $e > \frac{1}{2}$. This yielded the satisfying result of a pair of open sets, $\mathcal{O}_1, \mathcal{O}_2$, on which the Boltzmann entropy $B(e,t)$ and all subgraph densities are analytic, separated by the bounding curve $t = e^3$, on which the entropy is not even differentiable [30]. See figure 4.

Put another way, in [21] we proved the existence of two phases with $e > \frac{1}{2}$, one just above the curve $t = e^3$ and one just below, with a phase transition on the curve itself. As noted earlier, within each phase there must be a unique S-optimizing graphon for each (e,t). This optimal graphon is called the 'state' of the system. As discussed above, models of dense graphs with sharp competing constraints are a natural extension of extremal graph theory. We note there have been parallel studies within other parts of extremal combinatorics, for instance permutations and sphere packings; see section 8.

One may reasonably ask how we know that the two open sets $\mathcal{O}_1$ and $\mathcal{O}_2$ actually belong to different phases; how can we rule out the possibility that there is an analytic path between them, going *around* the phase transition? Neither the $\mathcal{O}_1$ phase nor the $\mathcal{O}_2$ phase seems to have any intrinsic property that clearly rules out an analytic continuation between them. Such an analytic continuation does not actually exist, since we previously showed [30] that $B(e,t)$

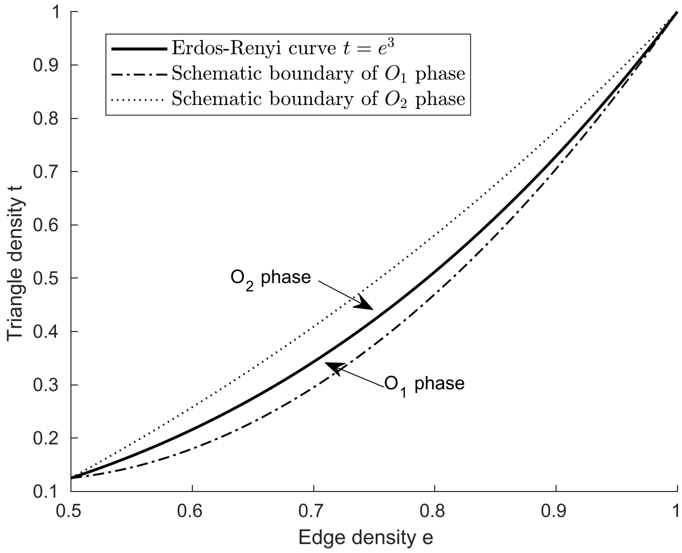

Figure 4. The phase transition between $\mathcal{O}_1$ and $\mathcal{O}_2$ was proven in [21].

cannot be differentiable at any point on the curve $t = e^3$ for any $0 < e < 1$. However that is not a very satisfying explanation.

The 'symmetric bipodal' phase whose existence we prove in this paper is different. The symmetry provides an *intrinsic* difference between the new phase and the $\mathcal{O}_1$ and $\mathcal{O}_2$ phases. In section 7 we discuss the connection between our symmetry argument and the use of 'order parameters' in equilibrium statistical mechanics, and as a by-product we clarify a problematic argument of Landau from the 1950s.

Another key difference between our new phase and the previously proven phases is that proof of the symmetric bipodal phase is not limited to a small neighborhood of the curve $t = e^3$. In [11], and again in [21], we studied small perturbations of the Erdős-Rényi graph $G(n,p)$ and attempted to get the greatest possible change in triangle count for the smallest possible entropy cost. The results are closely related to moderate deviation estimates [20]; depending on the sizes of n and $e^3 - t$, a finite graph with triangle density slightly less than e^3 can be viewed either as a typical (e,t) graph or as a deviation of an Erdős-Rényi graph. When $e > 1/2$, moderate deviations estimates that apply when $n^{-1} \ll e^3 - t \ll 1$ agree to leading order with large deviations estimates.

This difference is reflected in the different method of proof. In theorem 3, σ is not a small parameter. We can still do a power series expansion in σ, but we have to estimate all terms, not just the first few.

2. Precise statement of results

A graphon is said to be *bipodal* if it is equivalent to a graphon with the block structure shown in figure 5. It is *symmetric bipodal* if $c = 1/2$ and $a = b$. To avoid questions about graphons being equivalent under measure-preserving transformations of $[0,1]$, we restate the definition using arbitrary measurable subsets I_1 and I_2 of $[0,1]$, rather than intervals $[0,c]$ and $(c,1]$. In this description, a graphon is bipodal if there exist complementary measurable subsets I_1 and I_2 such that $g(x,y)$ is constant on $I_1 \times I_1$, constant on $I_2 \times I_2$, and constant on $I_1 \times I_2 \cup I_2 \times I_1$.

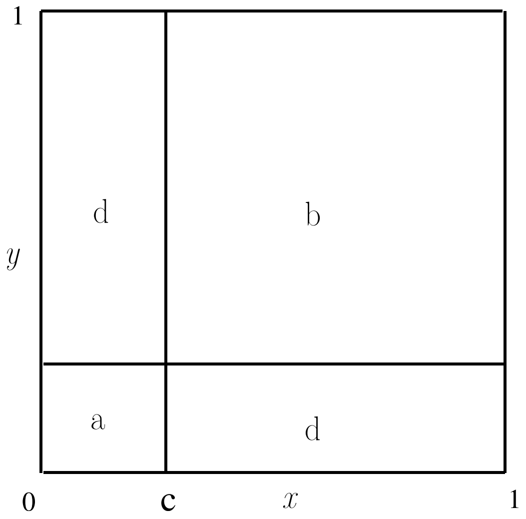

Figure 5. The parameters of a bipodal graphon.

It is symmetric bipodal if there exist complementary subsets I_1 and I_2, each of measure 1/2, and a positive number $\sigma \leqslant \min(e, 1-e)$, such that the graphon is

$$g(x,y) = \begin{cases} e-\sigma & (x,y) \in I_1 \times I_1 \cup I_2 \times I_2, \\ e+\sigma & (x,y) \in I_1 \times I_2 \cup I_2 \times I_1, \end{cases} \tag{7}$$

The edge density, triangle density and entropy of a symmetric bipodal graphon are

$$\varepsilon(g) = e, \qquad \tau(g) = e^3 - \sigma^3, \qquad S(g) = \frac{1}{2}(H(e+\sigma) + H(e-\sigma)). \tag{8}$$

The simple formula for $\tau(g)$ comes from the fact that the eigenvalues of g (viewed as an integral operator on $L^2([0,1])$ are e and $-\sigma$, and that $\tau(g)$ is the trace of the cube of that operator. Another characterization of a symmetric bipodal graphon is that it is a rank-1 perturbation of a constant graphon, with

$$g(x,y) = e - \sigma v_1(x) v_1(y),$$

where $\int_0^1 v_1(x)\mathrm{d}x = 0$ and $v_1(x)^2 = 1$ everywhere.

It was previously known that the unique optimal graphon on the open line interval $e = 1/2$, $t \in (0, 1/8)$ was symmetric bipodal [30]. Our main result, theorem 1, extends this to an open set $\mathcal{O}$ containing the line interval. It is convenient for our proofs to reformulate theorem 1 as

Theorem 4. *For fixed $\sigma \in \left(0, \frac{1}{2}\right)$ and for all sufficiently small δ (of either sign), the unique S-maximizing graphon with edge density $e = \frac{1}{2} + \delta$ and triangle density $t = e^3 - \sigma^3$ is symmetric bipodal. Furthermore, the size of the allowed interval of δ's varies continuously with σ.*

Corollary 5. *The Boltzmann entropy $B(e,t)$ and the densities of all subgraphs are real analytic functions of (e, t) in the open set thus defined.*

The region where the optimal graphon is symmetric bipodal is not limited to the small open set described in theorems 1 and 4. There is considerable numerical evidence that this region, called the A(2,0) phase in [13], is much bigger than that. However, there are provable limits to its extent. Theorem 3 says that it does not extend to the curve $t = e^3$ when $e < e_0 \approx 0.2113$. Theorem 1 from [21], which we restate here, says that it does not extend to the curve $t = e^3$ when $e > 1/2$. It is an open question whether the phase extends to the curve $t = e^3$ when $e_0 < e < 1/2$.

Theorem 6. *(Theorem 1 from [21]) There is an open subset $\mathcal{O}_1$ in the planar set of achievable parameters (e, t), whose upper boundary is the curve $t = e^3$, $1/2 < e < 1$, such that at (e, t) in $\mathcal{O}_1$ there is a unique entropy-optimizing graphon $g_{e,t}$. This graphon is bipodal and for fixed (e, t), the values of a, b, c, d can be approximated to arbitrary accuracy via an explicit iterative scheme. These parameters can also be expressed via asymptotic power series in $\sigma = (e^3 - t)^{1/3}$ whose leading terms are:*

$$\begin{aligned}
a &= 1 - e - \sigma + O(\sigma^2) \\
b &= e - \frac{\sigma^2}{2e-1} + O(\sigma^3) \\
c &= \frac{\sigma}{2e-1} - \frac{2\sigma^2}{2e-1} + O(\sigma^3) \\
d &= e + \sigma + \frac{\sigma^2}{eH'(e)}\left(H'(e) - \left(e - \frac{1}{2}\right)H''(e)\right) + O(\sigma^3).
\end{aligned} \tag{9}$$

Corollary 7. *$B(e,t)$ and the densities of all subgraphs are real analytic functions of (e, t) in the open set $\mathcal{O}_1$.*

This corollary was proven in the last paragraph of the proof of theorem 1 in [21], although not included in the statement of the theorem.

Theorems 4 and 6 and corollaries 5 and 7 are statements about entropy-maximizing *graphons*. So what do they tell us about actual random *graphs*?

Theorem 4 implies that the ensemble of all large graphs with edge density e close to 1/2 and triangle density t less than e^3 is dominated by graphs that have a block structure. In such graphs, roughly half the vertices are of one type (call them red) and roughly half are of the other type (blue). Two red vertices, or two blue vertices, are connected a fraction $e - \sigma$ of the time. A random red vertex and a random blue vertex are connected a fraction $e + \sigma$ of the time. From this description, the densities of all subgraphs (e.g. squares or pentagons or tetrahedra) can easily be computed. Theorem 6 is similar, except that in the region $\mathcal{O}_1$ only an $O(\sigma)$ fraction of the vertices are red and the probability a of a red-red edge is different from the probability b of a blue-blue edge.

It is important to note that **these are not block models**! The ensemble we are describing is that of *all* large graphs with the specified edge and triangle densities. We are not assigning colors to vertices or telling the edges to appear in some places but not in others. Rather, we prove that block structure emerges *on its own* in all but an exponentially small fraction of those graphs. The fact that such self-organization is a consequence of merely constraining the edge and triangle densities is truly remarkable.

To use a physical analogy where order appears but is not imposed, consider lowering the energy (or temperature) of a physical material. This often results in most of the particles arranging themselves (at least approximately) in a lattice. A continuum model in which this happens on its own is very different from a lattice model where this behavior is assumed from the start.

The bulk of this paper is devoted to proving theorem 4, which is tantamount to proving theorem 1. To explain the steps, we need some more notation. We diagonalize $g(x,y)-e$, viewed as an integral operator, and write

$$g(x,y) = e + \sum_{j=1}^{\infty} \lambda_j v_j(x) v_j(y), \tag{10}$$

where $|\lambda_1| \geqslant |\lambda_2| \geqslant \cdots$ and the functions $v_1, v_2, \cdots$ are orthonormal in $L^2([0,1])$. Let

$$g_1(x,y) = \lambda_1 v_1(x) v_1(y), \qquad g_2(x,y) = \sum_{j=2}^{\infty} \lambda_j v_j(x) v_j(y). \tag{11}$$

Our goal is to show that $g_2 = 0$ and that $v_1(x) = \pm 1$, taking each value on a set of measure 1/2. We do this in stages:

- In section 3, we prove *a priori* entropy bounds on any graphon having the given values of (e,t). We show that the symmetric bipodal graphon comes within $O(\delta^2)$ of saturating those bounds. This implies that any entropy-maximizing graphon must be L^2-close to a symmetric bipodal graphon. Specifically, g_2 must be L^2-small and v_1 must be L^2-close to the desired step function.
- In section 4 we show that g_2 is *pointwise* small and that $v_1(x)^2$ is pointwise close to 1. More precisely, the L^∞ norms of g_2 and $v_1^2 - 1$ must go to zero as $\delta \to 0$.
- In section 5 we expand the entropy $S(g)$ using a convergent Taylor series for $H(u)$ around $u = \frac{1}{2}$. Using the fact that g_2 is pointwise small, we express the difference between $S(g)$ and the entropy of a symmetric bipodal graphon as a quadratic function of the L^2 norm of g_2, the L^2-norm of $v_1(x)^2 - 1$ and the integral $\int_0^1 v_1(x)\mathrm{d}x$, plus higher-order corrections. We show that the quadratic function is negative-definite, implying that g_2 must be zero, $v_1(x)^2$ must be 1, and $\int_0^1 v_1(x)\mathrm{d}x$ must be zero. In other words, our graphon must be symmetric bipodal.
- In section 6 we turn our attention to theorem 3. We construct a family of tripodal graphons and we express the entropy of both this tripodal graphon and the symmetric bipodal graphon as power series in σ. When $e < e_0$, we can choose the parameters of the tripodal graphon such that the tripodal graphon has more entropy at order σ^2 than the symmetric bipodal graphon. This does *not* prove that the optimal graphon is tripodal! However, it does proves that, for σ sufficiently small, the symmetric bipodal graphon is not optimal.

We use big-O and little-o notation throughout. When we say that a certain quantity is $O(\delta^n)$, we mean that there exist positive numbers C and δ_0 (which may depend on σ) such that our quantity is bounded by $C|\delta|^n$ whenever $|\delta| < \delta_0$. When we say that a quantity is $o(\delta^n)$, we mean that there exists a constant δ_0 and function $f(\delta)$, going to zero as $\delta \to 0$, such that the quantity is bounded by $f(\delta)|\delta^n|$ when $|\delta| < \delta_0$.

3. *A priori* estimates

We begin with an upper bound on entropy.

Theorem 8. *If g is a graphon with edge density $e = \frac{1}{2} + \delta$ and triangle density $e^3 - \sigma^3$, with $\sigma > 0$, then*

$$S(g) \leqslant H\left(\frac{1}{2} + \sqrt{\delta^2 + \sigma^2}\right). \tag{12}$$

Proof. Let g be our arbitrary graphon, which we expand as in equations (10) and (11). For $i=1,2$, let $d_i(x)=\int_0^1 g_i(x,y)\mathrm{d}y$. A direct computation of the triangle density gives

$$\begin{aligned}\tau(g) &= \iiint g(x,y)\,g(y,z)\,g(z,x)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\\ &= e^3+\sum_{j=1}^{\infty}\lambda_j^3+3e\int_0^1\left(d_1(x)^2+d_2(x)^2\right)\mathrm{d}x,\end{aligned}$$

$$\sum_j \lambda_j^3 = -\left(\sigma^3+3e\int_0^1\left(d_1(x)^2+d_2(x)^2\right)\mathrm{d}x\right)\leqslant -\sigma^3. \tag{13}$$

Note that $\iint g(x,y)-e\,\mathrm{d}x\,\mathrm{d}y=0$, so there is term in the expansion of $\tau(g)$ that is linear in $g-e$,

The squared L^2 norm of g_1+g_2 is

$$\sum_j \lambda_j^2 \geqslant \left|\sum_j \lambda_j^3\right|^{2/3} \geqslant \sigma^2,$$

with equality if and only if $\lambda_1=-\sigma$, $\lambda_2=\lambda_3=\cdots=0$, and $d_1(x)=0$ for all x.

Next we maximize the entropy for a fixed $\|g_1+g_2\|_2^2$. We use an absolutely convergent power series for

$$H(u)=-\frac{1}{2}\left(u\ln(u)+(1-u)\ln(1-u)\right), \tag{14}$$

namely

$$H(u)=\sum_{n=0}^{\infty}\frac{H^{(n)}\left(\frac{1}{2}\right)}{n!}\left(u-\frac{1}{2}\right)^n. \tag{15}$$

The terms with n odd are identically zero, while the terms with n nonzero and even are negative. As a result,

$$S(g)=H\left(\frac{1}{2}\right)+\sum_{k=1}^{\infty}\frac{H^{(2k)}\left(\frac{1}{2}\right)}{(2k)!}\mu_{2k}, \tag{16}$$

where

$$\mu_{2k}=\iint\left(\delta+g_1(x,y)+g_2(x,y)\right)^{2k}\mathrm{d}x\,\mathrm{d}y. \tag{17}$$

The second moment depends only on the size of g_1 and g_2. Since $\iint g_1(x,y)+g_2(x,y)\,\mathrm{d}x\,\mathrm{d}y=0$, there are no cross terms between δ and g_1+g_2, leaving us with

$$\mu_2=\delta^2+\|g_1+g_2\|_2^2=\delta^2+\sum_{j=1}^{\infty}\lambda_j^2.$$

Maximizing $S(g)$ is equivalent to minimizing all of the higher moments μ_{2k} with $k>1$. Note that μ_{2k} is the kth moment of $(\delta+g_1(x,y)+g_2(x,y))^2$. By the concavity of the function x^n

when $n>1$, μ_{2k} is minimized when $(\delta+g_1(x,y)+g_2(x,y))^2$ is constant, equaling μ_2. In that case, g is everywhere equal to $\frac{1}{2}\pm\sqrt{\mu_2}$ and

$$S(g)=H\left(\frac{1}{2}+\sqrt{\delta^2+\|g_1+g_2\|_2^2}\right). \tag{18}$$

Since $\|g_1+g_2\|_2^2\geqslant\sigma^2$, and since $H(u)$ is a decreasing function of u for $u>1/2$, we conclude that

$$S(g)\leqslant H\left(\frac{1}{2}+\sqrt{\delta^2+\sigma^2}\right).$$

□

Corollary 9. *If g is an entropy-maximizing graphon, then $\int_0^1 v_1(x)\,\mathrm{d}x$ is $O(\delta)$, while $\|g_2\|_2^2$, $\int_0^1 d_2(x)^2\,\mathrm{d}x$ and $\int_0^1(v_1(x)^2-1)^2\,\mathrm{d}x$ are $O(\delta^2)$.*

Proof. The symmetric bipodal graphon has entropy

$$\begin{aligned}\frac{1}{2}\left[H(e+\sigma)+H(e-\sigma)\right] &= \frac{1}{2}\left[H\left(\frac{1}{2}+\delta+\sigma\right)+H\left(\frac{1}{2}+\delta-\sigma\right)\right]\\ &=\frac{1}{2}\left[H\left(\frac{1}{2}+\sigma+\delta\right)+H\left(\frac{1}{2}+\sigma-\delta\right)\right]\\ &=H\left(\frac{1}{2}+\sigma\right)+\frac{\delta^2}{2}H''\left(\frac{1}{2}+\sigma\right)+O\left(\delta^4\right).\end{aligned}$$

By contrast, the upper bound (12) is

$$H\left(\frac{1}{2}+\sigma\right)+\frac{\delta^2}{2\sigma}H'\left(\frac{1}{2}+\sigma\right)+O\left(\delta^4\right).$$

Since the symmetric bipodal graphon has an entropy within $O(\delta^2)$ of the upper bound (12), the entropy-maximizing graphon must also have an entropy within $O(\delta^2)$ of that bound. In particular, $\|g_1+g_2\|_2^2$ must be within $O(\delta^2)$ of σ^2 and the fourth moment μ_4 can be no more than $O(\delta^2)$ greater than $(\delta^2+\sigma^2)^2=\sigma^4+O(\delta^2)$.

Now

$$\|g_1+g_2\|_2^2=\sum_j\lambda_j^2\geqslant\left|\sum_j\lambda_j^3\right|^{2/3}=\left(\sigma^3+3e\int_0^1 d_1(x)^2+d_2(x)^2\,\mathrm{d}x\right)^{2/3}.$$

This can only be within $O(\delta^2)$ of σ^2 if $\int_0^1 d_1(x)^2\mathrm{d}x$ and $\int_0^1 d_2(x)^2\mathrm{d}x$ are both $O(\delta^2)$. However, $\int_0^1 d_1(x)^2\mathrm{d}x=\lambda_1^2\left(\int_0^1 v_1(x)\mathrm{d}x\right)^2$ and $\lambda_1\approx-\sigma$, so $\int_0^1 v_1(x)\mathrm{d}x$ must be $O(\delta)$.

We now turn to $\|g_2\|_2^2=\sum_{j=2}^{\infty}\lambda_j^2$. Since

$$\lambda_1^3=-\left(\sigma^3+3e\int_0^1 d_1(x)^2+d_2(x)^2\,\mathrm{d}x-\sum_{j=2}^{\infty}\lambda_j^3\right),$$

and since $|\sum_{j=2}^{\infty}\lambda_j^3| \leqslant \|g_2\|_2^3$, $\lambda_1 \leqslant -\sigma + O(\|g_2\|_2^3)$. But then

$$\sum_{j=1}^{\infty}\lambda_j^2 \geqslant \sigma^2 + \|g_2\|_2^2 + O\left(\|g_2\|_2^3\right).$$

Since this must be within $O(\delta^2)$ of σ^2, we must have $\|g_2\|_2^2 = O(\delta^2)$.

Finally, we consider the fourth moment μ_4. The leading contribution is

$$\lambda_1^4\left(\int_0^1 v_1(x)^4\,\mathrm{d}x\right)^2 = \lambda_1^4\left(1+\int_0^1\left(v_1(x)^2-1\right)^2\mathrm{d}x\right)^2.$$

For this to be within $O(\delta^2)$ of σ^4, $\int_0^1(v_1(x)^2-1)^2\mathrm{d}x$ must be $O(\delta^2)$. □

4. Pointwise estimates

The upshot of section 3 is that g must be L^2-close to a symmetric bipodal graphon, with λ_1 being close to $-\sigma$, with the sum of the other λ_j^2 being small, and with $v_1(x)$ being close to 1 on a set of measure approximately 1/2 and close to -1 on a set of measure approximately 1/2. In this section we upgrade those L^2 estimates into pointwise estimates:

Proposition 10. *If g is an entropy-maximizing graphon, then $\|g_2(x,y)\|_\infty$ is $o(1)$.*

Proposition 11. *If g is an entropy-maximizing graphon, then $\|v_1(x)^2-1\|_\infty$ is $o(1)$.*

We prove propositions 10 and 11 with a series of lemmas. We begin by showing that g_1 and g_2 are pointwise bounded.

Lemma 12. *Let g be an entropy-maximizing graphon. For all $x,y\in[0,1]$, the following bounds apply:*

$$|v_1(x)| \leqslant |\lambda_1|^{-1}, \qquad |g_1(x,y)| \leqslant |\lambda_1|^{-1}, \qquad |g_2(x,y)| \leqslant 1+|\lambda_1|^{-1}.$$

Proof. We use the fact that

$$0 \leqslant e + g_1(x,y) + g_2(x,y) \leqslant 1$$

for all (x,y). The only way for $g_2(x,y)$ be big and positive (resp. negative) is for $g_1(x,y)$ to be big and negative (resp. positive). This can only occur if $|v_1(x)|$ is large for some x.

Suppose that there is a point x_0 with $v_1(x_0) > 1/|\lambda_1|$. Let I_+ be the set of x for which $v_1(x) > 0.9|\lambda_1 v_1(x_0)|^{-1}$ and let I_- be the set of x for which $v_1(x) < -0.9|\lambda_1 v_1(x_0)|^{-1}$. We already know that the set of points with $v_1(x)$ close to ± 1 each have measure close to 1/2, since $\int_0^1(v_1(x)^2-1)\mathrm{d}x = O(\delta^2)$ and $\int_0^1 v_1(x)\mathrm{d}x = O(\delta)$, so I_+ and I_- also each have measure close to 1/2.

Since λ_1 is negative, $g_1(x,y)$ is less than or equal to -0.9 for all $y\in I_+$ and is greater than or equal to 0.9 for all $y\in I_-$. Since e is close to 1/2, $e+g_1(x_0,y)$ is close to or less than -1.4 when $y\in I_+$ (and in particular is less than -1.3) and is close to or greater than 1.4 (and in particular is greater than 1.3) when $y\in I_-$. As a result, $g_1(x_0,y)$ has magnitude at least 0.3, and sign opposite to that of $g_2(x_0,y)$, for all $y\in I_+\cup I_-$. This means that $g_1(x_0,y)g_2(x_0,y) < -0.27$ for all $y\in I_+\cup I_-$.

When $y \notin I_- \cup I_+$, $|g_2(x_0,y)| \leqslant 0.9 < 1$, so $|g_2(x_0,y)| < 2$, so $g_1(x_0,y)g_2(x_0,y) < 2$. Since $I_+ \cup I_-$ is a set of measure $m \approx 1$,

$$\int_0^1 g_1(x_0,y)\, g_2(x_0,y)\, \mathrm{d}y \leqslant -0.27m + 2(1-m) < 0.$$

However,

$$\int_0^1 g_1(x_0,y)\, g_2(x_0,y)\, \mathrm{d}y = \sum_{j=2}^{\infty} \lambda_1 \lambda_j v_1(x_0)\, v_j(x_0) \int_0^1 v_1(y)\, v_j(y)\, \mathrm{d}y = 0,$$

by the orthogonality of the functions $v_j(y)$ in $L^2([0,1])$. This is a contradiction, so x_0 does not exist.

The same argument, with signs reversed, rules out the possibility that $v_1(x)$ is ever less than $-1/|\lambda_1|$. Since $|v_1(x)|$ is bounded by $|\lambda_1|^{-1}$, $|g_1(x,y)| = |\lambda_1 v_1(x) v_1(y)|$ is also bounded by $|\lambda_1|^{-1}$. Finally, we have that

$$-e - g_1(x,y) \leqslant g_2(x,y) \leqslant 1 - e - g_1(x,y).$$

Since e and $1-e$ are both less than 1, this implies that $|g_2(x,y)| < |g_1(x,y)| + 1 \leqslant |\lambda_1|^{-1} + 1$. □

Lemma 12 is stated in terms of λ_1, which of course depends on the graphon g. However, $\lambda_1 = -\sigma + o(1)$, so for small δ we can replace our bounds involving λ_1 with uniform bounds in terms of σ, at the cost of replacing the constant 1 with a slightly smaller number. For instance,

$$|v_1(x)| < 1.1\sigma^{-1}, \qquad |g_1(x,y)| < 1.1\sigma^{-1}, \qquad |g_2(x,y)| \leqslant 1 + 1.1\sigma^{-1}$$

whenever δ is sufficiently small. In practice, we do not need the specific bounds of lemma 12. All we really need is for $v_1(x)$, $g_1(x,y)$ and $g_2(x,y)$ to be bounded.

We next turn to showing that $g_2(x,y)$ is not only bounded but small. Since $\|g_2\|_2^2$ is $O(\delta^2)$, the set of points where $|g_2(x,y)|$ is not small (say, smaller than a fixed ϵ) has measure $O(\delta^2)$. We now establish a similar result for vertical strips.

Lemma 13. *Let g be an entropy-maximizing graphon. For any $\epsilon > 0$ and any $x \in [0,1]$, the set of y-values for which $|g_2(x,y)| > \epsilon$ has measure $o(1)$.*

Proof. Let $G(x,y) = \int_0^1 g(x,z)g(z,y)\,\mathrm{d}z$. As operator, this is the square of g. Expanding that square using $g(x,y) = e + g_1(x,y) + g_2(x,y)$, we let G_1 be the portion of G that comes from $e + g_1$, and let G_2 be the additional contributions that involve g_2,

$$\begin{aligned} G_1(x,y) &= \int_0^1 (e + g_1(x,z))(e + g_1(y,z))\,\mathrm{d}z \\ &= e^2 + \lambda_1^2 v_1(x) v_1(y) + e\left(\int_0^1 v_1(z)\,\mathrm{d}z\right)(v_1(x) + v_1(y)). \\ &= e^2 + \lambda_1 g_1(x,y) + O(\delta), \\ &= \lambda_1(e + g_1(x,y)) + e^2 - \lambda_1 e + O(\delta), \end{aligned} \tag{19}$$

since $\int_0^1 v_1(z)\,\mathrm{d}z = O(\delta)$ and $v_1(x)$ and $v_1(y)$ are bounded.

We next turn to G_2. Since $\int_0^1 v_1(z) g_2(y,z)\,\mathrm{d}z = 0$, there is no contribution from the product of g_1 and g_2. We only have eg_2 and g_2^2 terms, specifically

$$G_2(x,y) = e(d_2(x) + d_2(y)) + \int_0^1 g_2(x,z) g_2(y,z)\,\mathrm{d}z.$$

The function $d_2(y)$ has small L^2-norm, and so must be $o(1)$ except on a set of small measure. (Since x is fixed, we cannot similarly argue that $d_2(x)$ is small.) Finally, since g_2 is bounded and has small L^2 norm, the integral $\int_0^1 g_2(x,z) g_2(y,z)\mathrm{d}z$ for fixed x is small except for a set of y's that has small measure. The result is an estimate

$$G_2(x,y) = ed_2(x) + o(1)$$

that is true for y in the complement of a set of measure $o(1)$, where that small set may depend on x.

Combining this with our estimate of G_1, we have

$$G(x,y) = \left(e^2 - \lambda_1 e\right) + \lambda_1 g_1(x,y) + ed_2(x) + o(1) \tag{20}$$

for all but a small set of y's.

Since g is assumed to maximize entropy subject to constraints on $\varepsilon(g)$ and $\tau(g)$, the functional derivative of $S(g)$ must be a linear combination of the functional derivatives of $\varepsilon(g)$ and $\tau(g)$. This yields the pointwise equations

$$H'(g(x,y)) = \Lambda_e + \Lambda_t G(x,y), \tag{21}$$

where Λ_e and Λ_t are Lagrange multipliers.

For all but a small set of y's, equation (21) takes the form

$$H'(e + g_1(x,y) + g_2(x,y)) = \mu + \nu(e + g_1(x,y)) + \rho d_2(x) + o(1), \tag{22}$$

where

$$\mu = \Lambda_e + \Lambda_t\left(e^2 - \lambda_1 e\right), \qquad \nu = \lambda_1 \Lambda_t, \qquad \rho = e\Lambda_t.$$

For most values of (x,y), $e + g_1(x,y)$ is close to $e \pm \sigma$ and $g_2(x,y)$ and $d_2(x)$ are small. This fixes Λ_e and Λ_t, and therefore μ, ν, and ρ, to within a small error. Since $e = \frac{1}{2} + \delta$ and $H'(\frac{1}{2} - \sigma) = -H'(\frac{1}{2} + \sigma)$, we obtain

$$\begin{aligned} H'\left(\frac{1}{2} + \sigma\right) &= \Lambda_e + \Lambda_t\left(\frac{1}{4} - \sigma^2\right) + o(1) \\ -H'\left(\frac{1}{2} + \sigma\right) &= \Lambda_e + \Lambda_t\left(\frac{1}{4} + \sigma^2\right) + o(1), \end{aligned} \tag{23}$$

with solution

$$\begin{aligned} &\Lambda_e \approx -\frac{1}{4\sigma^2} H'\left(\frac{1}{2} + \sigma\right), \Lambda_t \approx \frac{1}{\sigma^2} H'\left(\frac{1}{2} + \sigma\right), \\ &\mu \approx -\frac{1}{2\sigma} H'\left(\frac{1}{2} + \sigma\right), \nu \approx \frac{1}{\sigma} H'\left(\frac{1}{2} + \sigma\right), \quad \rho \approx -\frac{1}{2\sigma^2} H'\left(\frac{1}{2} + \sigma\right), \end{aligned} \tag{24}$$

where all of the approximations are '$+o(1)$' as $\delta \to 0$.

From the explicit form of $H'(g) = \ln(1-g) - \ln(g)$, we see that there are only three roots to the equation $H'(g) = \mu + \nu g$, which are located near $g = e \pm \sigma$ and $g = e$. Our immediate goal is to show that $v_1(x)$ only takes values close to 0 and ± 1, which implies that $g_1(x,y)$ is only close to e and $e \pm \sigma$.

Since $\int v_1(x)\mathrm{d}x$ and $\int (v_1(x)^2 - 1)^2 \mathrm{d}x$ are small, the function v_1 must be close to 1 on an interval (call it I_1) of measure close to 1/2, must be close to -1 on an interval I_2 of measure close to $1/2$, may be close to 0 on a third interval I_3 of small measure, and may take on other values on a fourth interval I_4 of small measure.

Let x be an arbitrary point in $[0,1]$, and let y_1 and y_2 be generic points in I_1 and I_2. Equation (22) then determines $a = g_2(x,y_1)$ and $b = g_2(x,y_2)$ in terms of $d_2(x)$. What's more, g_2 takes values close to a on all of $\{x\} \times I_1$ (excepting those values of y where equation (22) does not apply), and takes values close to b on all of $\{x\} \times I_2$. We then compute

$$\int_0^1 g_2(x,y)\, v_1(y)\, \mathrm{d}y \approx (a-b)/2.$$

However, this integral must be zero, since $v_1(y)$ is orthogonal to all of the functions $v_i(y)$ that make up $g_2(x,y)$. We conclude that $a \approx b$.

If $b \approx a$, then $d_2(x) = \int_0^1 g_2(x,y)\mathrm{d}y$ also equals a and our equations at (x,y_2) and (x,y_1) become

$$\begin{aligned} H'\left(\frac{1}{2} + \sigma v_1(x) + a\right) &\approx \mu + \nu\left(\frac{1}{2} + \sigma v_1(x)\right) + \rho a \\ H'\left(\frac{1}{2} - \sigma v_1(x) + a\right) &\approx \mu + \nu\left(\frac{1}{2} - \sigma v_1(x)\right) + \rho a, \end{aligned} \tag{25}$$

Adding these equations, and using the fact that $H'(\frac{1}{2} - \sigma v_1(x) + a) = -H'(\frac{1}{2} + \sigma v_1 - a)$, we get

$$H'\left(\frac{1}{2} + \sigma v_1(x) + a\right) - H'\left(\frac{1}{2} + \sigma v_1(x) - a\right) \approx 2\mu + \nu + 2\rho a \approx -\frac{a}{\sigma^2} H'\left(\frac{1}{2} + \sigma\right). \tag{26}$$

By the mean value theorem, the left hand side of equation (26) is $2aH''(u)$ for some u between $\frac{1}{2} + \sigma v_1(x) - a$ and $\frac{1}{2} + \sigma v_1(x) + a$. Regardless of the value of u, this is a negative multiple of a. However, the right hand side is a positive multiple of a, since $H'(\frac{1}{2} + \sigma) < 0$. Since a negative multiple of a equals a positive multiple, a must be (approximately) zero.

In particular, $g_2(x,y) \approx 0$ for all y such that equation (22) applies. That is, for fixed x, $g_2(x,y)$ is close to zero except on a set of y's of small measure. □

Proof of proposition 10. By lemma 12, $g_2(x,y)$ is bounded. By lemma 13, for each x, $g_2(x,y)$ is small for all but a small set of y's. Combining these results, we see that the degree function $d_2(x)$ is everywhere small, as is the integral $\int_0^1 g_2(x,z) g_2(y,z)\, \mathrm{d}z$. That is, $G_2(x,y)$ is pointwise small. But then equation (22) applies everywhere, so $g_2(x,y)$ is small everywhere. □

Proof of proposition 11. In the notation of the proof of lemma 13, we must show that the intervals I_3 and I_4 are empty, implying that $v_1(x)$ is everywhere close to ± 1.

Since $G_2(x,y)$ is small for all (x,y), we must have

$$H'\left(\frac{1}{2} + \sigma v_1(x)\right) = \mu + \nu\left(\frac{1}{2} + \sigma v_1(x)\right)$$

for all $y \in I_2$. However, the only solutions to this equation are (approximately) $\sigma v_1(x) = \pm\sigma$ or 0, implying that $v_1(x) \approx \pm 1$ or 0. In other words, $x \in I_1 \cup I_2 \cup I_3$ and I_4 is empty.

Showing that I_3 is empty requires a completely different argument, since equation (21) is indeed satisfied when $x \in I_3$. However, equation (21) only defines stationary points with respect to pointwise small changes in $g(x,y)$. We also have to consider infinitesimal changes in the boundary between I_1, I_2 and I_3.

So suppose that we increase the size of I_1 by an amount ϵ at the expense of I_3. That is, we change the value of $v_1(x)$ from near 0 to near 1 on a set of small measure ϵ. To first order in ϵ, the change in entropy is $2\epsilon(H(\frac{1}{2}+\sigma) - H(\frac{1}{2}))$, since we are changing $g(x,y)$ from near zero to near $\frac{1}{2} \pm \sigma$ on a set of measure $2\epsilon + O(\epsilon^2)$, and since $H(\frac{1}{2}-\sigma) = H(\frac{1}{2}+\sigma)$. The change in edge density is $-2\epsilon\lambda_1 \int_0^1 v_1(y)\mathrm{d}y = O(\delta\epsilon)$. To leading order, the change in the triangle density is $3\lambda_1^3\epsilon$, since the λ_1^3 contribution to $\tau(g)$ is actually $\lambda_1^3(\int_0^1 v_1(x)^2\mathrm{d}x)^3$, which changes from λ_1^3 to $\lambda_1^3(1+\epsilon)^3 \approx \lambda_1^3(1+3\epsilon)$.

The variational equations $\mathrm{d}S = \Lambda_e \mathrm{d}\varepsilon + \Lambda_t \mathrm{d}\tau/3$ then become

$$2\left[H\left(\frac{1}{2}+\sigma\right) - H\left(\frac{1}{2}\right)\right] = \lambda_1^3\Lambda_t = \sigma H'\left(\frac{1}{2}+\sigma\right). \tag{27}$$

We expand both sides of equation (27) as power series in σ. The left-hand side is

$$2\sum_{k=1}^{\infty} \frac{H^{(2k)}(1/2)}{(2k)!}\sigma^{2k}.$$

The right-hand side is

$$\sum_{k=1}^{\infty} \frac{H^{(2k)}(1/2)}{(2k-1)!}\sigma^{2k}.$$

The coefficients agree when $k=1$, but are strictly greater for the right-hand side when $k>1$. Since all terms are strictly negative (insofar as all even derivatives of H are negative-definite), the right-hand side strictly smaller than the left-hand side.

Since varying the size of I_3 does not satisfy the variational equation (27), we cannot be in the interior of our parameter space. Rather, the measure of I_3 must be zero. □

5. Cost-benefit analysis

In sections 3 and 4 we showed that $g_2(x,y)$ is pointwise small, as is $v_1(x)^2 - 1$. In this section we show that they are zero, completing the proof of theorem 1. The key measures of how far they are from being zero are

$$\alpha^2 = \sum_{i=2}^{\infty} \lambda_i^2, \qquad \beta^2 = \int_0^1 \left(v_1(x)^2 - 1\right)^2 \mathrm{d}x, \quad \text{and} \quad \gamma = \int_0^1 v_1(x)\,\mathrm{d}x. \tag{28}$$

By corollary 9, α, β and γ are all $O(\delta)$. The symmetric bipodal graphon is characterized by $\alpha = \beta = \gamma = 0$.

We use the expansion (16) and compare the moments μ_{2k} to those of the symmetric bipodal graphon. We will estimate costs (terms that increase μ_{2k}) and benefits (terms that decrease μ_{2k}). We will show that having α or β or γ nonzero comes with costs that go as α^2, β^2, and γ^2, while the benefits are $o(\alpha^2+\beta^2+\gamma^2)$. When δ is sufficiently small, the costs exceed the benefits, so the symmetric bipodal graphon has more entropy than any graphon that is not symmetric bipodal.

We first establish the costs. Our triangle density is

$$e^3-\sigma^3=t=e^3+3e\int_0^1\left(d_1(x)^2+d_2(x)^2\right)\mathrm{d}x+\lambda_1^3+\sum_{j=2}^{\infty}\lambda_j^3.$$

Now

$$\int_0^1 d_1^2(x)\,\mathrm{d}x=\lambda_1^2\left(\int_0^1 v_1(x)\,\mathrm{d}x\right)^2=\lambda_1^2\gamma^2,$$

while $\int_0^1 d_2(x)^2\,\mathrm{d}x>0$ and

$$\left|\sum_{j=2}^{\infty}\lambda_j^3\right|\leqslant\left(\sum_{j=2}^{\infty}\lambda_j^2\right)^{3/2}=\alpha^3.$$

This implies that

$$\lambda_1^3\leqslant-\sigma^3-3e\sigma^2\gamma^2+O\left(\alpha^3\right),$$

so

$$\lambda_1^2\geqslant\sigma^2+2e\sigma\gamma^2+O\left((\alpha,\beta,\gamma)^3\right)$$

and

$$\mu_2=\lambda_1^2+\alpha^2\geqslant\sigma^2+\alpha^2+2e\sigma\gamma^2+o\left(\alpha^2+\beta^2+\gamma^2\right).$$

That is, there are α^2 and γ^2 costs associated with μ_2.

We next look at μ_4. This contains a term

$$\iint g_1(x,y)^4\,\mathrm{d}x\,\mathrm{d}y=\lambda_1^4\left(\int_0^1 v_1(x)^4\,\mathrm{d}x\right)^2=\lambda_1^4\left(1+\beta^2\right)^2\geqslant\sigma^4\left(1+2\beta^2\right).$$

That is, there is a cost proportional to β^2. Having established costs proportional to α^2, β^2, and γ^2, we just have to show that the benefits of having α,β,γ nonzero are smaller.

We are looking at even moments

$$\mu_{2k}=\iint\left(\delta+g_1+g_2\right)^{2k}\mathrm{d}x\,\mathrm{d}y.$$

We expand out the power, getting terms proportional to a power of δ times a power of g_1 times a power of g_2. We make repeated use of the following trick:

$$u^{2m}=1+u^{2m}-1=1+\left(u^2-1\right)p_m(u),$$

where

$$p_m(u) = 1 + u^2 + u^4 + \cdots + u^{2m-1}.$$

This means that

$$\begin{aligned} g_1(x,y)^{2m} &= \lambda_1^{2m}\left[1 + \left(v_1(x)^2 - 1\right)p_m(v_1(x))\right]\left[1 + \left(v_1(y)^2 - 1\right)p_m(v_1(y))\right] \\ &= \lambda_1^{2m}\Big[1 + \left(v_1(x)^2 - 1\right)p_m(v_1(x)) + \left(v_1(y)^2 - 1\right)p_m(v_1(y)) \\ &\quad + \left(v_1(x)^2 - 1\right)\left(v_1(y)^2 - 1\right)p_m(v_1(x))p_m(v_1(y))\Big] \end{aligned} \tag{29}$$

and

$$\begin{aligned} g_1(x,y)^{2m+1} &= \lambda_1^{2m+1}\Big[v_1(y)\left(v_1(x) + \left(v_1(x)^3 - v_1(x)\right)p_m(v_1(x))\right) \\ &\quad + v_1(x)\left(v_1(y)^3 - v_1(y)\right)p_m(v_1(y)) \\ &\quad + \left(v_1(x)^2 - 1\right)\left(v_1(y)^2 - 1\right)v_1(x)v_1(y)p_m(v_1(x))p_m(v_1(y))\Big]. \end{aligned} \tag{30}$$

We divide the terms obtained by expanding $\iint(\delta + g_1 + g_2)^{2k}$ into several classes.

(1) Terms with three or more powers of g_2. Since g_2 is pointwise small and g_1 is bounded, these are bounded by small multiples of $\iint g_2^2 = \alpha^2$. In other words, they are $o(\alpha^2)$.

(2) Terms with two powers of g_2, an even number of powers of δ and an even number of powers of g_1. These are manifestly positive and so represent costs, not benefits. Aside from the $\iint \delta^0 g_1^0 g_2^2 = \alpha^2$ contribution to μ_2 that we already considered, we do not keep track of these.

(3) Terms with two powers of g_2, an odd number of powers of δ and an odd number of powers of g_1. The integrand is a positive power of δ times a bounded quantity times $g_2(x,y)^2$, making the integral $O(\delta\alpha^2)$.

(4) Terms with one power of g_2, an odd power of δ and an even power of g_1. We expand these using equation (29) and consider each piece separately. First, we compute

$$\iint g_2(x,y)\,\mathrm{d}x\,\mathrm{d}y = -\iint g_1(x,y)\,\mathrm{d}x\,\mathrm{d}y = -\lambda_1\left(\int_0^1 v_1(x)\,\mathrm{d}x\right)^2 = -\lambda_1\gamma^2.$$

Next, the L^2 norm of $v_1(x)^2 - 1$ is β and the L^2 norm of g_2 is α, so

$$\iint \left(v_1(x)^2 - 1\right)p_m(v_1(x))\,g_2\,\mathrm{d}x\,\mathrm{d}y = O(\alpha\beta).$$

The $(v_1(y)^2 - 1)$ piece is similar, while the $(v_1(x)^2 - 1)(v_1(y)^2 - 1)$ piece is $O(\alpha\beta^2)$. Thus $\iint g_1^{odd} g_2\,\mathrm{d}x\,\mathrm{d}y = O(\alpha\beta) + O(\gamma^2)$, so $\iint \delta^{odd} g_1^{odd} g_2\,\mathrm{d}x\,\mathrm{d}y = O(\delta\alpha\beta) + O(\delta\gamma^2) = o(\alpha^2 + \beta^2 + \gamma^2)$.

(5) Terms with one power of g_2, an even power of δ and an odd power of g_1. We expand these using equation (30), noting that the first line in (30) has a single factor of $v_1(y)$ and the second line has a single factor of $v_1(x)$. However,

$$\int_0^1 v_1(y)\,g_2(x,y)\,\mathrm{d}y = \int_0^1 v_1(x)\,g_2(x,y)\,\mathrm{d}x = 0,$$

where x is arbitrary in the first integral and y is arbitrary in the second. Thus the first two lines contribute nothing and $\iint g_1(x,y)^{2m+1} g_2(x,y)\,\mathrm{d}x\,\mathrm{d}y$ is equal to

$$\lambda_1^{2m+1} \iint \Big(v_1(x)^2-1\Big)\Big(v_1(y)^2-1\Big)\Big(v_1(x)v_1(y)p_m(v_1(x))p_m(v_1(y))\Big) g_2(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

The factor $(v_1(x)^2-1)(v_1(y)^2-1)$ has L^2 norm β^2, the factor $g_2(x,y)$ has L^2 norm α, and the factor $v_1(x)v_1(y)p_m(v_1(x))v_1(y)p_m(v_1(y))$ is bounded, so the integral is $O(\alpha\beta^2) = o(\alpha^2+\beta^2+\gamma^2)$.

(6) Terms with no powers of g_2, an even number of powers of δ and an even number of powers of g_1. These are all positive and are at least as big as the corresponding terms for the symmetric bipodal graphon. We have already taken into account the costs associated with μ_2 and μ_4. There are additional costs associated with higher moments, but they are not needed for this proof.

(7) Finally, there are terms with no powers of g_2, an odd number of powers of δ and an odd number of powers of g_1. Note that

$$\iint g_1(x,y)^{2m+1}\,\mathrm{d}x\,\mathrm{d}y = \lambda_1^{2m+1}\nu_{2m+1}^2,$$

where

$$\begin{aligned}\nu_{2m+1} &= \int_0^1 v_1(x)^{2m+1}\,\mathrm{d}x \\ &= \int_0^1 v_1(x)\,dx + \int_0^1 \Big(v_1(x)^2-1\Big) v_1(x)\,p_m(v_1(x))\,\mathrm{d}x \\ &= \gamma + O(\beta),\end{aligned}$$

since $v_1(x)^2-1$ has L^2-norm β and $v_1(x)p_m(v_1(x))$ is bounded. Squaring ν_{2m+1} and multiplying by an odd power of δ gives $O(\delta\gamma^2)+O(\delta\beta^2)+O(\delta\beta\gamma)$, which is $o(\alpha^2+\beta^2+\gamma^2)$.

Putting everything together, we have identified costs proportional to α^2, β^2 and γ^2. Other costs only add to that total, so the total cost is at least a constant times $\alpha^2+\beta^2+\gamma^2$. All of the potential benefits are smaller, either involving three or more powers of α, β and γ, or δ times a quadratic function of α, β and γ, or the sup norm of g_2 times α^2. When δ is sufficiently small, the costs outweigh the benefits, so the optimal graphon is symmetric bipodal.

6. The extent of the symmetric bipodal phase

We have proven that the entropy-maximizing graphon is unique and symmetric bipodal on a region containing the interval $e = 1/2$, $0 < t < 1/8$. It is natural to ask how far this symmetric bipodal phase extends. There is considerable evidence that this phase contains much of the region $e \leqslant 1/2$, $t < e^3$.

- The unique entropy-maximizing graphon when $e < 1/2$ and $t = 0$ is known to be symmetric bipodal, with $\sigma = e$ [29]. The entropy is $\frac{1}{2}H(2e)$.

- When $e < 1/2$ and $t < e^3$, the symmetric bipodal graphon has entropy strictly higher than any other graphon of the form $g(x,y) = e + \lambda_1 v_1(x) v_1(y)$. This is proposition 14, proven below.
- When $e < 1/2$ and t sufficiently close to but below e^3, the symmetric bipodal graphon has entropy strictly higher than any other bipodal graphon. This is proposition 16, proven below.
- Numerical investigations of the region $e < 1/2$, $t < 1/8$ [26] did not turn up *any* regions where the symmetric bipodal graphon was not optimal. That is not a proof that such regions do not exist, of course, but it does suggest that such regions are likely to be small.

Despite this evidence, theorem 3 says that there is a (possibly very small) open subset of the region $e < 1/2, t < 1/8$ on which the symmetric bipodal graphon is *not* optimal. In this section we state and prove propositions 14 and 16 and then prove theorem 3.

Proposition 14. *Suppose that $e < 1/2$ and that g is a graphon of the form*

$$g(x,y) = e + \lambda_1 v_1(x) v_1(y),$$

with edge density e and triangle density $t = e^3 - \sigma^3 < e^3$. Then $S(g)$ is bounded above by $\frac{1}{2}(H(e+\sigma) + H(e-\sigma))$, with equality if and only if g is symmetric bipodal.

Proof. First note that $\int_0^1 v_1(x)\,\mathrm{d}x = 0$, since the overall edge density is exactly e. The triangle density is then $e^3 + \lambda_1^3$, so $\lambda_1 = -\sigma$. Since $0 \leqslant g(x,y)$, $v_1(x)$ is bounded in magnitude by $\sqrt{e/\sigma}$ and $\lambda_1 v_1(x) v_2(x)$ is bounded in magnitude by e. This implies that the power series

$$H(g(x,y)) = \sum_{j=0}^{\infty} \frac{H^{(j)}(e)}{j!} (\lambda_1 v_1(x) v_1(y))^j$$

converges absolutely. Integrating over x and y then gives

$$S(g) = \sum_{j=1}^{\infty} (-\sigma)^j \frac{H^{(j)}(e)}{j!} \left(\int_0^1 v_1(x)^j \,\mathrm{d}x \right)^2.$$

Since $e < 1/2$, all the odd derivatives of H are positive at e, while all the even derivatives are negative. Multiplying by $(-\sigma)^j$, all of the terms with $j > 0$ are negative. We maximize the entropy by minimizing $\int_0^1 v_1(x)^j\,\mathrm{d}x$ for j even and by having $\int_0^1 v_1(x)^j\,\mathrm{d}x = 0$ for j odd. The second moment $\int_0^1 v_1(x)^2\,\mathrm{d}x$ is always equal to 1. The fourth and higher even moments are minimized when (and only when!) $v_1(x)^2$ is constant and equal to 1. Since $\int_0^1 v_1(x)\,\mathrm{d}x = 0$, this means that $v_1(x)$ is $+1$ on a set of measure 1/2 and -1 on a set of measure 1/2, which makes all of the odd integrals zero, as desired. In other words, the symmetric bipodal graphon is the unique entropy maximizer among graphons of this form. □

Before turning to what happens just below the line $t = e^3$, we establish constraints on the form of any entropy-maximizing bipodal graphon with $t < e^3$. We use the parameters (a,b,c,d) of figure 5 to describe bipodal graphons. Without loss of generality we can assume that $a \leqslant b$, since otherwise we could just swap c and $1-c$ while swapping a and b.

Proposition 15. *Suppose that $t = e^3 - \sigma^3 < e^3$ and that a graphon g maximizes entropy among all bipodal graphons with edge density e and triangle density t. (We do not assume that g maximizes entropy among all graphons, just that it is the best bipodal graphon.) Then either $a = b$ and $c = 1/2$ (a symmetric bipodal graphon) or $a < b < d$ and $c < 1/2$.*

Proof. In this setting, the variational equation (21) become

$$\begin{aligned} H'(a) &= \Lambda_e + \Lambda_t \left(ca^2 + (1-c)\, d^2\right), \\ H'(b) &= \Lambda_e + \Lambda_t \left(cd^2 + (1-c)\, b^2\right), \\ H'(d) &= \Lambda_e + \Lambda_t \left(cad + (1-c)\, bd\right). \end{aligned} \tag{31}$$

Subtracting the second equation from the first gives

$$H'(a) - H'(b) = \Lambda_t \left(c\left(a^2 - d^2\right) + (1-c)\left(d^2 - b^2\right)\right).$$

If $a = b$, then the left hand side is zero and the right hand side is a nonzero multiple of $(1-2c)(a^2-d^2)$, implying that either $c = 1/2$ or $a = d$. But if $a = b = d$, then the triangle density is exactly e^3, which is a contradiction. Thus $a = b$ implies that the graphon must by symmetric bipodal.

We now turn to the possibility that $a < b$. The left hand side is then positive, since H' is a decreasing function. Since $\Lambda_t = \frac{1}{3}\frac{\partial S}{\partial t}$ is positive, we must have

$$c\left(a^2 - d^2\right) + (1-c)\left(d^2 - b^2\right) > 0.$$

This either requires $d < a$, in which case $c > 1/2$ (since $a^2 - d^2 < b^2 - d^2$) or $d > b$, in which case $c < 1/2$.

Let λ_1 and λ_2 be the two nonzero eigenvalues of g. The trace of g is $\lambda_1 + \lambda_2 = ca + (1-c)b$, while the trace of G is $\lambda_1^2 + \lambda_2^2 = c^2a^2 + (1-c)^2b^2 + 2c(1-c)d^2$. From this we can compute

$$\lambda_1\lambda_2 = \frac{1}{2}\left((\lambda_1+\lambda_2)^2 - \left(\lambda_1^2 + \lambda_2^2\right)\right) = c\,(1-c)\left(ab - d^2\right).$$

If d were less than a and b, this would be positive, meaning that both eigenvalues would be positive. Moreover, one of the two eigenvalues is at least e, so the triangle density, which is $\lambda_1^3 + \lambda_2^3$, would be greater than e^3. This rules out the possibility that $d < a$, and we conclude that $a < b < d$ and $c < 1/2$. □

When $e > 1/2$ and t is slightly less than e^3, the optimal graphon has been proven to take this form, with $a \approx 1 - e$, b slightly less than e, and d slightly greater than e, and with c small. The situation is different when $e < \frac{1}{2}$.

Proposition 16. *Suppose that $e < 1/2$ and that g is a bipodal graphon with edge density e and triangle density $t = e^3 - \sigma^3 < e^3$. Then, for σ sufficiently small, $S(g)$ is bounded above by $\frac{1}{2}\left(H(e+\sigma) + H(e-\sigma)\right)$, with equality if and only if g is symmetric bipodal.*

Proof. Let

$$\Delta a := a - e, \qquad \Delta b := b - e, \qquad \Delta d := d - e$$

and let

$$\eta := \frac{c}{1-c}\Delta a + \Delta d, \qquad \alpha := 2c\eta - \frac{c}{1-c}\Delta a. \tag{32}$$

The leading term is α[1], while η measures the extent to which the degree function fails to be constant. We can then express all of our quantities in terms of α, η and c.

$$\begin{aligned}\Delta a &= -\frac{1-c}{c}\alpha + e(1-c)\eta,\\ \Delta b &= -\frac{c}{1-c}\alpha - 2c\eta,\\ \Delta d &= \alpha + (1-2c)\eta.\end{aligned} \tag{33}$$

In terms of these parameters, the triangle density works out to be

$$e^3 - \sigma^3 = \tau(g) = e^3 + 3(e-\alpha)c(1-c)\eta^2 - \alpha^3. \tag{34}$$

Since the triangle density is less than e^3, η^2 must be $O(\alpha^3)$. That is, η is much smaller than α.

We now compute the entropy

$$S(g) = \sum_{k=0}^{\infty}\frac{H^{(k)}(e)}{k!}\mu_k,$$

where

$$\mu_k = \iint (g(x,y)-e)^k\,\mathrm{d}x\,\mathrm{d}y = c^2(\Delta a)^k + (1-c)^2(\Delta b)^k + 2c(1-c)(\Delta d)^k.$$

Since $e < 1/2$, the odd derivatives of H at e are positive, while the even derivatives are negative, so we want to minimize the even moments and maximize the odd moments. The symmetric bipodal graphon (uniquely) minimizes the even moments and has all the odd moments equal to zero. For an asymmetric bipodal graphon to do as well, it must have some positive odd moments.

The moment μ_k is a kth order homogeneous polynomial in α and η with coefficients that depend on c. Since $\eta = O(\alpha^{3/2})$,

$$\mu_k = \left(2c(1-c) + (-1)^k\frac{(1-c)^k}{c^{k-2}} + (-1)^k\frac{c^k}{(1-c)^{k-2}}\right)\alpha^k + O\left(\alpha^{k+\frac{1}{2}}\right).$$

The coefficient of α^k is zero when $k=1$, is 1 when $k=2$, is negative when k is an odd number greater than 1, and is greater than 1 when k is an even number greater than 2. In other words, all moments with $k>2$ are worse, to leading order, than the moments of the symmetric bipodal graphon.

There is one more point we must account for. For k odd, $(2c(1-c) - \frac{(1-c)^k}{c^{k-2}} - \frac{c^k}{(1-c)^{k-2}})$ goes to zero as $(1-2c)^2$ as $c\to 1/2$. We must rule out the possibility that other contributions to μ_k might become greater than the α^k term as c approaches 1/2.

This requires estimates on η. From the formula for the triangle density, we have that

$$\alpha \approx \sigma + \frac{ec(1-c)\eta^2}{\sigma^2},$$

[1] This is not the same as the α in the proof of theorem 1. There are only so many Greek letters in the alphabet.

and hence that

$$\mu_2 = \alpha^2 + 2c(1-c)\eta^2 \approx \sigma^2 + \frac{2ec(1-c)}{\sigma}\eta^2.$$

That is, there is a cost proportional to η^2/σ that does not vanish as $c \to 1/2$. Meanwhile, all contributions to moments involving odd powers of η are proportional to $(1-2c)$. This is because the graphon is invariant under the transformation $\eta \to -\eta$, $c \to 1-c$, $a \leftrightarrow b$. The leading such contribution comes from μ_3 and goes as $(1-2c)\alpha^2\eta \approx (1-2c)\sigma^2\eta$ times a polynomial in c that does not vanish at $c = 1/2$. Setting the derivative of the entropy with respect to η equal to zero tells us that $\eta = O(\sigma^3(1-2c))$. All contributions from odd powers of η are thus $O(\sigma^5(1-2c)^2)$, and so are dominated by the α^3 contribution to μ_3, while all contributions from even powers of η are dominated by the μ^2/σ contribution to μ_2. □

Thanks to propositions 14 and 16, any graphon that does better than symmetric bipodal in the region just below $t = e^3$ with $e < 1/2$ must be at least tripodal (or perhaps not even multipodal at all) and the difference between that graphon and a constant graphon must have rank at least two. That is exactly what we construct in the proof of theorem 3.

Proof of theorem 3. We consider values of e and $t = e^3 - \sigma^3$ where $e < e_0$ and σ is sufficiently small. The number e_0 is defined by the equation

$$3H'''(e_0)^2 = H''(e_0)H''''(e_0), \tag{35}$$

which simplifies to $6e_0^2 - 6e_0 + 1 = 0$, or $e_0 = (3-\sqrt{3})/6 \approx 0.2113$. When $e < e_0$, $3H'''(e)^2$ is greater than $H''(e)H''''(e)$. In fact, as $e \to 0$, $3H'''(e)^2$ goes as $3e^{-4}$, while $H''(e)H''''(e)$ goes as $2e^{-4}$. However, as e approaches $1/2$, $3H'''(e)^2$ goes to zero while $H''(e)H''''(e)$ does not.

Let $A > B > 0$ and let

$$F(A,B) = \frac{H(e+A+B) + H(e-A+B) - 2H(e) - 2BH'(e)}{(A^3 - B^3)^{2/3}}. \tag{36}$$

We will eventually choose A and B to maximize $F(A,B)$. Pick a small number c and divide the interval $[0,1]$ into three pieces:

$$I_1 = [0, c/2], \qquad I_2 = (c/2, c], \qquad I_3 = (c, 1].$$

Consider the graphon

$$g(x,y) = \begin{cases} e - A + B(1-c) & (x,y) \in I_1 \times I_1 \cup I_2 \times I_2 \\ e + A + B(1-c) & (x,y) \in I_1 \times I_2 \cup I_2 \times I_1 \\ e - cB & (x,y) \in [(I_1 \cup I_2) \times I_3] \cup [I_3 \times (I_1 \cup I_2)] \\ e + \frac{c^2}{1-c}B & (x,y) \in I_3 \times I_3. \end{cases} \tag{37}$$

Equivalently, $g(x,y) = e - cAv_1(x)v_1(y) + cBv_2(x)v_2(y)$, where

$$v_1(x) = \begin{cases} c^{-1/2} & x \in I_1, \\ -c^{-1/2} & x \in I_2, \\ 0 & x \in I_3, \end{cases} \qquad v_2(x) = \begin{cases} \sqrt{(1-c)/c} & x \in I_1 \cup I_2, \\ -\sqrt{c/(1-c)} & x \in I_3. \end{cases}$$

This graphon has edge density $\varepsilon(g) = e$ and triangle density

$$\tau(g) = e^3 - c^3A^3 + c^3B^3.$$

Setting the triangle density equal to $e^3 - \sigma^3$ gives

$$c = \sigma\left(A^3 - B^3\right)^{-1/3}.$$

We now estimate the entropy

$$\begin{aligned} S(g) &= \frac{c^2}{2}H(e - A + (1-c)B) + \frac{c^2}{2}H(e + A + (1-c)B) \\ &\quad + 2c(1-c)H(e - cB) + (1-c)^2 H\left(e + \frac{c^2}{1-c}B\right) \end{aligned} \tag{38}$$

to order c^2, or equivalently to order σ^2. The first two terms already are $O(c^2)$, so we can simply replace $e \pm A + (1-c)B$ with $e \pm A + B$. For the remaining terms, we can use a linear approximation for $H(u)$. The result is

$$\begin{aligned} S(g) &= H(e) + \frac{c^2}{2}\left[H(e - A + B) + H(e + A + B) - 2H(e) - 2BH'(e)\right] + O\left(c^3\right) \\ &= H(e) + \frac{1}{2}F(A,B)\sigma^2 + O\left(\sigma^3\right). \end{aligned} \tag{39}$$

For comparison, the symmetric bipodal graphon has entropy

$$H(e) + \frac{1}{2}H''(e)\sigma^2 + O\left(\sigma^4\right).$$

If $F(A,B) > H''(e)$, and if σ is sufficiently small, then the tripodal graphon has more entropy than the symmetric bipodal graphon.

What remains is showing that we can get $F(A,B) > H''(e)$ when $e < e_0$. Let A be a small positive number and let

$$B = -\frac{H'''(e)}{2H''(e)}A^2.$$

Since $H'''(e) > 0 > H''(e)$, B is positive. Since $A^3 - B^3 = A^3 + O(A^6)$, $(A^3 - B^3)^{2/3} = A^2 + O(A^5)$. We compute the numerator of $F(A,B)$ to order A^4 by doing a 4th order Taylor series expansion of $H(e - A + B)$ and $H(e + A + B)$ around e and keeping terms proportional to A^2, A^4, B, B^2, and A^2B. (The expression is even in A, so we only get even powers of A.) The result is

$$\begin{aligned} F(A,B) &= \frac{\left(A^2 + B^2\right)H''(e) + A^2BH'''(e) + \frac{1}{12}A^4H''''(e) + O\left(A^6\right)}{\left(A^3 - B^3\right)^{2/3}} \\ &= \frac{A^2H''(e) + A^4\left(\frac{H''''(e)}{12} - \frac{\left(H'''(e)\right)^2}{4H''(e)}\right) + O\left(A^6\right)}{A^2 + O\left(A^5\right)} \\ &= H''(e) + \left(\frac{H''''(e)}{12} - \frac{\left(H'''(e)\right)^2}{4H''(e)}\right)A^2 + O\left(A^3\right). \end{aligned} \tag{40}$$

Since $e < e_0$, the coefficient of A^2 is positive, so $F(A,B) > H''(e)$ when A is small. $\square$

7. Symmetry as an order parameter

Question: considering figure 3 and the two open subsets $\mathcal{O}$ and $\mathcal{O}_1$ defined by figures 1 and 4, could $\mathcal{O}$ and $\mathcal{O}_1$ be part of the same phase?

Although there is no barrier between $\mathcal{O}$ and $\mathcal{O}_1$ like the curve $t = e^3$ between $\mathcal{O}_2$ and $\mathcal{O}_1$, the answer must be 'no', thanks to the following symmetry argument. On $\mathcal{O}$, the entropy-maximizing graphon has constant degree function $d(x) = \int_0^1 g(x,y)\,\mathrm{d}y = e$. The density T_2 of 2-stars is given by the integral $\int_0^1 d(x)^2\,\mathrm{d}x$, so $Q = T_2 - e^2$ is identically zero on $\mathcal{O}$. If $\mathcal{O}$ and $\mathcal{O}_1$ were part of the same phase and we had an analytic curve $c(s)$ running between them, then Q would have to be zero on the first part of the curve and then by analyticity it would have to be zero on the entire curve. However, it is easy to check that Q is not zero on all of $\mathcal{O}_1$, insofar as the degree function for the graphon of theorem 6 is not constant. Instead, Q is a nonzero multiple of σ^5 plus $O(\sigma^6)$. This contradiction proves our assertion. □

This argument has a very similar flavor to an argument that is common in statistical physics. There, if you can find an 'order parameter' [2, 32], a physical quantity which is identically zero on an open subset of the parameter space and nonzero in another, then it cannot be analytic along any path from the first region to the second, so the open subsets cannot be parts of the same phase. Finding such an order parameter can be very difficult, but once found it can be very useful, as we now show.

Of course the symmetry itself is not literally an order parameter, since order parameters are numerical quantities. Rather, the order parameter is a quantity such as Q that measures the extent to which the symmetry is broken. Within the symmetric phase it is zero, but elsewhere it is nonzero, so any path from the first region to the second must encounter a phase transition.

To appreciate the subtleties associated with some phase transitions in real materials, consider water in various common states. First consider *gaseous* water (steam) at temperature T_1 just above 100° Celsius and atmospheric pressure P, and *liquid* water at temperature T_2 just below 100° Celsius and again pressure P. The mass density is much lower in the gaseous state than in the liquid state, so in any reasonable sense there is a sharp transition in state corresponding to the (arbitrarily) small change in temperature. Now consider the pair of states: *liquid* water at temperature T_3 just above 0° Celsius and atmospheric pressure P and *solid* water (ice) at temperature T_4 just below 0° Celsius and again pressure P. Again the mass density is different between these two states (ice floats on water) so again in any reasonable sense there is a sharp transition in state corresponding to the (arbitrarily) small change in temperature.

However, there is a big difference between these two transitions. Consider a different way of changing the state from that gaseous state to its 'neighboring' liquid state. It is experimentally possible, by slowly accessing high temperatures and high pressures, to use a different path between liquid water and steam *without making any sharp change in state.* One says there is a 'critical' point in the gas/liquid transition. However, experiments indicate that there is NO critical point in the liquid/solid transition; no matter how you vary temperature and pressure to move slowly between a state of liquid water and a state of solid ice, you must go through a sharp transition. Percy Bridgman received the Nobel prize in 1946 for his extensive experiments on high pressure, one result of which was to demonstrate that there is no critical point on the transition between fluid and solid in *any* known material. It is an old problem to try to understand why this should be the case. Consider the following quote in [35, p 11]

- The most outstanding unsolved problem of equilibrium statistical mechanics is the problem of the phase transitions. Why do all substances occur in at least three phases, the solid, liquid, and vapor phase which can coexist in the triple point? Why is there, again for all substances, a critical point for the vapor-liquid equilibrium, while apparently there is no critical point for the fluid-solid transition. Note that since these are *general* phenomena, they must have a *general* explanation; the precise details of the molecular structure and of the intermolecular forces should not matter.

As described in [2, p 19], Lev Landau tried to use symmetry as an order parameter to solve this problem:

- It was Landau (1958) who, long ago, first pointed out the vital importance of symmetry in phase transitions. This, the First theorem of solid-state physics, can be stated very simply: it is impossible to change symmetry gradually. A given symmetry element is either there or it is not; there is no way for it to grow imperceptibly. This means, for instance, that there can be no critical point for the melting curve as there is for the boiling point: it will never be possible to go continuously through some high-pressure phase from liquid to solid.

While appealing, Landau's argument is not universally accepted; see [23, p 122]. Part of the problem is that there is no known model in equilibrium statistical mechanics which can be *proven* to exhibit both fluid and solid phases [5, 35]. Even if such phases were proven to exist, it is not at all clear how to define an appropriate order parameter.

Yet that is exactly what we have done in the context of random graphs at the beginning of this section: we contrasted the subset $\mathcal{O}$, part of a 'symmetric' phase in which the order parameter Q vanishes identically, with part of a phase in which Q is not zero.

Of course this is not a solution of the classic problem of proving a solid-fluid phase transition in a reasonable statistical mechanics model; we are working with random graphs, not with configurations of atoms. But that's actually the point! Graph models are a wonderful laboratory for developing, with full mathematical rigor, techniques that are simply too hard in statistical physics, a laboratory where important structural questions can be successfully solved.

8. Summary

This paper is part of a series [10–13, 19, 21, 26–30] studying combinatorial systems (graphs, permutations, sphere packing) under competing constraints, an extension of extremal combinatorics but concentrating on *nonextreme* states of the systems, that is, states under nonextreme constraints. We study asymptotically large systems and for graphs and permutations we use a large deviation principle (LDP) to analyze 'typical' (i.e. exponentially most) constrained states. (We do not know an LDP for sphere packing but analyze such systems using the hard sphere model [18] in equilibrium statistical mechanics.) In this paper we sharpened our notion of phase by the use of analyticity; see section 7.

Our goal in studying typical nonextremely-constrained states in these combinatorial systems is to analyze *emergent smoothness* in response to infinitesimal change in the constraints, and of the combinatorial systems we have considered we have found dense graphs the most amenable to development.

By design, our graph modeling has many features in common with that of the equilibrium statistical mechanics of particles with short range forces, but also has a significant difference: there is no 'distance' between edges, so each edge has the same influence on any other edge. In statistical mechanics, models with this feature of the influence of particles on one another are called 'mean-field', and although not part of the mathematical formalism [32], mean-field models such as Curie-Weiss and van der Waals [34] are used to study phase transitions where more realistic models prove too difficult. The random graph model we have been discussing has, in this sense, more in common with mean-field models of statistical mechanics, and, as seen by our success in proving the existence of phases and phase transitions, may be able to provide a mathematical formalism for studying the asymptotics of graphs and other combinatorial objects which will be as fruitful mathematically as statistical mechanics has been.

We conclude with the following open problems in this edge/triangle model.

(1) What is the actual structure of the optimal graphon when $e < e_0$ and t is slightly below e^3? Does it resemble the example given in section 6, or is its structure still wilder? How does the behavior of this graphon as $\sigma \to 0$ compare to the moderate deviations results for $G(n,m)$ in [20]?
(2) Is there a succession of phases as $e \to 0$ and t remains close to e^3, with tripodal graphons giving way to 4-podal, 5-podal, and so on?
(3) When $e_0 < e < 1/2$ and t is slightly less than e^3, is the optimal graphon symmetric bipodal, or is it something else?
(4) When $e < 1/2$ and $t = 0$, the optimal graphon is known to be symmetric bipodal. What if $e < 1/2$ and t is slightly positive?
(5) Proposition 16 is stated for t close to e^3. However, the only step in the proof that uses $t \approx e^3$ is the estimate that η is much smaller than α. Can the result be extended to the entire region $e < 1/2$, $t < e^3$?
(6) In [12] it is proven that there are two open sets with supersatured triangles, $t > e^3$, which extend to phases. One of these is bounded below by the curve $t = e^3, e < 1/2$, while the other is bounded by $t = e^3, e > 1/2$. Are these actually parts of the same phase, or are they distinct?
(7) In section 4 of [13], numerical evidence is given of phase transitions in the edge/triangle model along curves where the entropy-optimizing graphon is not unique. It would be of interest to prove this. In principle, it may also be possible to have regions of positive area on which the optimizing graphon is not unique, which would be a challenge to interpret.

Data availability statement

No new data were created or analysed in this study.

Acknowledgment

This work was partially supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC-2047/1 - 390685813, and by a fellowship from the Alfred P Sloan Foundation. This material is based upon work supported by the National Science Foundation under Grant Nos. 2145800 and 2204449. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

ORCID iDs

Joe Neeman https://orcid.org/0000-0001-8483-4632
Charles Radin https://orcid.org/0000-0002-4776-7335
Lorenzo Sadun https://orcid.org/0000-0002-2518-573X

References

[1] Aldous D 1981 Representations for partially exchangeable arrays of random variables *J. Multivar. Anal.* **11** 581–98
[2] Anderson P W 1984 *Basic Notions of Condensed Matter Physics* (Menlo Park)
[3] Borgs C, Chayes J and Lovász L 2010 Moments of two-variable functions and the uniqueness of graph limits *Geom. Funct. Anal.* **19** 1597–619
[4] Borgs C, Chayes J, Lovász L, Sós V T and Vesztergombi K 2008 Convergent graph sequences I: subgraph frequencies, metric properties and testing *Adv. Math.* **219** 1801–51
[5] Brush S G 1983 *Statistical Physics and the Atomic Theory of Matter, from Boyle and Newton to Landau and Onsager* (Princeton University Press) p 277
[6] Chatterjee S and Diaconis P 2013 Estimating and understanding exponential random graph models *Ann. Statist.* **41** 2428–61
Chatterjee S and Diaconis P 2011 arXiv:1102.2650
[7] Chatterjee S and Varadhan S R S 2011 The large deviation principle for the Erdős-Rényi random graph *Eur. J. Comb.* **32** 1000–17
Chatterjee S and Varadhan S R S 2010 arXiv:1008.1946
[8] Dembo A and Lubetzky E 2018 A large deviation principle for the Erdős-Rényi uniform random graph *Electron. Commun. Probab.* **23** 1–13
[9] Hoover D N 1982 *Row-Column Exchangeability and a Generalized Model for Probability, Exchangeability in Probability and Statistics (Rome, 1981)* (North-Holland) pp 281–91
[10] Kenyon R, Král' D, Radin C and Winkler P 2020 Permutations with fixed pattern densities *Random Struct. Algorithms* **56** 220–50
[11] Kenyon R, Radin C, Ren K and Sadun L 2017 Multipodal structures and phase transitions in large constrained graphs *J. Stat. Phys.* **168** 233–58
[12] Kenyon R, Radin C, Ren K and Sadun L 2016 Bipodal structure in oversaturated random graphs *Int. Math. Res. Not.* **2018** 1009–44
[13] Kenyon R, Radin C, Ren K and Sadun L 2017 The phases of large networks with edge and triangle constraints *J. Phys. A: Math. Theor.* **50** 435001
[14] Lovász L and Szegedy B 2006 Limits of dense graph sequences *J. Combin. Theory* B **98** 933–57
[15] Lovász L and Szegedy B 2007 Szemerédi's lemma for the analyst *GAFA Geom. Funct. Anal.* **17** 252–70
[16] Lovász L and Szegedy B 2011 Finitely forcible graphons *J. Combin. Theory* B **101** 269–301
[17] Lovász L 2012 *Large Networks and Graph Limits* (American Mathematical Society)
[18] Löwen H 2000 Fun with hard spheres *Statistical Physics and Spatial Statistics : The Art of Analyzing and Modeling Spatial Structures and Pattern Formation* (*Lecture Notes in Physics*) ed K R Mecke and D Stoyen (Springer-Verlag) pp 295–331
[19] Neeman J, Radin C and Sadun L 2020 Phase transitions in finite random networks *J. Stat. Phys.* **181** 305–28
[20] Neeman J, Radin C and Sadun L 2023 Moderate deviations in triangle count *Random Struct. Algorithms* **63** 1–42
[21] Neeman J, Radin C and Sadun L 2023 Typical large graphs with given edge and triangle densities *Probab. Theory Relat. Fields* **186** 1167–223
[22] Newman M E J 2010 *Networks: An Introduction* (Oxford University Press)
[23] Pippard A 1979 *The Elements of Classical Thermodynamics* (Cambridge University Press)
[24] Park J and Newman M E J 2005 Solution for the properties of a clustered network *Phys. Rev.* E **72** 026136
[25] Pikhurko O and Razborov A 2017 Asymptotic structure of graphs with the minimum number of triangles *Combin. Probab. Comput.* **26** 138–60
Pikhurko O and Razborov A 2012 arXiv:1204.2846
[26] Radin C, Ren K and Sadun L 2014 The asymptotics of large constrained graphs *J. Phys. A: Math. Theor.* **47** 175001
[27] Radin C, Ren K and Sadun L 2018 A symmetry breaking transition in the edge/triangle network model *Ann. Inst. Henri Poincare* D **5** 251–86
[28] Radin C, Ren K and Sadun L 2017 Surface effects in dense random graphs with sharp edge constraint (arXiv:1709.01036v2)
[29] Radin C and Sadun L 2013 Phase transitions in a complex network *J. Phys. A: Math. Theor.* **46** 305002 Radin C and Sadun L 2013 arXiv:1301.1256

[30] Radin C and Sadun L 2015 Singularities in the entropy of asymptotically large simple graphs *J. Stat. Phys.* **158** 853–65
[31] Radin C and Sadun L 2023 Optimal graphons in the edge-2star model (arXiv:2305.00333)
[32] Ruelle D 1969 *Statistical Mechanics; Rigorous Results* (Benjamin)
[33] Strauss D 1986 On a general class of models for interaction *SIAM Rev.* **28** 513–27
[34] Thompson C J 1972 *Mathematical Statistical Mechanics* (Princeton University Press)
[35] Uhlenbeck G E 1968 *Fundamental Problems in Statistical Mechanics II* ed E G D Cohen (Wiley)