Probability Theory and Related Fields
https://doi.org/10.1007/s00440-023-01187-8

Typical large graphs with given edge and triangle densities

Joe Neeman[1] · Charles Radin[1] · Lorenzo Sadun[1]

Received: 28 March 2022 / Revised: 23 November 2022 / Accepted: 17 December 2022

Abstract
The analysis of large simple graphs with extreme values of the densities of edges and triangles has been extended to the statistical structure of typical graphs of fixed intermediate densities, by the use of large deviations of Erdős-Rényi graphs. We prove that the typical graph exhibits sharp singularities as the constraining densities vary between different curves of extreme values, and we determine the precise nature of the singularities. The extension to graphs with fixed densities of edges and k-cycles for odd $k > 3$ is straightforward and we note the simple changes in the proof.

Mathematics Subject Classification 60B20 · 05C80 · 60F10

1 Introduction

Our results concern the nature of simple graphs on n vertices, for large n, constrained to have density ε of edges and τ of triangles. The range of achievable values of the pair (ε, τ) was an old problem in extremal combinatorics initiated by Turán in 1941 [27]. The extremal graph theory of these constraints was recently completed by Razborov et al. in [23, 26], which also contain a good history of this problem; see Fig. 1. The graphs associated with some parts of the boundary of this region are not unique, but it is not difficult to characterize those probabilistically using the graphon formalism

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 (Grant No. FG-2018-10533).

✉ Joe Neeman
joeneeman@gmail.com

Charles Radin
radin@math.utexas.edu

Lorenzo Sadun
sadun@math.utexas.edu

1 Department of Mathematics, The University of Texas at Austin, Austin, TX 78712, USA

Published online: 17 January 2023

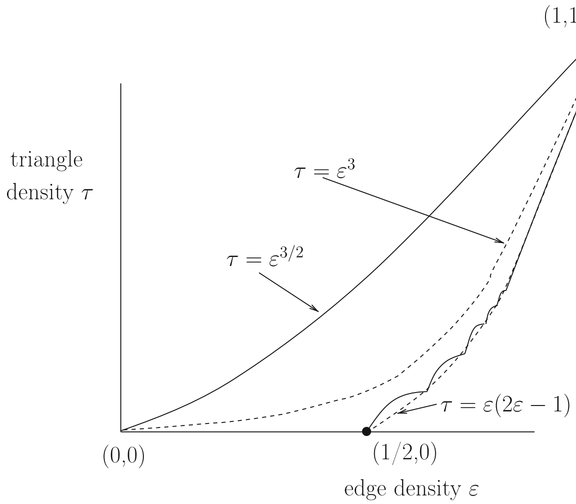

Fig. 1 The boundary of the achievable parameters is in solid lines. The figure is distorted to expose features

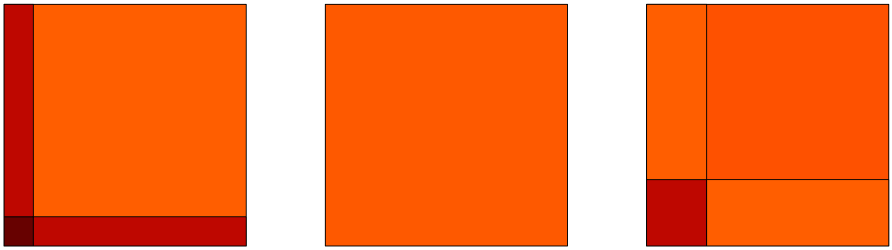

Fig. 2 Optimal graphons (not to scale) at $\varepsilon > \frac{1}{2}$ and at $\tau = \varepsilon^3 + \delta^3$ on the left, $\tau = \varepsilon^3$ in the middle, and $\tau = \varepsilon^3 - \delta^3$ on the right. On the left, the strip at the bottom has height $O(\delta^3)$; on the right it has height $O(\delta)$

of Borgs et al. [3, 4] and Lovász et al. [17–20]. See Sect. 4 in [24] for a discussion of those extremal graphs in terms relevant to this work.

The boundary of the parameter space depicted in Fig. 1 falls naturally into three curves: the upper boundary $\tau = \varepsilon^{3/2}$, the line segment on which $\tau = 0$, and the scalloped curve completed by Razborov et al. On the upper boundary $\tau > \varepsilon^3$, while on the latter two curves $\tau < \varepsilon^3$. The nature of the graphs associated with the points on each curve is similar, but those associated with different curves are not [24].

In this paper we analyze the statistical structure of 'typical' graphs with constraints on (ε, τ) in the *interior* of the parameter space of Fig. 1. We use the graphon formalism to describe asymptotic probabilistic structure, and use the rate function for certain large deviations of Erdős-Rényi graphs to interpret typicality, a notion central to our analysis. We give a careful discussion of typicality in Sect. 2.1, but informally it means 'all but exponentially few' graphs with the given density constraints, exponential in the number of vertices of the system.

Given the known graphons associated with the boundary, we focus on how a typical density-constrained graph changes as the densities (ε, τ) move from one of the three basic components of the boundary to another, in particular as (ε, τ) moves between

$\tau = \varepsilon^{3/2}$ and each of the other two curves: the line segment and the scallops. The statistical structure of a graph typical for constraints (ε, τ) is easily computed from the 'entropy-optimal graphon' associated by large deviations theory with (ε, τ), but we emphasize that as an element in any such study one must establish values of (ε, τ) at which there is a *unique* entropy-optimal graphon, since without uniqueness one cannot really speak of 'typical' behavior associated with (ε, τ). In fact, we have found that proving such uniqueness has been the most difficult part of the analysis, requiring stringent a priori knowledge of the possible optimal graphons. (There is numerical evidence of constraints (ε, τ) supporting multiple entropy-optimal graphons; see for instance the discussion of discontinuous transitions in [14]. There is also a proof, Theorem 5.1 in [13], of nonunique entropy-optimal graphons in the related model with fixed densities of edges and 2-stars.)

Our main result is the explicit determination, for any fixed $1/2 < \varepsilon < 1$, of the unique entropy-optimizing graphon associated with (ε, τ) as τ crosses through $\tau = \varepsilon^3$, and the corollary that the behavior is singular at $\tau = \varepsilon^3$. For a qualitative picture of the singularity see Fig. 2. In terms of large graphs, for these values of (ε, τ) we approximately determine the number of graphs with edge density approximately ε and triangle density approximately τ, and we show that most such graphs have a specific structure. For more details about the connection between entropy-optimizing graphons and large graphs, see Sect. 2.1.

To describe our results more quantitatively we need some notation (with more detail in Sect. 2.1). A graphon $g(x, y)$ on $[0, 1] \times [0, 1]$ is called *bipodal* if there is a decomposition of $[0, 1]$ into 2 intervals ('vertex clusters') C_1 and C_2 and constants a, b, d such that

$$\begin{aligned} g(x, y) = a \quad & \text{if } (x, y) \in C_1 \times C_1 \\ g(x, y) = b \quad & \text{if } (x, y) \in C_2 \times C_2 \\ g(x, y) = d \quad & \text{if } (x, y) \in C_1 \times C_2 \text{ or } (x, y) \in C_2 \times C_1. \end{aligned} \tag{1}$$

We denote the length of C_1 by c.

It is immediate that graphs with independent edges satisfy $\tau = \varepsilon^3$; see Fig. 1. It was previously proven [15] for $0 < \varepsilon < 1/2$ and for $1/2 < \varepsilon < 1$, that the entropy-optimal graphon for τ slightly greater than ε^3 is unique and bipodal and the structure was determined. In this paper we consider the more difficult case of the graphons with τ slightly less than ε^3. For $1/2 < \varepsilon < 1$ we prove that it again is unique and bipodal, and again determine its structure. We also determine the asymptotic behavior of the entropy as (ε, τ) approaches the curve $\tau = \varepsilon^3$, $1/2 < \varepsilon < 1$ from above and from below. Our main results can be summarized as follows, using the function:

$$H(p) = -[p \ln(p) + (1 - p) \ln(1 - p)]. \tag{2}$$

Theorem 1 *There is an open subset $\mathcal{O}_1$ in the planar set of achievable parameters (ε, τ), whose upper boundary is the curve $\tau = \varepsilon^3$, $1/2 < \varepsilon < 1$, such that at (ε, τ) in $\mathcal{O}_1$ there is a unique entropy-optimizing graphon $g_{(\varepsilon,\tau)}$. This graphon is bipodal and for fixed $(\varepsilon, \tau) = (e, e^3 - \delta^3)$, 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* δ *whose leading terms are:*

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

Theorem 2 *There is an open subset* $\mathcal{O}_2$ *in the planar set of achievable parameters* (ε, τ)*, whose lower boundary is the curve* $\tau = \varepsilon^3$, $1/2 < \varepsilon < 1$*, such that at each* (ε, τ) *in* $\mathcal{O}_2$ *there is a unique entropy-optimizing graphon* $g_{(\varepsilon,\tau)}$*. This graphon is bipodal and for fixed* $(\varepsilon, \tau) = (e, e^3 + \Delta\tau)$ *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* $\Delta\tau$ *whose leading terms are:*

$$\begin{aligned} a &= a_0 + O(\Delta\tau) \\ b &= e - \frac{2\Delta\tau}{3e(2e-1)} + O(\Delta\tau^2) \\ c &= \frac{\Delta\tau}{3e(2e-1)^2} + O(\Delta\tau^2) \\ d &= 1 - e + O(\Delta\tau), \end{aligned} \tag{4}$$

where a_0 *is the solution to*

$$H'(a_0) = \left(1 - \frac{2}{e}\right) H'(e). \tag{5}$$

Theorem 3 *The entropy function* $s(\varepsilon, \tau)$ *is real analytic in the variables* ε *and* τ *in the two subsets* $\mathcal{O}_2$ *and* $\mathcal{O}_1$ *of Theorems 1 and 2, which share the boundary curve* $\tau = \varepsilon^3$*. The entropy* $s(\varepsilon, \tau)$ *is continuous on their common boundary but is not differentiable at any point of the curve. As* τ *approaches* ε^3 *from above and below, the following estimates apply:*

- *For fixed* $\varepsilon = e \neq 1/2$*, as* $\tau = e^3 + \Delta\tau$ *approaches* e^3 *from above,*

$$\begin{aligned} s(e, e^3 + \Delta\tau) = H(e) &- \frac{2\Delta\tau}{3e(1-2e)} H'(e) \\ &+ \frac{\Delta\tau^2}{9e^2(1-2e)^4}\Bigg(H(a_0) - H(e) \\ &+ H'(e)\left(3(a_0 - e) + \frac{2(1-2e)}{e}(a_0 + 3e - 2)\right)\Bigg) \end{aligned}$$

$$+\frac{2\Delta\tau^2 H''(e)}{9e^2(1-2e)^2} + O(\Delta\tau^3). \tag{6}$$

- *For fixed $\varepsilon = e \in (1/2, 1)$, as $\tau = e^3 - \delta^3$ approaches e^3 from below,*

$$\begin{aligned} s(e, e^3 - \delta^3) = H(e) &+ \frac{\delta^2}{2e-1} H'(e) - \frac{2\delta^3 \nu}{(2e-1)^2} \\ &+\delta^4 \left(\frac{H'''(e)}{3(2e-1)} + \frac{4\nu}{(2e-1)^3} - \frac{2\nu^2}{eH'(e)(2e-1)^2} \right) \\ &+O(\delta^5), \end{aligned} \tag{7}$$

where $\nu = H'(e) - \left(e - \frac{1}{2}\right) H''(e)$.

In particular, $\partial s/\partial\tau$ diverges as δ^{-1} as τ approaches e^3 from below, as previously shown in [25]. As τ approaches e^3 from above, $\partial s/\partial\tau$ does not *diverge, instead approaching the finite (negative) value $2H'(e)/3e(2e-1)$. Furthermore, the second derivative $\partial^2 s/\partial\tau^2$ is negative.*

Theorem 1 and the second half of Theorem 3 generalize to models where we fix the densities of edges and k-cycles, where k is odd, instead of edges and triangles. The problem actually gets progressively easier as k increases, insofar as our concentration of degree estimates become sharper. Let τ_k denote the density of k-cycles.

Theorem 4 *Let $k > 3$ be an odd integer and let $1/2 < e < 1$.*

- *For sufficiently small $\delta > 0$, the entropy-maximizing graphon subject to the constraints $\varepsilon = e$ and $\tau = e^k - \delta^k$ is bipodal with parameters*

$$\begin{aligned} a &= 1 - e - \delta + O(\delta^2) \\ b &= e - \frac{\delta^2}{2e-1} + O(\delta^3) \\ c &= \frac{\delta}{2e-1} - \frac{2\delta^2}{2e-1} + O(\delta^3) \\ d &= e + \delta + O(\delta^{k-1}). \end{aligned} \tag{8}$$

The entropy is

$$\begin{aligned} s\left(e, e^k - \delta^k\right) = H(e) &+ \frac{\delta^2}{2e-1} H'(e) - \frac{2\delta^3 \nu}{(2e-1)^2} \\ &+\delta^4 \left(\frac{4\nu}{(2e-1)^3} + \frac{H'''(e)}{3(2e-1)} \right) + O(\delta^5). \end{aligned} \tag{9}$$

- *When $\Delta\tau$ is sufficiently small and positive, the optimizing graphon with $\varepsilon = e$ and $\tau_k = e^k + \Delta\tau$ is bipodal with*

$$
\begin{aligned}
a &= a_0 + O(\Delta\tau) \\
b &= e - \frac{2\Delta\tau}{ke^{k-2}(2e-1)} + O(\Delta\tau^2) \\
c &= \frac{\Delta\tau}{ke^{k-2}(2e-1)^2} + O(\Delta\tau^2) \\
d &= 1 - e + O(\Delta\tau),
\end{aligned}
\tag{10}
$$

where a_0 is the solution to (5). The entropy is

$$
s(e, e^3 + \Delta\tau) = H(e) - \frac{2\Delta\tau}{ke^{k-2}(1-2e)} H'(e) + O(\Delta\tau^2). \tag{11}
$$

Most of Theorem 2 was already proven in [15]; here we merely sharpen the estimates. The analysis for Theorem 1, dealing with $\tau < \varepsilon^3$, is the main part of this paper. A substantial portion of our analysis is devoted to proving that any optimal graphon for τ less than but sufficiently close to ε^3, with $1/2 < \varepsilon < 1$, is bipodal. Our argument is quite different from that of [15], which dealt with the case $\tau > \varepsilon^3$, because two of the main techniques in that paper do not apply to undersaturated graphs. Specifically, [15] begins with multipodality for graphs oversaturated with 2-stars, but there are no graphs that are *undersaturated* with 2-stars. Then [15] repeatedly applies the Euler-Lagrange equations; but besides the fact that it seems challenging to rigorously establish Euler-Langrange equations without first knowing multipodality, the Lagrange multipliers for undersaturated graphs are expected to explode as $\delta \to 0$. Finally, there is evidence of a different nature indicating why the situation for $\tau < \varepsilon^3$ is more complicated than for $\tau > \varepsilon^3$. The graphs associated with the boundary curve $\tau = \varepsilon^{3/2}$ are all similar, and [15] proves the same is true for τ just above ε^3, for all $0 < \varepsilon < 1/2$ and $1/2 < \varepsilon < 1$. But in [25] it is proven that for $\varepsilon = 1/2$ and any $0 < \tau < (1/2)^3$ there is a unique optimal graphon, bipodal and similar in kind to those on the boundary curve where $\tau = 0$, but quite *dissimilar* to what we have proven for $\varepsilon > 1/2$. In other words, the two different boundary curves with $\tau < \varepsilon^3$ – the part with $\tau = 0$ and the scallops – generate qualitatively different paths for crossing $\tau = \varepsilon^3$, and so far we have found the case $0 < \varepsilon < 1/2$ impervious to our techniques.

In Sect. 11 we prove Theorem 4. The proof follows the same steps as the proofs of Theorems 1 and 3, only with sharper bounds on the parameter μ. As a result, many of the most difficult steps in the proof of Theorem 1 can be streamlined or avoided entirely. Instead of repeating all the calculations, we merely note the changes needed to adapt the proof for triangles to higher values of k.

One motivation for our study of extremal graphs is an older problem in extremal combinatorics, the densest packing of spheres. In two and three dimensional Euclidean space the densest packing of congruent spheres has been proven, the latter by a celebrated tour de force by Hales et al. [11]. The densest packings have volume fraction $\pi/\sqrt{12} \approx 0.91$ in two dimensions and $\pi/\sqrt{18} \approx 0.74$ in three dimensions and in both cases the optimal packings are highly ordered. It has been an important open problem for many years to prove that a typical packing at fixed volume fraction, in both dimensions, loses its order at a sharp value of volume fraction below the optimum, though this has not yet been proven [21]. The mathematical setup for sphere packings is similar to the one we are using for graphs, with typicality defined through the entropy, the rate function of a large deviation variational principle [10, 16]. In sphere packing the singularity is known as a phase transition.

2 Regularity in the edge/triangle system

2.1 Notation for asymptotics

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 K of G, let $T_K(G)$ be the number of maps from $V(K)$ into $V(G)$ that send edges to edges. The *density* $\tau_K(G)$ of K in G is then defined to be

$$\tau_K(G) := \frac{|T_K(G)|}{n^{|V(K)|}},$$

where $n = |V(G)|$. An important special case is where K is a triangle. We use the letters e and t to denote specific values of the edge density ε and the triangle density τ. For $\alpha > 0$ and $\bar{\tau} = (e, t)$ we define $Z_{\bar{\tau}}^{n,\alpha}$ to be the number of graphs G on n vertices satisfying

$$\varepsilon(G) \in (e - \alpha, e + \alpha), \ \tau(G) \in (t - \alpha, t + \alpha).$$

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:

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

The double limit defining the entropy $s(\bar{\tau})$ is known to exist [24]. 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 [3, 4, 18–20]; see also the recent book [17].) 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 $d(x) = \int_0^1 g(x, y)dy$. The triangle density of a graphon g is $\tau(g) = \int g(x, y)g(y, z)g(z, x)\, dx\, dy\, dz$, and the edge density is $\varepsilon(g) = \int g(x, y)\, dx\, dy$. The *entropy* of a graphon g is $S(g) = \int\int H(g(x, y))\, dx\, dy$, where H is defined as in (2).

The following result is Theorem 3.1 in [24], itself a special case of Theorem 4.1 in [25]:

Theorem 5 (A Variational Principle) *For any values $\bar{\tau} = (e, t)$ in the parameter space we have $s(\bar{\tau}) = \max[S(g)]$, where S is maximized over all graphons g with $\varepsilon(g) = e$ and $\tau(g) = t$.*

The existence of an entropy-maximizing graphon $g = g_{\bar{\tau}}$ for any pair $\bar{\tau}$ of possible densities was proven in [24], again adapting a proof in [7].

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 [17]. 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 [17]. In the remaining sections of the paper, whenever we claim that a graphon has a property (e.g., uniqueness as an entropy maximizer), the caveat "up to relabeling" is implied.

Density-constrained graphons that maximize S, which we call 'entropy maximizing graphons', were introduced in [24] and have been studied slowly but steadily ever since. They can tell us what 'most' or 'typical' large density-constrained graphs are like: if $g_{\bar{\tau}}$ is the only reduced graphon maximizing S with $\bar{\tau}(g) = \bar{\tau}$, then as the number n of vertices diverges and $\alpha_n \to 0$, all but exponentially few graphs with densities $\bar{\tau}_i(G) \in (\tau_i - \alpha_n, \tau_i + \alpha_n)$ will have reduced graphons close to $g_{\bar{\tau}}$ [24]. This is based on large deviations of Erdős-Rényi graphs from [7]. We emphasize that this interpretation requires that the maximizer be unique; this has been difficult to prove in most cases of interest, is responsible for the slow advance of the study of 'typical' density-constrained graphs, and is an important focus of this work.

2.2 Related work

Recent developments in probabilistic combinatorics have dramatically expanded the scope of results like Theorem 5 from [24, 25]: the number of graphs with given edge and subgraph counts can be approximated by the solutions to certain entropic maximization problems [1, 5, 6, 8, 9, 12]. These results are particularly challenging for sparse graphs (i.e. for graphs with n vertices and $m_n = p_n\binom{n}{2}$ edges for $p_n \to 0$). On the other hand, it is quite rare that these entropic maximization problems can be solved explicitly. We only know of one other such case where the optimizers are non-constant graphons: for the upper tail of sparse random graphs, the maximization problem was solved by [2]. Thanks to even more recent results in the theory of large deviations [1, 6, 12], it is now known that for $1 \gg p_n \gg (\log^2 n)/\sqrt{n}$ and any fixed $u > 0$, a random graph G with n vertices and $m_n = p_n\binom{n}{2}$ edges satisfies

$$\Pr\left(\tau(G) \geq (1+u)p_n^3\binom{n}{3}\right) = \exp\left(-(1+o(1))n^2 p_n^2 \log\frac{1}{p_n} \min\left\{\frac{u^{2/3}}{2}, \frac{u}{3}\right\}\right).$$

Moreover, graphs with bipodal structure saturate the bound on the right hand side ([8] call them "clique" or "hub" graphs depending on whether $u^{2/3}/2$ or $u/3$ is smaller). Some structural results in the sparse case are also known: [12] show that conditioned on having triangle density $\tau(G)$ at least $(1+u)p_n^3\binom{n}{3}$, it is likely that the graph contains a large clique, a large hub, or both.

3 Strategy of proof

As discussed above, the proof of Theorem 1 involves two very different sorts of arguments. First, we must show that any optimal graphon is bipodal. Then we must solve the finite-dimensional optimality problem for bipodal graphons. The details are complicated, with many technical estimates, so we present an overview of the argument here.

We start the proof that the optimal graphon is bipodal in Sect. 4. We begin by computing an upper bound for the entropy, based on the fact that any graphon with $\varepsilon = e$ and $\tau = e^3 - \delta^3$ must have

$$\iint [g(x, y) - e]^2 \, dx \, dy \geq \delta^2, \tag{12}$$

and from the maximum possible entropy of any graphon satisfying (12). We then exhibit an explicit "model" bipodal graphon that comes within $O(\delta^3)$ of achieving the upper bound. (This is where the assumption that $e > 1/2$ comes in. The same upper bound applies when $e < 1/2$, but isn't nearly as sharp.) Since (12) is sharp only when $g(x, y) - e$ has rank 1, we conclude that $g - e$ is close to rank 1 and also that it concentrates mainly on two values. This shows that g is close in L^2 to a bipodal graphon: there are well-defined quadrants of the unit square on which g has L^2-small fluctuations.

To show that any optimal graphon is bipodal, we assume that it isn't and then construct explicit competitors by first averaging g on each quadrant (which maintains the edge density while possibly changing the triangle density), and then making small adjustments to some parameters to recover the original triangle density. We aim to show that if g wasn't bipodal to begin with, it would be possible to increase the entropy with such a perturbation.

This step requires estimates on the best bipodal graphon. The space of bipodal graphons is only 4-dimensional, so maximizing the entropy becomes a problem in 4-variable calculus, which we tackle in Sect. 5. We use the constraints on ε and τ to eliminate two of the variables, writing the entropy as a function of the value a of the graphon in one quadrant and a parameter μ that measures how far the degrees are from being constant. Taking derivatives of the entropy with respect to a and μ, and setting them equal to zero, yields the estimates in Theorem 1.

This analysis is complicated by the fact that we do not know, a priori, that the parameters can be expressed as power series in δ. When using a Taylor series to approximate values of the function $H(p)$ near $p = e$ or $p = 1-e$, or when estimating the quantity $\frac{c\Delta a}{1-c}$ in terms of μ, it is not immediately clear which terms must be kept and which can be ignored. We get around this with a bootstrap, using initial estimates to establish which terms must be kept, and then using the revised expansions to get more accurate estimates. In particular, we use the concentration-of-degree estimate from Sect. 4 to claim that $\mu = O(\delta^{3/2})$, which we then use to prove that μ is in fact $O(\delta^2)$. Aside from that concentration-of-degree estimate, this part of the proof is completely independent of the proof that the optimizing graphon is bipodal.

In Sect. 5.3 we turn to the uniqueness of the optimizing bipodal graphon. We compute the Hessian of the entropy with respect to a and μ, obtaining a matrix of the form

$$\begin{pmatrix} K_1\delta^2 + O(\delta^3) & O(\delta^2) \\ O(\delta^2) & K_2\delta^{-1} + O(1) \end{pmatrix},$$

for negative constants K_1 and K_2 (which depend on e but not on δ). Thus the equations for critical points of S, subject to the constraints (3), are well-approximated by a non-singular linear problem, which of course has a unique solution.

Having established the properties of the optimal bipodal graphon, we return in Sects. 6 and 7 to showing that the optimal graphon is in fact bipodal. In Sect. 6 we assume that the optimizing graphon is not bipodal and we estimate the change in the entropy, and the change in the triangle count, from averaging the optimal graphon over each quadrant to obtain a bipodal graphon. In Sect. 7 we use the results of Sect. 5 to show that the parameters of this bipodal graphon can then be perturbed to recover the original triangle count, with higher than the original entropy. Since the original graphon was assumed to be optimal, this is a contradiction, implying that the original optimal graphon was in fact already bipodal.

In other words, the problem of finding the optimal graphon reduces to the finite-dimensional problem of finding the optimal *bipodal* graphon, which we already solved in Sect. 5. The entropy function S is analytic in the parameters (a, b, c, d) and the constraints $(\varepsilon, \tau) = (e, e^3 - \delta^3)$ are analytic (actually algebraic) in (a, b, c, d, e, δ). This implies that the set of critical points is a 2-dimensional analytic variety in $\mathbb{R}^6$. Away from singularities, the implicit function theorem says that (a, b, c, d) are analytic functions of (e, δ). The analysis of Sect. 5.3 shows that there are no singularities in the region defined by (3) where the actual optimal graphon lives, so the parameters of the optimal graphon, and the entropy, are analytic in (e, δ), and therefore in (ε, τ), for τ strictly less than, and sufficiently close to, ε^3. That completes the proof of Theorem 1.

4 Initial approximation

4.1 Notation

Working at a specific edge density $\varepsilon = e$ between $\frac{1}{2}$ and 1, we write g for the graphon $g(x, y)$ and also $g_x(y) = g(x, y)$, and define $\Delta g(x, y) = g(x, y) - e$. We consider $\tau = e^3 - \delta^3$ for sufficiently small δ (depending on e). We take g to be a maximizer of $S(g)$ subject to $\varepsilon(g) = e$ and $\tau(g) = e^3 - \delta^3$. In our asymptotic notation, we treat e as fixed and consider $\delta \to 0$. That is, the hidden constants in $O(\delta)$ are allowed to depend on e, but on nothing else.

Define $D(p) = p \ln(p/e) + (1 - p) \ln[(1 - p)/(1 - e)]$. We will write $\|\cdot\|_2$ for either the L^2 norm on $[0, 1]$ (with respect to Lebesgue measure) or the L^2 norm on $[0, 1]^2$ (with respect to the Lebesgue measure). It should be clear from the context which of these is the case. For a function $h \in L^2([0, 1]^2)$, we write T_h for the integral operator with kernel h:

$$(T_h u)(x) = \int_y h(x, y)u(y)\,dy,$$

which is compact and Hilbert-Schmidt because $h \in L^2$.

The first step towards proving that g is bipodal is to show that g is well approximated in L^2 by a bipodal graphon of a certain form. Specifically, we show:

Proposition 6 *There is a function $v : [0, 1] \to \mathbb{R}$ such that*

(1) *v takes only two values, one of them $\sqrt{2e-1} \pm O(\delta)$, and one of them $\pm O(\delta)$;*
(2) $\int_0^1 v(x)\,dx = 0$*;*
(3) $\|v\|_2^2 = \delta + O(\delta^2)$*;*
(4) $\|\Delta g(x, y) + v(x)v(y)\|_2^2 = \|g(x, y) - (e - v(x)v(y))\|_2^2 = O(\delta^3)$*;*
(5) *If $C_2 = \{x : v(x) = O(\delta)\}$ and $C_1 = [0, 1] \setminus C_2$ then for every $x \in C_2$,*

$$\|g_x - (1-e)\|_{C_1}^2 \geq \|g_x - e\|_{C_1}^2 - O(\delta^2); \tag{13}$$

and

(6) *for every $x \in C_1$,*

$$\|g_x - (1-e)\|_{C_1}^2 \leq \|g_x - e\|_{C_1}^2 + O(\delta^2). \tag{14}$$

The point here is that $e - v(x)v(y)$ is a bipodal graphon with triangle density $e - \|v\|_2^6 = e - \delta^3 \pm O(\delta^4)$; Proposition 6 shows that g is close to this bipodal graphon in L^2. The regions C_2 and C_1 are called *podes*. C_1 is the small pode and C_2 is the big pode, and the last two points of Proposition 6 show that no $x \in [0, 1]$ is classified into a clearly wrong pode.

4.2 Entropy cost

Our first step towards Proposition 6 is a pretty good estimate (at least for small δ) for the entropy cost of reducing the triangle density.

Define

$$C(e) = \frac{\ln \frac{e}{1-e}}{2e-1}. \tag{15}$$

The relevance of $C(e)$ is that it characterizes the optimal trade-off, in some sense, between entropy and L^2 mass.

Lemma 7

$$C(e) = \inf_{p \in [0,1]} \frac{D(p)}{(p-e)^2}.$$

Moreover, the infimum above is uniquely attained at $p = 1-e$*, and it is a second-order minimum in the sense that for any* $e \in (0, 1)$ *there is a constant* $c(e) > 0$ *such that for any* $p \in [0, 1]$,

$$\frac{D(p)}{(p-e)^2} \geq C(e) + c(e)(p-(1-e))^2.$$

The proof of Lemma 7 can be found in [22].

Recall that $H(e)$ is the entropy of the constant graphon. By the concavity of H, $S(g) < H(e)$; the following lemma gives a bound on just how much smaller it must be.

Lemma 8

$$C(e)\delta^2 \leq C(e)\|\Delta g\|_2^2 \leq H(e) - S(g) \leq C(e)\delta^2 + O(\delta^3).$$

Proof Note that $\int D(g) = H(e) - \int H(g) = H(e) - S(g)$. By Lemma 7,

$$\int D(g) \geq \frac{D(1-e)}{(2e-1)^2} \int (g-e)^2 = C(e)\|\Delta g\|_2^2.$$

Moreover, $\mathrm{Tr}[(T_{\Delta g})^3] = \tau(g) - e^3 - 3e \int (d(x)-e)^2\, dx \leq \tau(g) - e^3 = -\delta^3$, and Cauchy-Schwarz implies that $\delta^3 \leq -\mathrm{Tr}[(T_{\Delta g})^3] \leq \|\Delta g\|_2^3$. This proves the first two claimed inequalities.

To prove the final inequality, we construct a graphon having triangle density $e - \delta^3$ and entropy cost $C(e)\delta^2 + O(\delta^3)$, and then the inequality follows from the fact that g has minimal entropy cost among graphons with triangle density $e - \delta^3$. Let $v(x)$ be a function taking the value $-\sqrt{2e-1}$ on a set of measure a, and the value $\sqrt{2e-1}\,a/(1-a)$ on a set of measure $1-a$. Then $\int v = 0$ and $\int v^2 = (2e-1)a/(1-a)$. For the graphon $h(x, y) = e - v(x)v(y)$ (which has edge density e), the fact that our perturbation is rank-1 (and orthogonal to constants) implies that

$$\tau(h) - e^3 = -\|v\|_2^6 = -(2e-1)^3 \frac{a^3}{(1-a)^3}.$$

Now we fix a so that $\tau(h) = e^3 - \delta^3$. Then $a = \delta/(2e-1) + O(\delta^2)$ as $\delta \to 0$, and

$$\begin{aligned} H(e) - S(h) = \int D(h) &= a^2 D(1-e) + (1-a)^2 D(e - \Theta(a^2)) \\ &\quad + 2a(1-a)D(e + \Theta(a)). \end{aligned}$$

Since $D(e) = D'(e) = 0$ and D is twice differentiable, we conclude that

$$\begin{aligned} H(e) - S(h) &= a^2 D(1-e) + O(a^3) = a^2(2e-1)\ln\frac{e}{1-e} + O(a^3) \\ &= C(e)\delta^2 + O(\delta^3). \end{aligned}$$

Recalling that $S(g) \geq S(h)$, this completes the proof of the claimed upper bound. □

4.3 Closeness to ideal values

We saw that $\|\Delta g\|_2^2 = \|g - e\|_2^2 = O(\delta^2)$, but we can get a better bound if we look at the distance of g from *either* e or $1 - e$. A useful interpretation of this is that most of the L^2 mass of Δg is spent at values near $1 - 2e$. This notion – that most of the L^2 mass of something is spent near some particular value – will be used repeatedly. We will therefore study some basic properties of this notion.

Let

$$V_a(x) = \min\left\{x^2, (x-a)^2\right\}. \tag{16}$$

The point of this definition is that "most of the L^2 mass of u is near a" can be encoded as $\int V_a(u) \ll \int u^2$. The basic homogeneity property of V_a is that for any $a, x \in \mathbb{R}$, $V_a(x) = a^2 V_1(x/a)$. This means that it mostly suffices to study properties of V_1.

Next, we show two stability properties: the notion of mass concentration is stable under small perturbations of the function u, and also under small changes to the ideal value a.

Lemma 9 *For any $u, w \in L_2(\mu)$, $\int V_1(u+w)\,d\mu \leq 2\int V_1(u)\,d\mu + 2\|w\|_{L^2(\mu)}^2$.*

Proof If $u^2(x) \leq (u(x)-1)^2$ then $(u(x)+w(x))^2 \leq 2u^2(x)+2w^2(x) = 2V_1(u(x))+2w^2(x)$. Similarly, if $(u(x)-1)^2 \leq u^2(x)$ then $(u(x)-1+w(x))^2 \leq 2(u(x)-1)^2+2w^2(x) = 2V_1(u(x))+2w^2(x)$. Taking the minimum of these two inequalities, $V_1(u(x)+w(x)) \leq 2V_1(u(x))+2w^2(x)$, and the claim follows by integrating. □

Lemma 10 *For any $u \in L_2(\mu)$ and any $\eta \geq 0$,*

$$\int V_{1+\eta}(u)\,d\mu \leq \int V_1(u)\,d\mu + 4\eta(1+\eta)\|u\|_{L^2(\mu)}^2.$$

Proof If $u \leq 1/2$ then $V_{1+\eta}(u) = V_1(u) = u^2$. On the other hand, if $u \geq 1/2$ then $V_{1+\eta}(u) \leq (u-1-\eta)^2 \leq (u-1)^2+\eta^2+\eta = V_1(u)+\eta^2+\eta$. Markov's inequality implies that $\mu\{u \geq 1/2\} \leq 4\|u\|_{L^2(\mu)}^2$, and so

$$\begin{aligned}\int V_{1+\eta}(u)\,d\mu &\leq \int V_1(u) + 1_{\{u\geq\frac{1}{2}\}}(\eta^2+\eta)\,d\mu \\ &\leq \int V_1(u)\,d\mu + 4\eta(1+\eta)\|u\|_{L^2(\mu)}^2.\end{aligned}$$

□

Our final property of V will allow us to show that if $u(x)+u(y)$ puts most of its mass near 1, then the same is true of u.

Lemma 11 *Let u be a function on a finite measure space (Ω, μ). If $\|u\|_{L^2(\mu)}^2 \leq \mu(\Omega)/32$ and*

$$\int_{\Omega\times\Omega} V_1(u(x)+u(y))\, d\mu^{\otimes 2} \leq K$$

for some $0 < K < \mu(\Omega)^2/256$ then

$$\int_{\Omega} V_1(u(x))\, d\mu \leq \frac{CK}{\mu(\Omega)}$$

for some universal constant C.

Note that the restriction $\|u\|_{L^2(\mu)}^2 \leq \mu(\Omega)/32$ is required to rule out the situation where u is close to the constant function $1/2$, in which case our desired conclusion wouldn't hold.

Proof We can assume without loss of generality that μ is a probability measure. Let $A = \{x : u(x) \leq 1/4\}$. By Markov's inequality and the fact that $\int_\Omega u^2\, d\mu \leq 1/32$, $\mu(A) \geq 1/2$. On $A\times A$, $u(x)+u(y) \leq 1/2$ and so $V_1(u(x)+u(y)) = (u(x)+u(y))^2$. Hence,

$$\int_{A\times A} V_1(u(x)+u(y))\, d\mu^{\otimes 2} = \int_{A\times A} (u(x)+u(y))^2 \geq 2\mu(A)\int_A u^2(x)\, d\mu.$$

On the other hand,

$$\int_{A\times A} V_1(u(x)+u(y))\, d\mu^{\otimes 2} \leq K,$$

and so

$$\int_A V_1(u(x))\, d\mu = \int_A u^2(x)\, d\mu \leq \frac{K}{2\mu(A)} \leq K. \tag{17}$$

Since $K \leq 1/256$, applying Markov's inequality to $u 1_A$ gives

$$\mu\left(\left\{|u1_A| \geq \frac{1}{8}\right\}\right) \leq 64\int_A u^2\, d\mu \leq 64K \leq \frac{1}{4},$$

meaning that $\mu(\{|u| \leq 1/8\}) \geq 1/4$; let $B = \{x : |u(x)| \leq 1/8\}$. For $y \notin A$ and $x \in B$, $u(x)+u(y) \geq 1/4 - 1/8 = 1/8$, meaning that $(u(x)+u(y)-1)^2 \leq 64V_1(u(x)+u(y))$. Hence,

$$\begin{aligned}\int_{B\times A^c} (u(x)+u(y)-1)^2\, d\mu(x)\, d\mu(y) &\leq 64\int_{\Omega\times\Omega} V_1(u(x)+u(y))\, d\mu(x)\, d\mu(y)\\ &\leq 64K.\end{aligned}$$

On the other hand, Cauchy-Schwarz gives $(u(x)+u(y)-1)^2 \geq (u(y)-1)^2/2-u^2(x)$, and so

$$\begin{aligned}\int_{B\times A^c} (u(x)+u(y)-1)^2\, d\mu(x)\, d\mu(y) &\geq \frac{1}{2}\mu(B)\int_{A^c}(u(y)-1)^2\, d\mu(y)\\ &\quad -\mu(A^c)\int_B u^2(x)\, d\mu(x)\\ &\geq \frac{1}{8}\int_{A^c}(u(y)-1)^2\, d\mu(y) - K,\end{aligned}$$

where we used (17) for the last inequality, noting that $B\subseteq A$. Rearranging, we have

$$\frac{1}{8}\int_{A^c}(u(y)-1)^2\, d\mu(y) \leq 65K.$$

Since $V_1(u(y)) \leq 16(u(y)-1)^2$ for $y \notin A$, this shows that

$$\int_{A^c} V_1(u(y))\, d\mu(y) \leq 520K. \tag{18}$$

Combined with (17), this completes the proof. □

We return to studying the perturbations of our graphon g. In Lemma 8, we saw that the most entropy-efficient way to perturb e in L^2 was to set some of the values of g to $1-e$, which is equivalent to having Δg equal to $1-2e$ at some points. We strengthen this by showing that most of the mass of Δg must be spent near $1-2e$.

Lemma 12

$$\int V_{1-2e}(\Delta g(x,y))\, dx\, dy = O(\delta^3).$$

Proof By the second part of Lemma 7,

$$\begin{aligned}H(e)-S(g) = \int D(g) &= \int \frac{D(g)}{(g-e)^2}(g-e)^2\\ &\geq C(e)\int (1+\Theta((g-(1-e))^2))(g-e)^2\\ &= C(e)\|\Delta g\|_2^2 + \Theta\left(\int (g-(1-e))^2(g-e)^2\right)\\ &= C(e)\|\Delta g\|_2^2 + \Theta\left(\int V_{1-2e}(\Delta g)\right).\end{aligned}$$

On the other hand, Lemma 8 implies that

$$H(e)-S(g) \leq C(e)\delta^2 + O(\delta^3) \leq C(e)\|\Delta g\|_2^2 + O(\delta^3),$$

and comparing this to the previous bound proves the claim. □

4.4 Concentration of degrees

We define the "degree" of $x \in [0, 1]$ to be $d(x) = \int g(x, y)\, dy$. Note that $\int d(x)\, dx = \varepsilon(g) = e$. It turns out that having a non-constant degree function increases the triangle density, so our optimal graphon g must have an almost-constant degree function.

Lemma 13

$$\int (d(x) - e)^2\, dx = O(\delta^4)$$

Recalling that $\delta^4 = \Theta(\|\Delta g\|_2^4)$, this is better than the trivial bound (coming from Jensen's inequality) of

$$\int (d(x) - e)^2\, dx \le \|\Delta g\|_2^2.$$

Proof Start by observing that

$$\mathrm{Tr}[(T_{\Delta g})^3] = t - e^3 + 3e^3 - 3e \int d^2(x)\, dx = -\delta^3 + 3e \int (d(x) - e)^2\, dx.$$

Cauchy-Schwarz gives $\mathrm{Tr}[(T_{\Delta g})^3] \ge -\|\Delta g\|_2^3$, and so

$$\delta^3 \le \|\Delta g\|_2^3 - 3e \int (d(x) - e)^2\, dx.$$

By the concavity of the function $t \mapsto t^{2/3}$, if $s < t$ then $(t - s)^{2/3} \le t^{2/3} - \frac{2}{3} t^{-1/3} s$. Therefore,

$$\delta^2 \le \|\Delta g\|_2^2 - \frac{2e \int (d(x) - e)^2\, dx}{\|\Delta g\|_2}.$$

Comparing this to Lemma 8 gives

$$\begin{aligned} C(e)\|\Delta g\|_2^2 &\le H(e) - S(g) \le C(e)\delta^2 + O(\delta^3) \\ &\le C(e)\|\Delta g\|_2^2 - 2eC(e)\frac{\int (d(x) - e)^2\, dx}{\|\Delta g\|_2} + O(\delta^3), \end{aligned}$$

and so we conclude that

$$\int (d(x) - e)^2\, dx \le O(\|\Delta g\|_2 \delta^3) = O(\delta^4).$$

□

4.5 Rank

In this section, we will prove Proposition 6. We'll start by just considering an eigenfunction (which will not necessarily take only two values). Later, we'll round it.

Lemma 14 *There is a function $\tilde{v}(x)$ such that*

$$\|\Delta g(x, y) + \tilde{v}(x)\tilde{v}(y)\|_2^2 = O(\delta^3).$$

Proof Recall that

$$H(e) - S(g) \geq C(e)\|\Delta g\|_2^2, \tag{19}$$

and then we used the fact that (if λ_i are the eigenvalues of $T_{\Delta g}$) $\sum_i \lambda_i^2 \geq (\sum_i |\lambda_i|^3)^{2/3}$ to compare this to δ^3. We can sharpen this eigenvalue comparison: if we write the eigenvalues λ_i so that their absolute values are non-increasing, and if $\epsilon > 0$ is chosen so that $\sum_{i\geq 2} \lambda_i^2 = \epsilon\lambda_1^2$, then

$$\|\lambda\|_3^3 \leq \|\lambda\|_\infty \|\lambda\|_2^2 = |\lambda_1|\|\lambda\|_2^2 \leq \frac{\|\lambda\|_2^3}{\sqrt{1+\epsilon}}$$

and so

$$\|\Delta g\|_2^2 = \|\lambda\|_2^2 \geq (1+\epsilon)^{1/3}\|\lambda\|_3^2.$$

Recalling from Lemma 13 that $\|\lambda\|_3^3 \geq -\mathrm{Tr}[(T_{\Delta g})^3] = \delta^3 + 3e\|d-e\|_2^2$ and that $\|d-e\|_2^2 = O(\delta^4)$, we have

$$\|\Delta g\|_2^2 \geq (1+\epsilon)^{1/3}(\delta^3 - O(\delta^4))^{2/3} = \delta^2 + \Omega(\epsilon\delta^2) - O(\delta^3).$$

Combining this estimate with (19) gives

$$H(e) - S(g) \geq C(e)\delta^3 + \Omega(\epsilon\delta^2) - O(\delta^3).$$

Compared to Lemma 8, this shows that $\epsilon = O(\delta)$. In other words, we have $\sum_{i\geq 2} \lambda_i^2 = O(\delta\lambda_1^2)$.

On the other hand, $\sum_{i\geq 1} \lambda_i^2 = \|\Delta g\|_2^2 = \Theta(\delta^2)$, and so $\lambda_1^2 = \Theta(\delta^2)$ and $\sum_{i\geq 2} \lambda_i^2 = O(\delta^3)$. In particular, if $u(x)$ is an eigenfunction of Δg with eigenvalue λ_1, normalized so that $\|u\|_2 = 1$, then

$$\|\Delta g - \lambda_1 u(x)u(y)\|_2^2 = O(\delta^3).$$

Finally, note that $\lambda_1 < 0$, because $\sum_i \lambda_i^3 = t - e^3 + O(\delta^4) = -\delta^3 + O(\delta^4)$, and since $\sum_{i\geq 2} |\lambda_i|^3 \leq \left(\sum_{i\geq 2} \lambda_i^2\right)^{3/2} = O(\delta^{9/2})$, we must have $\lambda_1^3 = -\delta^3 + O(\delta^4)$. Setting $\tilde{v}(x) = \sqrt{|\lambda_1|}u(x)$ completes the proof. □

From now on, we fix a function $\tilde{v}$ satisfying Lemma 14. The following bound just comes from combining Lemma 14 with Lemma 12 and the triangle inequality (in the form of Lemma 9).

Corollary 15

$$\int V_{2e-1}(\tilde{v}(x)\tilde{v}(y))\,dx\,dy = O(\delta^3).$$

Our next goal is to show that we can replace $\tilde{v}(x)$ by a rounded version. We'll start by ignoring the sign of $\tilde{v}$.

Lemma 16 *Let $\bar{v}(x)$ be either 0 or $\sqrt{2e-1}$, whichever is closer to $|\tilde{v}(x)|$. There is a universal constant C such that*

$$\|\bar{v}(x)\bar{v}(y) - |\tilde{v}(x)\tilde{v}(y)|\|_2^2 \le C \int V_{2e-1}(\tilde{v}(x)\tilde{v}(y))\,dx\,dy.$$

Proof Let $c = \sqrt{2e-1}$, and let $w(x, y)$ be either 0 or c^2, whichever is closer to $|\tilde{v}(x)\tilde{v}(y)|$. Since $\bar{v}(x)\bar{v}(y)$ is always either 0 or c^2, we have the pointwise bound

$$\Big| |\tilde{v}(x)\tilde{v}(y)| - w(x, y) \Big| \le \Big| |\tilde{v}(x)\tilde{v}(y)| - \bar{v}(x)\bar{v}(y) \Big|.$$

Our first goal is to show the reverse inequality for most points x and y:

$$\Big| |\tilde{v}(x)\tilde{v}(y)| - \bar{v}(x)\bar{v}(y) \Big| \le C \Big| |\tilde{v}(x)\tilde{v}(y)| - w(x, y) \Big| = C V_{2e-1}(\tilde{v}(x)\tilde{v}(y)). \quad (20)$$

Let

$$A_1 = \left\{ x : |\tilde{v}(x)| \le \frac{1}{2}c \right\} = \{\bar{v} = 0\}$$
$$A_2 = \left\{ x : |\tilde{v}(x)| \ge \frac{4}{3}c \right\}.$$

If $x \in A_1$ then $\bar{v}(x)\bar{v}(y) = 0$ no matter the value of y. Now if $x \in A_1$ and $|\tilde{v}(y)| \le c$ then $w(x, y) = 0$ and so $\bar{v}(x)\bar{v}(y) = w(x, y)$. If $x \in A_1$ and $c < |\tilde{v}(y)| \le 4c/3$ then $w(x, y) = c^2$ but $|\tilde{v}(x)\tilde{v}(y)| \le 2c^2/3$, meaning that

$$\Big| |\tilde{v}(x)\tilde{v}(y)| - w(x, y) \Big| \ge \frac{c^2}{3} \ge \frac{1}{2}|\tilde{v}(x)\tilde{v}(y)| = \frac{1}{2}\Big| \bar{v}(x)\bar{v}(y) - |\tilde{v}(x)\tilde{v}(y)| \Big|$$

To summarize: if $x \in A_1$ and $y \notin A_2$ then (20) holds with $C = 2$. Of course, the same holds if $y \in A_1$ and $x \notin A_2$.

If $x, y \notin A_1$ then $\bar{v}(x)\bar{v}(y) = c^2$ and $w(x, y)$ could be either 0 or c^2. If $w(x, y) = c^2$ then $\bar{v}(x)\bar{v}(y) = w(x, y)$ and so (20) holds with $C = 1$; while if $w(x, y) = 0$ then

$$\Big||\tilde{v}(x)\tilde{v}(y)| - w(x,y)\Big| = |\tilde{v}(x)\tilde{v}(y)| \geq \frac{1}{8}(|\tilde{v}(x)\tilde{v}(y)| + c^2)) \geq \frac{1}{8}\Big||\tilde{v}(x)\tilde{v}(y)| - \bar{v}(x)\bar{v}(y)\Big|,$$

and so (20) holds with $C = 8$.

The only case where we have *not* shown (20) is when $x \in A_1$ and $y \in A_2$ (or vice versa). In particular, by integrating out (20) on every set where we have proven it, we get

$$\int_{[0,1]^2\setminus((A_1\times A_2)\cup(A_2\times A_1))} (|\tilde{v}(x)\tilde{v}(y)| - \bar{v}(x)\bar{v}(y))^2\,dx\,dy \leq C\||\tilde{v}(x)\tilde{v}(y)| - w\|_2^2. \tag{21}$$

Finally, we consider the case $x \in A_1$, $y \in A_2$: here the pointwise bound (20) is not necessarily true, and so we give an integral bound instead. Note that

$$\begin{aligned}\||\tilde{v}(x)\tilde{v}(y)| - w\|_2^2 &\geq \int_{A_1\times A_1} (|\tilde{v}(x)\tilde{v}(y)| - w(x,y))^2\,dx\,dy \\ &= \int_{A_1\times A_1} \tilde{v}^2(x)\tilde{v}^2(y)\,dx\,dy \\ &= \left(\int_{A_1} \tilde{v}^2(x)\,dx\right)^2\end{aligned}$$

and similarly

$$\begin{aligned}\||\tilde{v}(x)\tilde{v}(y) - w\|_2^2 &\geq \int_{A_2\times A_2} (|\tilde{v}(x)\tilde{v}(y)| - w(x,y))^2\,dx\,dy \\ &= \int_{A_2\times A_2} (|\tilde{v}(x)\tilde{v}(y)| - c^2)^2\,dx\,dy \\ &\geq \int_{A_2\times A_2} (|\tilde{v}(x)\tilde{v}(y)| - \frac{9}{16}|\tilde{v}(x)||\tilde{v}(y)|)^2\,dx\,dy \\ &\geq \frac{1}{16}\int_{A_2\times A_2} \tilde{v}^2(x)\tilde{v}^2(y)\,dx\,dy \\ &= \left(\frac{1}{4}\int_{A_2} \tilde{v}^2(x)\,dx\right)^2.\end{aligned}$$

So then

$$\begin{aligned}\int_{A_1\times A_2} (|\tilde{v}(x)\tilde{v}(y)| - \bar{v}(x)\bar{v}(y))^2\,dx\,dy &= \int_{A_1\times A_2} \tilde{v}^2(x)\tilde{v}^2(y)\,dx\,dy \\ &= \int_{A_1} \tilde{v}^2(x)\,dx \int_{A_2} \tilde{v}^2(y)\,dy\end{aligned}$$

$$\leq 4\||\tilde{v}(x)\tilde{v}(y)| - w\|_2^2.$$

Of course, we have the same bound if we replace $A_1 \times A_2$ with $A_2 \times A_1$. Together with (21), this shows that

$$\||\tilde{v}(x)\tilde{v}(y)| - \bar{v}(x)\bar{v}(y)\|_2^2 \leq C\||\tilde{v}(x)\tilde{v}(y)| - w\|_2^2.$$

This is almost the same as the claim; the difference is that the right hand side above is

$$C \int \min\left\{\tilde{v}^2(x)\tilde{v}^2(y), (|\tilde{v}(x)\tilde{v}(y)| - c^2)^2\right\} dx\, dy,$$

whereas right hand side in the claim has no absolute values. But since $\Big||v(x)v(y)| - c^2\Big| \leq |v(x)v(y) - c^2|$, the claim follows. □

Next, we handle the signs. Of course, $\tilde{v}(x)$ can be negated without changing $\tilde{v}(x)\tilde{v}(y)$, but the rounding to $\{0, \sqrt{2e-1}\}$ is affected by the sign. Therefore, we may need to replace $\tilde{v}$ by $-\tilde{v}$ in order to give bounds for the rounded version.

Lemma 17 *After possibly replacing $\tilde{v}$ by $-\tilde{v}$, the following holds. Let $v(x)$ be either 0 or $\sqrt{2e-1}$, whichever is closer to $\tilde{v}(x)$. Then*

$$\|v(x)v(y) - \tilde{v}(x)\tilde{v}(y)\|_2^2 \leq C \int V_{2e-1}(\tilde{v}(x)\tilde{v}(y))\, dx\, dy.$$

Proof Let $c = \sqrt{2e-1}$ and recall the definition of $\bar{v}$ from Lemma 16: $\bar{v}(x)$ is either 0 or c, whichever is closer to $|\tilde{v}(x)|$. In particular,

$$\begin{aligned}
&v(x)v(y) = \bar{v}(x)\bar{v}(y) && \text{if } \tilde{v}(x) \geq -\frac{c}{2} \text{ and } \tilde{v}(y) \geq -\frac{c}{2}\\
&v(x)v(y) = \bar{v}(x)\bar{v}(y) = 0 && \text{if } |\tilde{v}(x)| \leq \frac{c}{2} \text{ or } |\tilde{v}(y)| \leq \frac{c}{2}\\
&v(x)v(y) = 0 \neq c^2 = \bar{v}(x)\bar{v}(y) && \text{if } \tilde{v}(x) < -\frac{c}{2} \text{ and } |\tilde{v}(y)| > \frac{c}{2}, \text{ or vice versa.}
\end{aligned}$$

Let $A_1 = \{x : \tilde{v}(x) < -c/2\}$ and $A_2 = \{x : \tilde{v}(x) > c/2\}$, so that $v(x)v(y) \neq \bar{v}(x)\bar{v}(y)$ only on $A_1 \times A_1$, $A_1 \times A_2$, and $A_2 \times A_1$. Hence,

$$\begin{aligned}
&\|\tilde{v}(x)\tilde{v}(y) - v(x)v(y)\|_2^2 - \|\tilde{v}(x)\tilde{v}(y) - \bar{v}(x)\bar{v}(y)\|_2^2\\
&\quad\leq \int_{A_1 \times A_1} (\tilde{v}(x)\tilde{v}(y) - v(x)v(y))^2\, dx\, dy\\
&\qquad + 2\int_{A_1 \times A_2} (\tilde{v}(x)\tilde{v}(y) - v(x)v(y))^2\, dx\, dy\\
&\quad= \left(\int_{A_1} v^2(x)\, dx\right)^2 + 2\int_{A_1} v^2(x)\, dx \int_{A_2} v^2(x)\, dx.
\end{aligned}$$

Because $\bar{v}$ is non-negative, $|\tilde{v}(x)\tilde{v}(y) - \bar{v}(x)\bar{v}(y)| \leq \Big||\tilde{v}(x)\tilde{v}(y)| - \bar{v}(x)\bar{v}(y)\Big|$, and then Lemma 16 implies that

$$\|\tilde{v}(x)\tilde{v}(y) - v(x)v(y)\|_2^2 \leq C \int V_{2e-1}(\tilde{v}(x)\tilde{v}(y))\,dx\,dy + \left(\int_{A_1} v^2(x)\,dx\right)^2 + 2\int_{A_1} v^2(x)\,dx \int_{A_2} v^2(x)\,dx \tag{22}$$

and it remains to bound the last two terms.

Now, on $A_1 \times A_2$, we have $\tilde{v}(x)\tilde{v}(y) \leq -c^2/4$, meaning that $V_{2e-1}(v(x)v(y)) = v^2(x)v^2(y)$. Hence,

$$\int_{[0,1]^2} V_{2e-1}(v(x)v(y))\,dx\,dy \geq \int_{A_1\times A_2} v^2(x)v^2(y)\,dx\,dy = \int_{A_1} v^2(x)\,dx \int_{A_2} v^2(y)\,dy.$$

Moreover, we may assume that $\int_{A_1} v^2(x)\,dx \leq \int_{A_2} v^2(y)\,dy$ (if not, this becomes true when we replace v by $-v$). Then we can remove the last term of (22) at the cost of increasing C by 3. □

Proof of Proposition 6 Let u be the function that we called v in Lemma 17, and define $v = u - \int u\,dx$. Then v trivially satisfies item 2.

By Lemma 17, Lemma 14 and the triangle inequality,

$$\|u(x)u(y) - \Delta g\|_2^2 \leq C \int V_{2e-1}(\tilde{v}(x)\tilde{v}(y))\,dx\,dy + O(\delta^3) = O(\delta^3),$$

with the second inequality coming from Corollary 15. Since $\|\Delta g\|_2^2 = \delta^2 \pm O(\delta^3)$ by Lemma 8, we have

$$\|\Delta g\|_2 - \|u(x)u(y) - \Delta g\|_2 \leq \|u(x)u(y)\|_2 = \|u\|_2^2 \leq \|\Delta g\|_2 + \|u(x)u(y) - \Delta g\|_2,$$

and so $\|u\|_2^2 = \delta \pm O(\delta^2)$. Since u only takes 2 values (one of them 0, one of them bounded away from zero), we have $a := \int u\,dx = O(\delta)$. Hence $\|v\|_2^2 = \|u\|_2^2 - a^2 = \delta \pm O(\delta^2)$, proving item 3. The bound $a = O(\delta)$ also proves item 1.

For item 4,

$$\|v(x)v(y) - \Delta g\|_2 \leq \|u(x)u(y) - \Delta g\|_2 + \|v(x)v(y) - u(x)u(y)\|_2;$$

the first term on the right is $O(\delta^{3/2})$, while the second term is $\|a^2 - 2au\|_2 = O(\delta^{3/2})$ also.

Having proven the first four claims, we will show that the last two claims follow by perturbing C_2 and C_1 slightly. That is, we are going to redefine v: it will still take two values, but we will change the sets on which it takes those two values, and then we will recenter it to maintain the property $\int v = 0$. If we only change the values of v on a set of size $O(\delta^2)$, the triangle inequality implies that items 1–4 still hold for the modified function v.

Let $A_2 \subset C_1$ consist of those x for which

$$\int_{C_1} (g(x, y) - (1 - e))^2 \, dy > \int_{C_1} (g(x, y) - e)^2 \, dy,$$

and let $A_1 \subset C_2$ consist of those x for which

$$\int_{C_1} (g(x, y) - (1 - e))^2 \, dy < \int_{C_1} (g(x, y) - e)^2 \, dy.$$

Note that for $x, y \in C_1$ we have $g(x, y) = (1 - e) + \Theta(\delta) + r(x, y)$, and so if $x \in A_2$ then $\int_{C_1} r(x, y)^2 \, dy = \Omega(\delta)$. Similarly, if $x \in C_2$ and $y \in C_1$ then $g(x, y) = e + \Theta(\delta) + r(x, y)$ and so if $x \in A_1$ then $\int_{C_1} r(x, y)^2 \, dy = \Omega(\delta)$. Since $\|r\|_2^2 = O(\delta^3)$, it follows that $|A_1 \cup A_2| = O(\delta^2)$.

We now redefine Ω_i as follows: set $\tilde{C}_2 = C_2 \cup A_2 \setminus A_1$ and $\tilde{C}_1 = C_1 \cup A_1 \setminus A_2$. Since g is bounded and $|\Omega_i \Delta \tilde{\Omega}_i| = O(\delta^2)$, (13) and (14) hold with $\tilde{\Omega}_i$ in place of Ω_i. Redefining the function v to be $\sqrt{2e - 1} + O(\delta)$ on $\tilde{C}_1$ and $O(\delta)$ on $\tilde{C}_2$ (with the $O(\delta)$ terms chosen so that $\int v = 0$), Proposition 6 holds with this modified function v. □

5 Estimating the bipodal parameters

At this point, we have only shown that an entropy-optimal triangle-deficient graphon is *approximately* bipodal. For this section, we will temporarily switch to studying truly bipodal graphons. Up to measure-preserving transformations of [0, 1], every bipodal graphon takes the form

$$g(x, y) = \begin{cases} a & x, y < c, \\ b & x, y > c, \\ d & x < c < y \text{ or } y < c < x. \end{cases}$$

That is, the pode sizes are c and $1 - c$, and without loss of generality we can assume that $c \leq 1 - c$. We define

$$\Delta a = a - e, \qquad \Delta b = b - e, \qquad \Delta d = d - e, \qquad \mu = \frac{c \Delta a}{1 - c} + \Delta d.$$

The main result of this section is that in the class of bipodal graphons with edge density e, triangle density $\tau = e^3 - \delta^3$, and parameters $\Delta b = o(1)$, $c = o(1)$, $\Delta d = o(1)$, $\mu = o(\delta)$, and $|\Delta a| = \Omega(1)$, there is a unique entropy-optimal bipodal graphon and we get good estimates on its parameters.

To be precise, Let $\mathcal{G}_{e,\delta,\eta}$ be the set of bipodal graphons with edge-density e, triangle density $e^3 - \delta^3$, and parameters a, b, c and d satisfying $|b-e| < \eta\sqrt{\delta}$, $c < \eta$, $|d-e| < \eta$, $|a-e| > \eta$, and $|\mu| = O(\delta^{3/2})$.

Proposition 18 *For every $\frac{1}{2} < e < 1$ there exists $\eta > 0$ such that for any $\delta < \eta$ there is at most one graphon $g \in \mathcal{G}_{e,\delta,\eta}$ maximizing $S(g)$. Moreover, any such optimal graphon has parameters satisfying*

$$\begin{aligned} a &= 1 - e - \delta + O(\delta^2) \\ b &= e - \frac{\delta^2}{2e-1} + O(\delta^3) \\ c &= \frac{\delta}{2e-1} - \frac{2\delta^2}{2e-1} + O(\delta^3) \\ d &= e + \delta + \frac{\delta^2}{eH'(e)}\left(H'(e) - \left(e - \frac{1}{2}\right)H''(e)\right) + O(\delta^3). \end{aligned} \tag{23}$$

To prove Proposition 18, we first show that the parameters of any optimal graphon must satisfy the claimed estimates, so assume that g is optimal. We define

$$\Delta a = a - e, \qquad \Delta b = b - e, \qquad \Delta d = d - e, \qquad \mu = \frac{c\Delta a}{1-c} + \Delta d.$$

The edge density, triangle density and entropy of the graphon are:

$$\varepsilon(g) = c^2 a + 2c(1-c)d + (1-c)^2 b, \tag{24}$$

$$\tau(g) = c^3a^3 + 3c^2(1-c)ad^2 + 3c(1-c)^2bd^2 + (1-c)^3b^3, \tag{25}$$

$$S(g) = c^2H(a) + 2c(1-c)H(d) + (1-c)^2H(b). \tag{26}$$

5.1 Expressing all quantities in terms of a and μ

Our constraint on edge count is then

$$0 = \varepsilon - e = c^2\Delta a + 2c(1-c)\Delta d + (1-c)^2\Delta b.$$

Combined with the definition of μ, this gives

$$\Delta b = \frac{c}{1-c}\left(\frac{c\Delta a}{1-c} - 2\mu\right), \qquad \Delta d = \mu - \frac{c\Delta a}{1-c}. \tag{27}$$

We then turn to the triangle count. Plugging $a = e + \Delta a$, $b = e + \Delta b$, and $d = e + \Delta d$ into equation (25) gives

$$\begin{aligned} \tau - e^3 &= 3e^2(c^2\Delta a + 2c(1-c)\Delta d + (1-c)^2\Delta b) \\ &\quad +3e\left(c(c\Delta a + (1-c)\Delta d)^2 + (1-c)(c\Delta d + (1-c)\Delta b)^2\right) \end{aligned}$$

$$+c^3\Delta a^3 + 3c^2(1-c)\Delta a\Delta d^2 + 3c^2(1-c)^2\Delta b\Delta d^2 + (1-c)^3\Delta b^3. \tag{28}$$

The first line is $3e^2(\varepsilon - e) = 0$. The second line works out to

$$3e\left[c((1-c)\mu)^2 + (1-c)(-c\mu)^2\right] = 3ec(1-c)\mu^2.$$

The terms in the last line are

$$\begin{aligned}
&(1-c)^3\left(\frac{c\Delta a}{1-c}\right)^3, \\
&3c(1-c)^2\left[\left(\frac{c\Delta a}{1-c}\right)^3 - 2\mu\left(\frac{c\Delta a}{1-c}\right)^2 + \mu^2\left(\frac{c\Delta a}{1-c}\right)\right], \\
&3c^2(1-c)\left[\left(\frac{c\Delta a}{1-c}\right)^3 - 4\mu\left(\frac{c\Delta a}{1-c}\right)^2 + 5\mu^2\left(\frac{c\Delta a}{1-c}\right) - 2\mu^3\right],
\end{aligned}$$

and

$$c^3\left[\left(\frac{c\Delta a}{1-c}\right)^3 - 6\mu\left(\frac{c\Delta a}{1-c}\right)^2 + 12\mu^2\left(\frac{c\Delta a}{1-c}\right) - 8\mu^3\right].$$

Setting the sum of the second line of (28) and these four terms equal to $-\delta^3$ gives

$$\begin{aligned}
-\delta^3 = &\left(\frac{c\Delta a}{1-c}\right)^3 - 6c\mu\left(\frac{c\Delta a}{1-c}\right)^2 \\
&+3c\mu^2\left((1+3c)\left(\frac{c\Delta a}{1-c}\right) + e(1-c)\right) - (6c^2+2c^3)\mu^3.
\end{aligned} \tag{29}$$

The definition of $\mathcal{G}_{e,\delta,\eta}$ implies that $\mu = O(\delta^{3/2})$. This means that the terms in (29) involving μ do not affect $\tau - e^3$ to leading order, so we have

$$\left(\frac{c\Delta a}{1-c}\right)^3 = -\delta^3 + o(\delta^3).$$

In particular, c is of order δ. This in turn implies that the $o(\delta^3)$ terms in (29), which go as $c^3\mu$ and $c\mu^2$ and higher powers of c and μ, are actually $O(\delta^{9/2})$, so

$$\begin{aligned}
\frac{c\Delta a}{1-c} &= -\delta + O(\delta^{5/2}), \\
c &= \frac{-\delta}{\Delta a} + O(\delta^2).
\end{aligned}$$

The leading corrections to the approximation $c\Delta a/(1-c) \approx -\delta$ come from the terms

$-6c\mu\,(c\Delta a/(1-c))^2$ and $3ec(1-c)\mu^2$ in the expansion of $\tau - e^3$. A priori we don't know which is larger, so for now we will keep both. (They will both turn out to be of order δ^5). However, all other terms are at least one power of δ smaller than one (or both) or these terms. We can use the approximation $c \approx -\delta/\Delta a$ to simplify these higher-order corrections:

$$\left(\frac{c\Delta a}{1-c}\right)^3 = -\delta^3 + \frac{3e\mu^2\delta - 6\mu\delta^3}{\Delta a} + O(\mu\delta^4, \mu^2\delta^2),$$
$$\frac{c\Delta a}{1-c} = -\delta + \frac{e\mu^2 - 2\mu\delta^2}{\delta\Delta a} + O(\mu\delta^2, \mu^2),$$

where $O(\mu\delta^4, \mu^2\delta^2)$ is shorthand for $O(\mu\delta^4) + O(\mu^2\delta^2)$. From that we compute c:

$$\begin{aligned} c &= \frac{-\delta + \frac{e\mu^2 - 2\mu\delta^2}{\delta\Delta a}}{\Delta a - \delta + \frac{e\mu^2 - 2\mu\delta^2}{\delta\Delta a}} + O(\mu\delta^2, \mu^2) \\ &= \frac{\delta + \frac{2\mu\delta^2 - e\mu^2}{\delta\Delta a}}{\delta - \Delta a} + O(\mu\delta^2, \mu^2). \end{aligned} \tag{30}$$

This then determines Δb and Δd:

$$\begin{aligned} \Delta b &= \frac{c}{1-c}\left(\frac{c\Delta a}{1-c} - 2\mu\right) \\ &= \frac{\delta^2}{\Delta a} + \frac{2\mu\delta}{\Delta a} + \frac{4\mu\delta^2 - 2e\mu^2}{\Delta a} + O(\mu\delta^3, \mu^2\delta), \\ \Delta d &= \mu + \delta + \frac{2\mu\delta^2 - \mu^2 e}{\delta\Delta a} + O(\mu\delta^2, \mu^2). \end{aligned}$$

Note that the constraints on ε and τ are algebraic, so we could have expressed b, c and d as power series in a and μ. However, in a power series the derivative of a term with respect to μ is at most an order μ^{-1} larger than the term itself, while the derivative with respect to a is at most of the same order as the term. We can therefore turn our estimates of (b, c, d) into estimates of $(\partial_a b, \partial_a c, \partial_a d)$ and $(\partial_\mu b, \partial_\mu c, \partial_\mu d)$. Specifically:

$$\begin{aligned} \partial_a b &= \frac{-\delta^2}{\Delta a^2} - \frac{2\mu\delta}{\Delta a^2} + \frac{4\mu^2 e - 8\mu\delta^2}{\Delta a^3} + O(\mu^2\delta, \mu\delta^3) \\ &= -\frac{c^2}{(1-c)^2} - \frac{2\mu\delta}{\Delta a^2} + \frac{2\mu^2 e - 8\mu\delta^2}{\Delta a^3} + O(\mu^2\delta, \mu\delta^3), \\ \partial_a c &= \frac{c}{\delta - \Delta a} + O(\mu\delta, \mu^2/\delta), \\ \partial_a d &= \frac{e\mu^2 - 2\mu\delta^2}{\delta\Delta a^2} + O(\mu\delta^2, \mu^2). \end{aligned}$$

$$\partial_\mu b = \frac{2\delta}{\Delta a} + \frac{4\delta^2 - 4e\mu}{\Delta a^2} + O(\mu\delta, \delta^3),$$
$$\partial_\mu c = \frac{2\mu e - 2\delta^2}{\delta \Delta a^2} + O(\mu, \delta^2),$$
$$\partial_a d = 1 + \frac{2\delta^2 - 2e\mu}{\delta \Delta a} + O(\delta^2, \mu).$$

5.2 Solving $\partial_a S = \partial_\mu S = 0$

We now solve the equations $\partial_a S = \partial_\mu S = 0$ in three passes. First we solve $\partial_a S = 0$ to lowest order, obtaining a to within $O(\delta)$. Using this value of a, we solve $\partial_\mu S = 0$, showing that μ is a specific constant times δ^2, up to an $O(\delta^3)$ error. Finally, we solve $\partial_a S = 0$ more precisely, determining a to order δ, with an $O(\delta^2)$ error. This then determines (a, b, c, d) to the accuracy specified in Proposition 18.

Since

$$\begin{aligned} S &= c^2 H(a) + 2c(1-c)H(d) + (1-c)^2 H(b), \\ \partial_a S &= c^2 H'(a) + 2c(1-c)H'(d)\partial_a d + (1-c)^2 H'(b)\partial_a b \\ &\quad + 2\partial_a c(H(d) - H(b) + c(H(a) + H(b) - 2H(d))). \end{aligned} \tag{31}$$

Keeping terms through $O(\delta^2)$, and noting that all discarded terms are of order δ^3 or higher, we have

$$\begin{aligned} \partial_a S &= c^2 H'(a) - c^2 H'(b) - \frac{2c}{\Delta a}\big(\delta H'(e) + c(H(a) - H(e)\big) + O(\delta^3) \\ &= c^2(H'(a) + H'(e)) - 2c^2 \frac{H(a) - H(e)}{a - e} + O(\delta^3). \end{aligned}$$

Setting this equal to zero and dividing by δ^2 gives

$$H'(a) + H'(e) - 2\frac{H(a) - H(e)}{a - e} = O(\delta).$$

This implies that a is within $O(\delta)$ of either e or $1 - e$. Since $a - e$ is assumed not to be $o(1)$, this means that we can write

$$a = 1 - e + a_1,$$

where a_1 is $O(\delta)$. This allows us to expand $H(a)$ and $H'(a)$ a power series in a_1 with quantified errors and also to quantify how much $-1/\Delta a$ and $1/(\delta - \Delta a)$ differ from $1/(2e - 1)$. In particular,

$$H(d) - H(b) + c(H(a) + H(b) - 2H(d)) = H'(e)\delta + O(\mu, \delta^2).$$

We now evaluate

$$\partial_\mu S = 2c(1-c)H'(d)\partial_\mu d + (1-c)^2 H'(b)\partial_\mu b \\ +2\partial_\mu c(H(d) - H(b) + c(H(a) + H(b) - 2H(d)))$$

through order μ, δ^2:

$$\begin{aligned} 2c(1-c)H'(d)\partial_\mu d &= 2\left(\frac{\delta}{2e-1} + \frac{a_1\delta - 2\delta^2}{(2e-1)^2}\right)\left(H'(e)+\delta H''(e)\right) \\ &\quad \left(1+\frac{2e\mu-2\delta^2}{\delta(2e-1)}\right) + o(\mu,\delta^2) \\ &= H'(e)\left(\frac{2\delta}{2e-1} + \frac{2a_1\delta - 8\delta^2 + 4e\mu}{(2e-1)^2}\right) + \frac{2H''(e)\delta^2}{2e-1} + o(\mu,\delta^2), \\ (1-c)^2 H'(b)\partial_\mu b &= H'(e)\left(1 - \frac{2\delta}{2e-1}\right)\left(\frac{-2\delta}{2e-1} + \frac{4\delta^2 - 4e\mu - 2a_1\delta}{(2e-1)^2}\right) + o(\mu,\delta^2) \\ &= H'(e)\left(\frac{-2\delta}{2e-1} + \frac{8\delta^2 - 4e\mu - 2a_1\delta}{(2e-1)^2}\right) + o(\mu,\delta^2). \\ 2\partial_\mu c(H(d) - H(b) + c(H(a) + H(b) - 2H(d))) &= 2\left(\frac{2e\mu - 2\delta^2}{\delta(2e-1)^2}\right) H'(e)\delta + o(\mu,\delta^2). \end{aligned}$$

Adding up the terms in $\partial_\mu S$ and setting the total equal to zero, we have

$$\begin{aligned} \frac{4H'(e)(\mu e - \delta^2)}{(2e-1)^2} + \frac{2H''(e)\delta^2}{2e-1} &= o(\mu,\delta^2), \\ e\mu - \delta^2 &= -\frac{H''(e)\delta^2(2e-1)}{2H'(e)} + o(\mu,\delta^2), \\ \mu &= \frac{\delta^2}{e}\left(1 - \frac{H''(e)(2e-1)}{2H'(e)}\right) + o(\delta^2,\mu). \quad (32) \end{aligned}$$

Now that we have established that $\mu = O(\delta^2)$, we can check the order of the error terms in our estimate of $\partial_\mu S$. They are all $O(\delta^3)$, not just $o(\mu,\delta^2)$, so we have actually estimated μ to within $O(\delta^3)$.

Using our known value of μ, we can restate our estimates for the derivatives of (b, c, d) as

$$\begin{aligned} \partial_a b &= \frac{-\delta^2}{\Delta a^2} - \frac{2\mu\delta}{\Delta a^2} + O(\delta^4) \\ &= \frac{-c^2}{(1-c)^2} - \frac{2c\mu}{2e-1} + O(\delta^4), \\ \partial_a c &= \frac{c}{\delta - \Delta a} + O(\delta^3) \\ &= \frac{c^2}{\delta} + O(\delta^3), \end{aligned}$$

$$\partial_a d = O(\delta^3).$$

We then compute the first three terms in $\partial_a S$:

$$\begin{aligned} c^2 H'(a) &= c^2(-H'(e) + a_1 H''(e)) + O(\delta^4), \\ (1-c)^2 \partial_a b H'(b) &= -c^2 H'(e) - \frac{2c\mu}{2e-1} H'(e) + O(\delta^4), \\ 2c(1-c)\partial_a d H'(d) &= O(\delta^4). \end{aligned}$$

The last term, namely $2\partial_a c(H(d) - H(b) + c(H(a) + H(b) - 2H(d)))$, works out to

$$\begin{aligned} &\frac{2c^2}{\delta}\left(\frac{H''(e)\delta^2}{2} + H'(e)\left(\delta + \mu + \frac{\delta^2}{2e-1} + c(-a_1 - 2\delta)\right)\right) + O(\delta^4) \\ &= \frac{2c^2}{\delta}\left(\frac{H''(e)\delta^2}{2} + H'(e)(\delta + \mu - c(a_1 + \delta)\right) + O(\delta^4). \end{aligned}$$

Adding everything together, we have

$$\partial_a S = c^2(a_1 + \delta)\left(H''(e) - \frac{2H'(e)}{2e-1}\right) + O(\delta^4),$$

so

$$a_1 = -\delta + O(\delta^2).$$

Applying this value, and the computed value of μ, to our expressions for (a, b, c, d), we obtain the estimates of Proposition 18.

5.3 Uniqueness

Having proven that any optimizing bipodal graphon must have certain parameter estimates, we now show that it is unique. We compute the Hessian of S with respect to a and μ in a neighborhood of the optimal graphon, specifically in the neighborhood defined by the estimates in Proposition 18, and show that is is well-approximated by a fixed non-degenerate matrix. That is, the equations $\partial_a S = \partial_\mu S = 0$ are approximately linear and non-degenerate in this region, and so have a unique solution.

The partial derivatives of b, c, and d are, to leading order,

$$\begin{aligned} \partial_a b &= \frac{-\delta^2}{(2e-1)^2} + O(\delta^3), \partial_\mu b = \frac{-2\delta}{2e-1} + O(\delta^2), \\ \partial^2_{aa} b &= \frac{-2\delta^2}{(2e-1)^3} + O(\delta^3), \partial^2_{\mu\mu} b = \frac{-4e}{(2e-1)^2} + O(\delta), \partial^2_{a\mu} b = \frac{-2\delta}{(2e-1)^2} + O(\delta^2). \\ \partial_a c &= \frac{\delta}{(2e-1)^2} + O(\delta^2), \partial_\mu c = \frac{2e\mu - 2\delta^2}{\delta(2e-1)^2} + O(\delta^2), \end{aligned}$$

$$\partial^2_{aa}c = \frac{2\delta}{(2e-1)^3} + O(\delta^2),\ \partial^2_{\mu\mu}c = \frac{2e}{\delta(2e-1)} + O(1),\ \partial^2_{a\mu}c = \frac{4e\mu - 4\delta^2}{\delta(2e-1)} + O(\delta^2).$$
$$\partial_a d = \frac{e\mu^2 - 2\mu\delta^2}{\delta(2e-1)^2} + O(\delta^4),\ \partial_\mu d = 1 + O(\delta),$$
$$\partial^2_{aa}d = \frac{2e\mu^2 - 4\mu\delta^2}{\delta(2e-1)^3} + O(\delta^4),\ \partial^2_{\mu\mu}d = \frac{2e}{\delta(2e-1)} + O(1)\partial^2_{a\mu}d = \frac{2e\mu - 2\delta^2}{\delta(2e-1)^2} + O(\delta^2).$$

We compute $\partial^2_{aa}S$ to order δ^2. Four terms contribute to that order, namely

$$2\partial^2_{aa}c(H(d) - H(b)) \approx \frac{4\delta^2 H'(e)}{(2e-1)^3}, \quad 4c\partial_a c H'(a) \approx \frac{-4\delta^2 H'(e)}{(2e-1)^3},$$
$$\partial^2_{aa}bH'(b) \approx \frac{-2\delta^2 H'(e)}{(2e-1)^3}, \quad c^2 H''(a) \approx \frac{\delta^2 H''(e)}{(2e-1)^2}.$$

Adding up these terms gives

$$\partial^2_{aa}S = \frac{\delta^2}{(2e-1)^2}\left(H''(e) - \frac{2H'(e)}{2e-1}\right) + O(\delta^3).$$

In the expansion of $\partial^2_{\mu\mu}S$, the unique $O(\delta^{-1})$ term is $2(1-2c)\partial^2_{\mu\mu}dH'(d)$, giving us

$$\partial^2_{\mu\mu}S = \frac{4eH'(e)}{\delta(2e-1)} + O(1).$$

Finally, in computing $\partial^2_{a\mu}S$ there are two $O(\delta)$ terms, namely $(1-c)^2\partial^2_{a\mu}bH'(b)$ and $2(1-2c)\partial_a c\partial_\mu dH'(d)$. The first gives $-2\delta H'(e)/(2e-1)^2 + O(\delta^2)$ while the second gives $+2\delta H'(e)/(2e-1)^2 + O(\delta^2)$, so

$$\partial^2_{a\mu}S = O(\delta^2).$$

Since the Hessian of S takes the form

$$\begin{pmatrix} K_1\delta^2 + O(\delta^3) & O(\delta^2) \\ O(\delta^2) & K_2\delta^{-1} + O(1) \end{pmatrix},$$

for negative constants K_1 and K_2, the entropy has a unique maximizer, and indeed a unique critical point, in the region governed by the estimates of Proposition 18. At first glance, the off-diagonal terms appear as big as the $\partial^2_{aa}S$ term. However, replacing the variable μ with $\mu\delta^{-3/2}$ would convert the Hessian to the form

$$\begin{pmatrix} K_1\delta^2 + O(\delta^3) & O(\delta^{7/2}) \\ O(\delta^{7/2}) & K_2\delta^2 + O(\delta^3) \end{pmatrix},$$

in which the known diagonal terms more manifestly dominate the error terms.

This completes the proof of Proposition 18.

6 Averaging

We now return to the study of entropy-optimal graphons that are not necessarily bipodal: let g be a graphon maximizing $S(g)$ subject to $\varepsilon(g) = e$, $\tau(g) = e^3 - \delta^3$, and recall from Proposition 6 that we can partition [0, 1] into podes C_1 and C_2 so that g is approximately bipodal with respect to these podes. Recall also the definition of v from Proposition 6, and that $g(x, y) \approx e - v(x)v(y)$. Let h be the graphon obtained by averaging g on the podes $C_i \times C_j$. We will write $h = g + \Delta h$.

In this section we show that g must be *exactly* bipodal. Specifically, we show that if $g \neq h$ then we can get a contradiction by constructing a $\tilde{h}$ with $\tau(\tilde{h}) = \tau(g)$ but $S(\tilde{h}) > S(g)$. This argument comes in four parts:

(1) We give a lower bound on $S(h) - S(g)$, and we show that if the bound is almost sharp then Δh has a particular structure;
(2) We give an upper bound on $\tau(h) - \tau(g)$, and we show that if the bound is almost sharp then Δh has a particular structure;
(3) We find a perturbation $\tilde{h}$ of h that trades off entropy for triangles at essentially the best possible ratio of the preceding two bounds, and it follows that if g is optimal then both of the preceding two bounds must be almost sharp; and finally,
(4) We show that the structure of Δh implied by the first two parts is incompatible with Proposition 6.

One basic observation before we start is that because $v(x)v(y)$ is constant on $\Omega_i \times \Omega_j$, and because the mean of any function is the constant with the smallest L^2 distance to that function,

$$\|\Delta h\|_2^2 \leq \|g - v(x)v(y)\|_2^2 = O(\delta^3), \tag{33}$$

where the last inequality follows from Proposition 6.

6.1 The entropy change

Let's compute the change in entropy that results from replacing each pode in g by its average in h. Recall that $C_2 \subset [0, 1]$ is the set on which $v(x) \approx 0$ and $C_1 = [0, 1] \setminus C_2$ is the set on which $v(x) \approx \sqrt{2e - 1}$. By Proposition 6 and the fact that $|C_1| = \Theta(\delta)$,

$$h(x, y) = \begin{cases} h_{11} = 1 - e + O(\delta^{1/2}) & \text{on } C_1 \times C_1 \\ h_{12} = e + O(\delta) & \text{on } C_2 \times C_1 \text{ and } C_1 \times C_2 \\ h_{22} = e + O(\delta^{3/2}) & \text{on } C_2 \times C_2 \end{cases}$$

In terms of the notation of Sect. 5, $h_{11} = a$, $h_{12} = d$, $h_{22} = b$, and $|C_1| = c$.

Because h is obtained by averaging g, it has larger entropy. Our first bound shows that it must be larger by at least about the "optimal entropy-L^2 tradeoff constant," $C(e)$.

Lemma 19

$$S(g) \le S(h) - C(e)(1 - O(\sqrt{\delta}))\|g - h\|_2^2.$$

Proof. Let $D(p,q) = p \ln \frac{p}{q} + (1-p)\ln\frac{1-p}{1-q}$, and note that for any set $A \subset [0,1]^2$,

$$\int_A D(g(x,y),q)\,dx\,dy = H(q)|A| - \int_A H(g(x,y))\,dx\,dy.$$

In particular, if we apply this to $A = \Omega_i \times \Omega_j$ then

$$\int_{\Omega_i \times \Omega_j} H(h(x,y)) - H(g(x,y))\,dx\,dy = \int_A D(g(x,y),h_{ij})\,dx\,dy.$$

Recalling the definition of $C(e)$ from (15), we have

$$\int_A D(g(x,y),h_{ij})\,dx\,dy \ge C(h_{ij}) \int_A (g(x,y) - h_{ij})^2\,dx\,dy. \tag{34}$$

Since $C(e) = C(1-e)$ and C is continuous and differentiable at e, $C(h_{ij}) = C(e) + O(\delta^{1/2})$ for every i and j. Hence,

$$\int_A D(g(x,y),h_{ij})\,dx\,dy \ge C(e)(1 - O(\sqrt{\delta})) \int_A (g(x,y) - h(x,y))^2\,dx\,dy,$$

and summing over i and j gives

$$S(g) \le S(h) - C(e)(1 - O(\sqrt{\delta}))\|g - h\|_2^2. \quad \square$$

Next, we show that unless g approximately takes the value e and $1-e$, the entropy gain of h is even larger than in Lemma 19. The argument here is similar to that of Lemma 8; the difference is that the cost here is measured in terms of $\|\Delta h\|_2^2$, instead of in terms of $\|\Delta g\|_2^2$ (which is much larger).

Lemma 20 *Fix* $\eta > 0$. *For sufficiently small* δ *(depending on* η*), if*

$$\int_{C_1 \times C_1} V_{2e-1}(-\Delta h)\,dx\,dy \ge \eta \|\Delta h\|_2^2$$

or

$$\int_{(C_1 \times C_1)^c} V_{2e-1}(\Delta h)\,dx\,dy \ge \eta \|\Delta h\|_2^2$$

then

$$S(g) \le S(h) - C(e)(1 + \Omega(\eta))\|\Delta h\|_2^2.$$

Proof We follow the same argument as in Lemma 19, but in (34) we use the improved bound of Lemma 7 to obtain

$$\begin{aligned}&\int_{\Omega_i \times \Omega_j} D(g(x,y), h_{ij})\,dx\,dy\\ &\geq \int_{\Omega_i \times \Omega_j} (C(h_{ij}) + \Omega((g-1+h_{ij})^2)(g-h_{ij})^2\,dx\,dy.\end{aligned}$$

In particular, compared to (34), the right hand side is increased by

$$\begin{aligned}&\Omega\left(\int_{\Omega_i \times \Omega_j} \min\{|g-1+h_{ij}|, |g-h_{ij}|\}^2\,dx\,dy\right)\\ &\quad = \Omega\left(\int_{\Omega_i \times \Omega_j} V_{1-2h_{ij}}(\Delta h)\,dx\,dy\right).\end{aligned}$$

Recalling that $h_{22} = 1 - e + o(1)$, and $h_{ij} = e + o(1)$ otherwise, Lemma 10 implies that

$$\begin{aligned}\int_{C_1 \times C_1} D(g(x,y), h_{22})\,dx\,dy &\geq C(h_{22})|C_1|^2\\ &\quad + \int_{C_1 \times C_1} V_{2e-1}(-\Delta h)\,dx\,dy - o(\|\Delta h\|_2^2),\end{aligned}$$

and

$$\begin{aligned}\int_{\Omega_i \times \Omega_j} D(g(x,y), h_{ij})\,dx\,dy &\geq C(h_{ij})|\Omega_i||\Omega_j|\\ &\quad + \int_{\Omega_i \times \Omega_j} V_{2e-1}(\Delta h)\,dx\,dy - o(\|\Delta h\|_2^2),\end{aligned}$$

for $(i, j) \neq (2, 2)$. Now we simply sum over i, j as in the proof of Lemma 19: under our assumptions, at least one of the $\int_{\Omega_i \times \Omega_j} V_{2e-1}(\pm\Delta h)\,dx\,dy$ terms gives an additional contribution of $\Omega(\eta\|\Delta h\|_2^2)$ compared to Lemma 19. □

6.2 The triangle change

In this section we will control the triangle change between g and h. Recall that $T_{\Delta h} : L^2([0,1]) \to L^2([0,1])$ denotes the integral operator with kernel Δh. Recall that

$$\begin{aligned}\tau(h) - \tau(g) = 3\int h(x,y)\Delta h(y,z)\Delta h(z,x) + 3\int h(x,y)h(y,z)\Delta h(z,x)\\ + \int \Delta h(x,y)\Delta h(y,z)\Delta h(z,x)\end{aligned}$$

$$= 3\int h(x,y)\Delta h(y,z)\Delta h(z,x) + O(\|\Delta h\|_2^3)$$
$$\geq -3\|T_{\Delta h}v\|_2^2 + O(\delta^{3/2}\|\Delta h\|_2^2), \tag{35}$$

where the last line follows because (by Proposition 6) $h = e - v(x)v(y) - \tilde{r}(x,y)$ for some $\|\tilde{r}\|_2 \leq \delta^{3/2}$, and the term $\int e\Delta h(y,z)\Delta h(z,x)\, dx\, dy\, dz$ is non-negative.

Lemma 21

$$\langle v, T_{\Delta h}v\rangle = 0$$

Proof We can write

$$\langle v, T_{\Delta h}v\rangle = \int \Delta h(x,y)v(x)v(y)\, dx\, dy;$$

recall that $v(x)v(y)$ is constant on every pode and $\Delta h(x,y)$ integrates to zero on each pode. □

Now define $u = \|v\|_2^{-2}T_{\Delta h}v$, define $w(x,y)$ by $\Delta h(x,y) = u(x)v(y)+u(y)v(x)+w(x,y)$, and write T_w for the integral operator with kernel w. By Lemma 21, $\langle u, v\rangle = 0$. Then $u = T_{\Delta h}v = u + T_w v$, and so $T_w v = 0$. It follows that

$$\|T_{\Delta h}\|_2^2 = \|\Delta h\|_2^2 = 2\|u\|_2^2\|v\|_2^2 + \|w\|_2^2 = 2\frac{\|T_{\Delta h}v\|_2^2}{\|v\|_2^2} + \|w\|_2^2.$$

Since $\|w\|_2^2 \geq 0$, $\|T_{\Delta h}v\|_2^2 \leq \frac{1}{2}\|\Delta h\|_2^2\|v\|_2^2$, and then (35) gives the following bound:

Lemma 22

$$\tau(h) \leq \tau(g) + \frac{3}{2}\|\Delta h\|_2^2\|v\|_2^2 - 3\|w\|_2^2\|v\|_2^2 + O(\delta^{3/2}\|\Delta h\|_2^2).$$

We should interpret Lemma 22 as saying that $\tau(h) \lesssim \tau(g) + \frac{3}{2}\|\Delta h\|_2^2\|v\|_2^2$, with approximate tightness only if $\|w\|_2^2$ is small compared to $\|\Delta h\|_2^2$.

Let's also note that $\|u\|_2$ must be small:

Lemma 23

$$\|u\|_2^2 = O(\delta^2).$$

Proof

$$2\|u\|_2^2\|v\|_2^2 \leq \|\Delta h\|_2^2 = O(\delta^3),$$

where the second inequality follows from (33). Finally, $\|v\|_2^2 = \Theta(\delta)$. □

In order for Lemma 22 to be sharp, $\|w\|_2$ must be small compared to $\|\Delta h\|_2$. On the other hand, Lemma 20 implies that if the entropy change inequality of Lemma 19 is sharp then Δh must spend most of its L^2 mass at particular values. Combining these two pieces of tells us that u must spend most of its L^2 mass at particular values.

Lemma 24 *Fix $\eta \geq \delta > 0$ and suppose that $\|w\|_2^2 \leq \eta \|\Delta h\|_2^2$.*
If $\int_{C_1 \times C_1} V_{2e-1}(-\Delta h)\,dx\,dy \leq \eta \|\Delta h\|_2^2$ then

$$\int_{C_1} V_{\sqrt{2e-1}}(-u) \leq O(\eta)\|u\|_2^2.$$

If $\int_{(C_1 \times C_1)^c} V_{2e-1}(\Delta h)\,dx\,dy \leq \eta \|\Delta h\|_2^2$ then

$$\int_{C_2} V_{\sqrt{2e-1}}(u) \leq O(\eta)\|u\|_2^2.$$

Proof Let $v_2 = \sqrt{2e-1} \pm O(\delta)$ be the value that v takes on C_1. On $C_1 \times C_1$, we have $\Delta h = v_2(u(x) + u(y)) + w(x, y)$. If $\int_{C_1 \times C_1} V_{2e-1}(-\Delta h)\,dx\,dy \leq \eta \|\Delta h\|_2^2$ then (by Lemma 9)

$$\int_{C_1 \times C_1} V_{2e-1}(-v_2(u(x) + u(y)))\,dx\,dy \leq O(\eta)\|\Delta h\|_2^2 + O\left(\|w\|_2^2\right) = O(\eta)\|\Delta h\|_2^2.$$

Since $(2e-1)/v_2 = \sqrt{2e-1} + O(\delta)$, Lemma 10 implies that

$$\int_{C_1 \times C_1} V_{\sqrt{2e-1}}(-u(x) - u(y))\,dx\,dy \leq O(\eta)\|\Delta h\|_2^2 + O\left(\delta^2 \|u\|_2^2\right) \leq O(\eta)\|\Delta h\|_2^2.$$

We apply Lemma 11 with $K = O(\eta)\|\Delta h\|_2^2$, $\Omega = C_1$, and μ the Lebesgue measure on Ω. (Note that the hypotheses of Lemma 11 are satisfied when δ is small, because $|C_1| = \Theta(\delta)$, while $\|u\|_2^2 = O(\delta^2)$ and $\|\Delta h\|_2^2 = O(\delta^3)$). Since $|C_1| = \Theta(\delta)$, this gives

$$\int_{C_1} V_{\sqrt{2e-1}}(-u(x))\,dx = O\left(\eta \|\Delta h\|_2^2/\delta\right) = O\left(\eta \|u\|_2^2\right),$$

as claimed.

Now we prove the second claim. For $(x, y) \in C_2 \times C_1$, $\Delta h(x, y) = v_2 u(x) + w(x, y) \pm O(\delta u(y))$. Since $\|\delta u(y)\|_2^2 = \delta^2 \|u\|_2^2 = \Theta(\delta \|\Delta h\|_2^2)$, if we set $\tilde{w}(x, y) = w(x, y) \pm O(\delta u(y))$ then $\Delta h(x, y) = v_2 u(x) + \tilde{w}(x, y)$, and $\|\tilde{w}\|_2^2 \leq (\eta + O(\delta))\|\Delta h\|_2^2 = O(\eta)\|\Delta h\|_2^2$. Assuming that

$$\int_{C_2 \times C_1} V_{2e-1}(\Delta h)\,dx\,dy \leq \eta \|\Delta h\|_2^2,$$

Lemma 9 implies that

$$\int_{C_2\times C_1} V_{2e-1}(v_2 u(x))\,dx\,dy \le O(\eta)\|\Delta h\|_2^2,$$

and since $|C_1| = \Theta(\delta)$, we have

$$\int_{C_2} V_{(2e-1)/v_2}(u(x))\,dx \le O(\delta\eta)\|\Delta h\|_2^2 = O(\eta\|u\|_2^2).$$

The second claim follows from Lemma 10 and the fact that $v_2 = \sqrt{2e-1} + O(\delta)$. □

7 The improvement lemma

The goal of this section is to show that if we have a bipodal graphon with approximately the expected parameter values, then we can trade off triangle density for entropy at a prescribed rate. The idea behind the perturbation is simple: we lower the triangle density by increasing the size of the small pode, and tweaking the other parameters to keep the edge density constant.

Recall that (up to measure-preserving transformations), a bipodal graphon h may be parametrized as

$$h(x, y) = \begin{cases} a & x, y < c, \\ b & x, y > c, \\ d & x < c < y \text{ or } y < c < x, \end{cases}$$

and we may assume that $c \le 1 - c$.

Recall that $\mathcal{G}_{e,\delta,\eta}$ is the set of bipodal graphons with edge-density e, triangle density $e^3 - \delta^3$, and parameters a, b, c and d satisfying $|b - e| < \eta\sqrt{\delta}$, $c < \eta$, $|d - e| < \eta$, $|a - e| > \eta$, and $|\mu| = O(\delta^{3/2})$. The point of this definition is that our earlier estimates imply that for any e, some $\eta > 0$ depending on e, and all sufficiently small $\delta > 0$ depending on e, the bipodal graphon h obtained by starting with an optimal graphon g and averaging on podes belongs to $\mathcal{G}_{e,\delta,\eta}$.

Proposition 25 *For any graphon $h \in \mathcal{G}_{e,\delta,\eta}$ and for any $\tau(h) > t \ge \tau(h) - \delta^3$, there is a bipodal graphon $\tilde{h}$ with edge density e, triangle density t, and entropy*

$$S(\tilde{h}) \ge S(h) - C(e)\frac{2(\tau(h) - t)}{3\delta}(1 + O(\eta)).$$

Out of the four parameters a, b, c, and d, the edge-density constraint $\varepsilon(h) = e$ can be used to eliminate one. Defining

$$\Delta a = a - e, \qquad \Delta d = d - e, \qquad \mu = \frac{c\Delta a}{1-c} + \Delta d,$$

we can change parameters to express everything in terms of Δa, c, and μ; and in (29) we showed that the triangle deficit can be expressed as

$$-\delta^3 = c^3\alpha^3 - 6c^3\mu\alpha^2 + 3c\mu^2((1+3c)\alpha + e(1-c)) - (6c^2+2c^3)\mu^3, \quad (36)$$

where $\alpha = \Delta a/(1-c)$. To prove Proposition 25 we will simply increase c while keeping Δa and μ constant. In terms of the original parameters, this is equivalent to setting

$$\begin{aligned} c(s) &= c + s \\ a(s) &= a \\ d(s) &= e + \mu - \frac{c(s)\Delta a}{1-c(s)} \\ b(s) &= e - \frac{2c(s)(1-c(s))d(s) + c^2(s)a(s)}{(1-c(s))^2}. \end{aligned}$$

In particular, everything is a rational function of s and is smooth for $s < 1-c$. Let h_s be the bipodal graphon with these parameters.

If $\mu = O(\delta^{3/2})$, we see immediately from (36) that

$$\frac{d}{ds}\tau(h_s) = 3c^2(s)\alpha^3 + O(\delta^3) = 3\delta^2(2e-1) + O(\delta^3 + \delta^2 s).$$

Moreover, the $O(\delta^3)$ term is uniform over $0 \le s \le \delta/(2e-1)$, because μ is constant in s and $|c(s)| \le (1+(2e-1)^{-1})\delta$ for s in this range. Since $-\delta^3$ is continuous in s, for any $t \in [\tau(h) - \eta\delta^3, \tau(h))$ there is an

$$s_* = \frac{\tau(h) - t}{3\delta^2(2e-1)}(1 + O(\eta))$$

such that $\tau(h_{s_*}) = t$.

Next, we consider the change in entropy as a function of s. Recall that if $\int h = e$ then $\int H(h) = H(e) - \int D(h)$. Therefore, the change in entropy is the same as the change in $-\int D(h)$; that is,

$$\frac{d}{ds}S(h_s) = -\frac{d}{ds}\int D(h_s(x,y))\,dx\,dy.$$

We can write

$$\int D(h_s(x,y))\,dx\,dy = c^2(s)D(a) + 2c(s)(1-c(s))D(d(s)) + (1-c(s))^2 D(b(s)).$$

Now, D is differentiable with $D(e) = 0$, and D is locally quadratic around e. That is, $D'(e+\epsilon) = O(\epsilon)$, $D(e+\epsilon) = O(\epsilon^2)$, and $D(1-e+\epsilon) = D(1-e) + O(\epsilon)$. In particular, because $|b(s) - e| = O(\eta\sqrt{\delta})$ and $|d(s) - e| = O(\eta)$, the derivative in s is bounded by the contribution of the first term:

$$\frac{d}{ds}\int D(h_s(x,y))\,dx\,dy = 2c(s)c'(s)D(a) + O(\delta\eta) = 2\frac{\delta}{2e-1}D(1-e) + O(\delta\eta),$$

where the second equality holds for $s \leq O(\eta\delta)$. Plugging in $s = s_*$ and applying the Fundamental Theorem of Calculus, we obtain

$$D(h_{s_*}) = D(h) - \frac{2\delta D(1-e)}{2e-1}(1+O(\eta))s_* = D(h) - C(e)\frac{2(\tau(h)-t)}{3\delta}(1+O(\eta)),$$

completing the proof of Proposition 25.

8 Completing the proof of bipodality

Recall that h was obtained by averaging g on podes. Define $\epsilon = \|g-h\|_2^2$, and recall that $\epsilon = O(\delta^3)$. We assume for a contradiction that g is optimal but not bipodal, and hence $\epsilon > 0$.

First, note that Proposition 6 and Lemma 13 imply that for any $\eta_1 > 0$, if $\delta > 0$ is sufficiently small then $g \in \mathcal{G}_{e,\delta,\eta_1}$. Indeed, part (3) of Proposition 6 implies the required estimates on the parameters a, b, c, and d, while Lemma 13 implies the required estimate on μ. In particular, we may apply Proposition 25 to construct a graphon $\tilde{h}$ with $\tau(\tilde{h}) = \tau(g)$. Lemma 22 implies that $\tau(h) \leq \tau(g) + (3/2)\epsilon\delta + o(\epsilon\delta)$, and so Proposition 25 with $t = \tau(g)$ gives a graphon $\tilde{h}$ with $\tau(\tilde{h}) = \tau(g)$ and

$$S(g) \geq S(\tilde{h}) = S(h) - \frac{2}{3}C(e)\frac{\tau(h)-\tau(g)}{\delta}(1-O(\eta_1)) \geq S(h) - C(e)\epsilon(1+O(\eta_1)).$$

In particular, for any fixed $\eta > 0$ we can find $\eta_1 > 0$ small enough so that the conclusion of Lemma 20 fails to hold for all sufficiently small δ.

On the other hand, for any fixed $\eta > 0$, if $\|w\|_2^2 \geq \eta\|\Delta h\|_2^2$ then Lemma 22 gives the improved bound $\tau(h) \leq \tau(g) + (3/2)\epsilon\delta(1-\Omega(\eta))$, meaning that (by Proposition 25, if η_1 is small compared to η)

$$S(g) \geq S(h) - C(e)\epsilon(1-\Omega(\eta)),$$

which contradicts Lemma 19 (for δ sufficiently small). We conclude that

$$\|w\|_2^2 \leq \eta\|\Delta h\|_2^2, \tag{37}$$

and that the conclusion of Lemma 20 fails, meaning that

$$\max\left\{\int_{C_1\times C_1} V_{2e-1}(-\Delta h)\,dx\,dy, \int_{(C_1\times C_1)^c} V_{2e-1}(\Delta h)\,dx\,dy\right\} \leq \eta\epsilon. \tag{38}$$

Finally, we will show that (37) and (38) together contradict Proposition 6. Fix $x \in C_1$. Recall that $v : [0, 1] \to \mathbb{R}$ takes two values; let $v_1 \approx \sqrt{2e-1}$ be the value v takes on C_1 and let $v_2 = \Theta(\delta)$ be the value v takes on C_2. Then $\Delta h_x = v_1 u + u(x) v + w_x$ and $g_x = h_x - \Delta h_x$, where $h_x(y) = (1-e) + o(1)$ for $y \in C_1$. Note that h_x and v are constant on C_1, with $h_x = 1 - e + o(1)$ and $v = \sqrt{2e-1} + o(1)$. Recalling that $u = \|v\|_2^{-2} T_{\Delta h} v$, note that if $u(x) \approx -\sqrt{2e-1}$ then $h_x - u(x)v \approx e$ on C_1 (and is constant on C_1). To be precise, we consider two cases.

If $u(x) \le -(1/2)(\sqrt{2e-1}+\eta)$ then $h_x - u(x)v \ge 1/2 + \Omega(\eta)$ on C_1. In particular, $h_x - u(x)v$ is closer to e than it is to $1-e$, and so item 6 of Proposition 6 implies that w_x will need to be large to compensate: we have $g_x = h_x - u(x)v - v_1 u - w_x \ge 1/2 + \Omega(\eta) - v_1 u - w_x$ and so by the triangle inequality,

$$\left(\int_{C_1} (g_x(y) - e)^2 \, dy\right)^{1/2} \le |e - \frac{1}{2} - \Omega(\eta)| + \|v_1 u + w_x\|_2$$
$$\le |e - \frac{1}{2} - \Omega(\eta)| + \|w_x\|_2 + O(\delta). \tag{39}$$

On the other hand,

$$\left(\int_{C_1} (g_x(y) - (1-e))^2 \, dy\right)^{1/2} \ge |1 - e - \frac{1}{2} - \Omega(\eta)| - \|v_1 u + w_x\|_2$$
$$\ge |1 - e - \frac{1}{2} - \Omega(\eta)| - \|w_x\|_2 - O(\delta). \tag{40}$$

Item 6 of Proposition 6 implies that (39) can be at most $O(\delta^2)$ smaller than (40). Since $e > 1/2$, for $\eta > 0$ sufficiently small (and $\delta > 0$ sufficiently small depending on η),

$$\|w_x\|_2 \ge \Omega(\eta) - O(\delta) = \Omega(\eta) = \Omega(\eta |u(x)|).$$

For the other case, if $u(x) \ge -(1/2)(\sqrt{2e-1} + \eta)$ then (for η sufficiently small) u is bounded away from $-\sqrt{2e-1}$ and so $V_{\sqrt{2e-1}}(-u(x)) \ge \Omega(u^2(x))$.

Let $A = \{x \in C_1 : u(x) \le -(1/2)(\sqrt{2e-1} + \eta)\}$ (i.e. the set of x for which the first case happens), and let $B = C_1 \setminus A$. Then $\|w_x\|_2^2 \ge \Omega(\eta^2 u^2(x))$ for $x \in A$ and $u^2(x) \le O(V_{\sqrt{2e-1}}(-u(x)))$ for $x \in B$. If $\int_A u^2(x)\, dx \ge (1/2) \int_{C_1} u^2(x)\, dx$ then

$$\int_{C_1} \|w_x\|_2^2 \, dx \ge \Omega(\eta^2) \int_{C_1} u^2(x) \, dx; \tag{41}$$

and if $\int_B u^2(x)\, dx \ge (1/2) \int_{C_1} u^2(x)\, dx$ then

$$\int_{C_1} V_{\sqrt{2e-1}}(-u(x)) \, dx \ge \Omega(1) \int_{C_1} u^2(x) \, dx. \tag{42}$$

We will give a similar argument for C_2: fix $x \in C_2$ and note that $g_x = h_x - u(x)v - v_2 u - w_x$, where $h_x \approx e$ on C_1 and $v \approx \sqrt{2e-1}$ on C_1, and $v_2 = O(\delta)$. This time, if

$u(x) \geq (1/2)(\sqrt{2e-1}+\eta)$, item 5 of Proposition 6 will imply that $\|w_x\|_2$ is large, while in the other case we will have $V_{\sqrt{2e-1}}(u(x))$ bounded below by $u^2(x)$.

Let $A = \{x \in C_2 : u(x) \geq (1/2)(\sqrt{2e-1}+\eta)\}$ and let $B = C_2 \setminus A$. Analogously to (39) and (40), if $x \in A$ then $h_x - u(x)v$ takes a constant value of at most $1/2 - \Omega(\eta)$, which is $\Omega(\eta)$ closer to $1-e$ than it is to e. Then $g_x \leq 1/2 - \Omega(\eta) - v_2 u - w_x$, and so

$$\left(\int_{C_1} (g_x(y) - (1-e))^2\, dy\right)^{1/2} \leq |1 - e - \frac{1}{2} + \Omega(\eta)| + \|w_x\|_2 + O(\delta)$$

and

$$\left(\int_{C_1} (g_x(y) - e)^2\, dy\right)^{1/2} \leq |e - \frac{1}{2} + \Omega(\eta)| - \|w_x\|_2 - O(\delta),$$

and it follows from item 5 of Proposition 6 that

$$\|w_x\|_2 \geq \Omega(\eta |u(x)|).$$

On the other hand, if $x \in B$ then $V_{\sqrt{2e-1}}(u(x)) \geq \Omega(u^2(x))$ (for η sufficiently small). As before, depending on whether $\int_A u^2$ or $\int_B u^2$ is larger, we have either

$$\int_{C_2} \|w_x\|_2^2\, dx \geq \Omega(\eta^2) \int_{C_2} u^2(x)\, dx; \tag{43}$$

or

$$\int_{C_2} V_{\sqrt{2e-1}}(u(x))\, dx \geq \Omega(1) \int_{C_2} u^2(x)\, dx. \tag{44}$$

Finally, we consider whether $\int_{C_2} u^2\, dx$ or $\int_{C_1} u^2\, dx$ is larger. In the former case, we apply (43) or (44); in the latter case, we apply (41) or (42). We conclude that either $\|w\|_2^2 \geq \Omega(\eta^2)\|u\|_2^2$, in which case $\|\Delta h\|_2^2 = \|u\|_2^2\|v\|_2^2 + \|w\|_2^2$ and $\|v\|_2^2 = \Theta(\delta)$ imply that

$$\|w\|_2^2 \geq \Omega(\epsilon),$$

or else at least one of $\int_{C_2} V_{\sqrt{2e-1}}(u(x))\, dx$ or $\int_{C_1} V_{\sqrt{2e-1}}(u(x))\, dx$ is $\Omega(\|u\|_2^2)$, which by Lemma 24 implies that

$$\max\left\{\int_{C_1\times C_1} V(\Delta h)\, dx\, dy, \int_{(C_1\times C_1)^c} V(\Delta h)\, dx\, dy\right\} \geq \Omega(\epsilon).$$

The first case contradicts (37); the second contradicts (38). In either case, this contradiction implies that $\epsilon = 0$ and so g is bipodal. This concludes the proof of Theorem 1.

9 The region $\mathcal{O}_2$

Having established the properties of the region $\mathcal{O}_1$, where τ is slightly less than ε^3, we turn our attention to the region where the triangle density τ is slightly greater than ε^3, the region $\mathcal{O}_2$. In this region, the optimizing graphon was already proven to be bipodal in [15]. To complete the proof of Theorem 2, we must determine the parameters of that optimal bipodal graphon.

Remark 26 The results of [15] apply when $0 < \varepsilon < 1/2$ as well as when $1/2 < \varepsilon < 1$, as do all the computations in this section. However, they do not apply *at* $\varepsilon = 1/2$, insofar as many quantities go as negative powers of $1 - 2\varepsilon$. By contrast, our results for $\tau < \varepsilon^3$ only apply when $\varepsilon > 1/2$. The situation when $\varepsilon < 1/2$ and $\tau < \varepsilon^3$ is qualitatively different and is the subject of ongoing work.

Our analysis of $\mathcal{O}_2$ is qualitatively similar to that of $\mathcal{O}_1$, with one very important difference. Instead of everything being a power series in δ, where $(\varepsilon, \tau) = (e, e^3 - \delta^3)$, everything is a power series in $\Delta\tau$, where $(\varepsilon, \tau) = (e, e^3 + \Delta\tau)$. We use the constraint $\varepsilon = e$ to express b in terms of a, c, and d, and then use the value of τ to express c in terms of a, d and $\Delta\tau$. We then solve the equations $\partial_d S = 0$ and $\partial_a S = 0$ iteratively. Knowing a and d to a given order allows us to compute c to a certain order, which then allows us to compute a and d to one more power of $\Delta\tau$ than before, which allows us to compute c to one more power of $\Delta\tau$, and the process repeats. We will exhibit the first few steps in this process, enough to compute the entropy up to an $O(\Delta\tau^3)$ error.

Extended to all orders, the result would be a set of asymptotic series for (a, b, c, d), and thus for the entropy S. However, the analyticity of of parameters only implies that we have convergent Taylor series expansions around points in the interior of $\mathcal{O}_2$. Our expansion in powers of $\Delta\tau$ is around a point $(\varepsilon, \tau) = (e, e^3)$ on the boundary of $\mathcal{O}_2$, so convergence of our series is not guaranteed.

We therefore need a separate iterative method to use at fixed (non-infinitesimal) values of $\Delta\tau$. As before, we linearize the equations $\partial_a S = \partial_d S = 0$. As long as the Hessian of S is well-approximated by a fixed matrix M_0 in a neighborhood of the approximate values of (a, b, c, d) that we have derived for $(\varepsilon, \tau) = (e, e^3 + \Delta\tau)$, then the iteration

$$\begin{pmatrix} a_{new} \\ d_{new} \end{pmatrix} = \begin{pmatrix} a_{old} \\ d_{old} \end{pmatrix} - M_0^{-1} \begin{pmatrix} \partial_a S(a_{old}, d_{old}) \\ \partial_d S(a_{old}, d_{old}) \end{pmatrix} \tag{45}$$

is guaranteed to converge to the solution to $\partial_a S = \partial_d S = 0$[1].

9.1 Expressing quantities in terms of *a* and *d*

As before, we begin with the two identities:

$$\Delta b = -\left(\frac{c}{1-c}\right)^2 \Delta a - 2\left(\frac{c}{1-c}\right) \Delta d \tag{46}$$

[1] Using the fixed matrix M_0 is less efficient, but more robust, than applying Newton's method.

and

$$\Delta\tau = 3ec(1-c)\left(\frac{c}{1-c}\Delta a + \Delta d\right)^2 \\ +c^3\Delta a^3 + 3c^2(1-c)\Delta a\Delta d^2 + 3c(1-c)^2\Delta b\Delta d^2 + (1-c)^3\Delta b^3. \tag{47}$$

The difference is that now Δd is of order 1, so the first term in the expansion of $\Delta\tau$, which is $\Omega(c)$, dominates the remaining terms, which are $O(c^2)$. Using (46), we have

$$3c(1-c)^2\Delta b\Delta d^2 = -3c^3\Delta a\Delta d^2 - 6c^2(1-c)\Delta d^3,$$
$$(1-c)^3\Delta b^3 = -8c^3\Delta d^3 - \frac{12c^4}{1-c}\Delta a\Delta d^2 - \frac{6c^5}{(1-c)^2}\Delta a^2\Delta d - \frac{c^6}{(1-c)^3}\Delta a^3$$
$$3ec\left(\frac{c}{1-c}\Delta a + \Delta d\right)^2 = 3ec\left(\Delta d^2 + c\Delta d(2\Delta a - \Delta d) + \frac{c^2}{(1-c)^2}\Delta a^2\right),$$

which in turn gives

$$\Delta\tau = 3ec\Delta d^2 \\ +c^2\left(6e\Delta a\Delta d - 3e\Delta d^2 + 3\Delta a\Delta d^2 - 6\Delta d^3\right) \\ +c^3\left(\frac{3e\Delta a^2}{1-c} + \Delta a^3 - 6\Delta a\Delta d^2 - 2\Delta d^3\right) \\ -\frac{12c^4}{1-c}\Delta a\Delta d^2 - \frac{6c^5}{(1-c)^2}\Delta a^2\Delta d - \frac{c^6}{(1-c)^3}\Delta a^3. \tag{48}$$

This expression is exact. Given estimates of Δd to order $\Delta\tau^k$ and Δa to order $\Delta\tau^\ell$, it determines c to order $\Delta\tau^{k+1}$ or $\Delta\tau^{\ell+2}$, whichever is larger. As long as the estimates of Δa and Δd only involve integer powers of $\Delta\tau$, so will the estimates of c.

In particular, given values of Δa and Δd, we have

$$c = \frac{\Delta\tau}{3d\Delta d^2} - c^2\left(\frac{2\Delta a}{\Delta d} - 1 + \frac{\Delta a - 2\Delta d}{e}\right) + O(\Delta\tau^3)$$

The fact that c can be expanded as a polynomial in c whose coefficients are rational functions of a and d means that we can compute $\partial_a c$ and $\partial_d c$ to within $O(\Delta\tau^3)$ by taking the derivative of this expression. Differentiating implicitly, we have

$$0 = \partial_a c\left(3e\Delta d^2 + 6c(2e\Delta a\Delta d - e\Delta d^2 + \Delta a\Delta d^2 - 2\Delta^2) + O(c^2)\right) \\ +c^2(6e\Delta d + 3\Delta d^2) + O(\Delta\tau^3),$$

which gives

$$\partial_a c = -c^2\left(\frac{2}{\Delta d} + \frac{1}{e}\right) + O(\Delta\tau^3).$$

Similarly,

$$\begin{aligned}0 = \partial_d c &\left(3e\Delta d^2 + 6c(2e\Delta a\Delta d - e\Delta d^2 + \Delta a\Delta d^2 - 2\Delta\tau^2) + O(c^2)\right)\\ &+6ec\Delta d + 6c^2(e\Delta a - e\Delta d + \Delta a\Delta d - 3\Delta d^3) + O(\Delta\tau^3).\end{aligned}$$

After a little algebra, this yields

$$\partial_d c = \frac{-2c}{\Delta d}\left(1 + c\left(1 - 3\frac{\Delta a}{\Delta d} - \frac{\Delta a}{e} + \frac{\Delta d}{e}\right)\right) + O(\Delta\tau^3).$$

Taking the derivative of (46) with respect to a and d then gives

$$\begin{aligned}\partial_a b &= c^2\left(3 + 2\frac{\Delta d}{e}\right) + O(\Delta\tau^3),\\ \partial_d b &= 2c\left(1 + c\left(5 - 4\frac{\Delta a}{\Delta d} - 2\frac{\Delta a}{e} + 2\frac{\Delta d}{e}\right)\right) + O(\Delta\tau^3).\end{aligned}$$

9.2 Solving $\partial_a S = \partial_d S = 0$

We now compute the partial derivatives of S, beginning with $\partial_d S$ to order $\Delta\tau$, obtaining:

$$\partial_d S = 2cH'(d) + 2cH'(b) - \frac{4c}{\Delta d}(H(d) - H(b)) + O(\Delta\tau^2),$$

Setting this equal to zero and dividing by $2c$ gives

$$H'(d) + H'(b) - \frac{2}{\Delta d}(H(d) - H(b)) = O(\Delta\tau),$$

implying that

$$d = 1 - b + O(\Delta\tau) = 1 - e + O(\Delta\tau).$$

We therefore write

$$d = 1 - e + d_1, \qquad \Delta d = 1 - 2e + d_1,$$

where d_1 is a quantity of order $\Delta\tau$. This allows us to express quantities like $H(d)$, $H'(d)$, $H''(d)$, Δb, $H(b)$, $H'(b)$ and $H''(b)$ in terms of d_1.

We next compute $\partial_a S$ to order $\Delta\tau^2$.

$$\begin{aligned}\partial_a S &= c^2 H'(a) + (1-c)^2 \partial_a b H'(b) \\ &\quad + 2\partial_a c(H(d) - H(b) + c(H(a) + H(b) - 2H(d))) \\ &= c^2 H'(a) + c^2 \left(3 + \frac{2\Delta d}{e}\right) H'(b) + O(\Delta\tau^3).\end{aligned}$$

Setting $\partial_a S$ equal to zero gives

$$\begin{aligned}H'(a) &= -\left(3 + \frac{2\Delta d}{e}\right) H'(b) + O(\Delta\tau) \\ &= -\left(3 + \frac{2(1-2e)}{e}\right) H'(e) + O(\Delta\tau) \\ &= \left(1 - \frac{2}{e}\right) H'(e) + O(\Delta\tau).\end{aligned}$$

Let a_0 be the solution to

$$H'(a_0) = \left(1 - \frac{2}{e}\right) H'(e),$$

which happens to be

$$a_0 = \left(1 + \left(\frac{e}{1-e}\right)^{\frac{2}{e}-1}\right)^{-1}.$$

We then have $a = a_0 + O(\Delta\tau)$.

To complete the proof of Theorem 2, we must show that there is an iterative scheme for approximating (a, b, c, d), and therefore S, for fixed values of $(e, \Delta\tau)$ with $\Delta\tau$ sufficiently small. As with our analysis of $\mathcal{O}_1$, we do this by computing the Hessian of S.

Our previously computed first derivatives of b and c are, to leading order:

$$\begin{aligned}\partial_a b &= c^2 \left(3 + \frac{2\Delta d}{e}\right) + O(\Delta\tau^3), \\ \partial_a c &= -c^2 \left(\frac{2}{\Delta d} + \frac{1}{e}\right) + O(\Delta\tau^3), \\ \partial_d b &= 2c + O(\Delta\tau^2), \\ \partial_d c &= -\frac{2c}{\Delta d} + O(\Delta\tau^2).\end{aligned}$$

Our second partials are:

$$\partial_{aa}^2 b = O(\Delta\tau^3),$$

$$\partial^2_{ad} b = -6\frac{c^2}{e} - 12\frac{c^2}{\Delta d} + O(\Delta\tau^3),$$
$$\partial^2_{dd} b = -4\frac{c}{\Delta d} + O(\Delta\tau^2),$$
$$\partial^2_{aa} c = O(\Delta\tau^3),$$
$$\partial^2_{ad} c = \frac{c^2}{\Delta d^2}\left(10 + \frac{4\Delta d}{e}\right) + O(\Delta\tau^3),$$
$$\partial^2_{dd} c = 6\frac{c}{\Delta d^2} + O(\Delta\tau^2).$$

Next we compute the partial derivatives of $S = c^2 H(a) + 2c(1-c)H(d) + (1-c)^2 H(b)$. Since $\partial_a b$ and $\partial_a c$ are $O(\Delta\tau^2)$, we have

$$\partial_a S = \partial_a b H'(b) + 2\partial_a c(H(d) - H(b)) + c^2 H'(a) + O(\Delta\tau^3).$$

Taking another derivative with respect to a, only one term contributes at order $\Delta\tau^2$:

$$\partial^2_{aa} S = c^2 H''(a_0) + O(\Delta\tau^3).$$

If instead we take a derivative of $\partial_a S$ with respect to d, three terms contribute at order $\Delta\tau^2$:

$$\begin{aligned}\partial^2_{ad} S &= 2c\partial_d c H'(a) + \partial^2_{ad} b H'(b) + 2\partial_d c H'(d) + O(\Delta\tau^3)\\ &= -\frac{4c^2}{\Delta d} H'(a) - 6c^2\left(\frac{1}{e} + \frac{2}{\Delta d}\right) H'(b) - 2c^2\left(\frac{1}{e} + \frac{2}{\Delta d}\right) H'(d) + O(\Delta\tau^3).\end{aligned}$$

Using the fact that $H'(b) \approx H'(e)$, $H'(d) \approx -H'(e)$ and $H'(a) \approx (1 - 2/e)\, H'(e)$, this simplifies to

$$\partial^2_{ad} S = \frac{4c^2(1-e)}{e(1-2e)} H'(e) + O(\Delta\tau^3).$$

Finally, we compute $\partial^2_{dd} S$. We have

$$\begin{aligned}\partial_d S &= \partial_d b H'(b) + 2cH'(d) + 2\partial_d c(H(d) - H(b)) + O(\Delta\tau^2).\\ \partial^2_{dd} S &= \partial^2_{bb} b H'(b) + 4\partial_d c H'(d) + 2cH''(d) + 2\partial^2_{dd} c(H(d) - H(b)) + O(\Delta\tau^2)\\ &= -\frac{4c}{\Delta d} H'(b) - \frac{8c}{\Delta d} H'(d) + 2cH''(d) + O(\Delta\tau^2)\\ &= 2c\left(H''(e) + \frac{2}{1-2e} H'(e)\right) + O(\Delta\tau^2).\end{aligned}$$

The upshot is that $\partial^2_{aa} S$ and $\partial^2_{ad} S$ are nonzero constants times c^2 plus $O(\Delta\tau^3)$, while $\partial^2_{dd} S$ is a nonzero constant times c plus $O(\Delta\tau^2)$. Replacing c with $\Delta\tau/(3e(1-2e)^2)$

and throwing away the error terms, we obtain a non-singular matrix M_0, with no explicit dependence on a or d, that comes within

$$\begin{pmatrix} O(\Delta\tau^3) & O(\Delta\tau^3) \\ O(\Delta\tau^3) & O(\Delta\tau^2) \end{pmatrix}$$

of the actual Hessian for all (a, d) within $O(\Delta\tau)$ of $(a_0, 1-e)$. In particular the iteration (45), beginning at $(a, d) = (a_0, 1-e)$, converges to the unique solution to $\partial_a S = \partial_d S = 0$ for all sufficiently small values of $\Delta\tau$. This completes the proof of Theorem 2.

10 Entropy estimates

Having established the properties of the optimizing graphons in $\mathcal{O}_1$ and $\mathcal{O}_2$ we now consider the asymptotic values of the the entropy function as τ approaches ε^3, both from above and from below.

10.1 The $\mathcal{O}_2$ region with $\varepsilon = e, \tau > e^3, e \neq 1/2$.

Let

$$\begin{aligned} c_1 &= \frac{\Delta\tau}{3e(1-2e)^2} \\ c_2 &= -2\frac{c_1 d_1}{1-2e} - c_1^2\left(\frac{2(a_0-e)-(1-2e)}{1-2e} + \frac{a_0-e-2(1-2e)}{e}\right). \end{aligned} \tag{49}$$

By our previous estimates, we have

$$c = c_1 + c_2 + O(\Delta\tau^3).$$

We estimate the terms in (31) using

$$\begin{aligned} H(b) &= H(e) + H'(e)\Delta b + \frac{H''(e)}{2}\Delta b^2 + O(\Delta\tau^3) \\ &= H(e) - H'(e)\left(\frac{2c}{1-c}(1-2e+d1) + c^2\Delta a\right) \\ &\quad +2c^2(1-2e)^2 H''(e) \\ &= H(e) - 2c(1-2e)H'(e) - 2cd_1 H'(e) \\ &\quad +c^2(2H''(e)(1-2e)^2 \\ &\quad -H'(e)(a_0 - e + 2(1-2e))) + O(\Delta\tau^3) \\ H(d) - H(b) &= H'(1-e)d_1 - H'(e)\Delta b + O(\Delta\tau^2) \\ &= H'(e)(2c(1-2e) - d_1) + O(\Delta\tau^2) \\ H(a) + H(b) - 2H(d) &= H(a) - H(e) + O(\Delta\tau). \end{aligned}$$

Plugging these terms into (31) yields

$$S = H(e) - 2(c_1 + c_2)(1 - 2e + d_1)H'(e) - 2c_1 d_1 H'(e) \\ + c_1^2 \left(H(a_0) - H(e) + H'(e)(2(1-2e) - (a_0 - e)) + 2H''(e)(1-2e)^2 \right) \\ + O(\Delta\tau^3)$$

Substituting for c_2 from (49), we find that the $c_1 d_1$ terms cancel, leaving us with

$$\begin{aligned} S = {} & H(e) - 2c_1(1-2e)H'(e) + 2c_1^2 H'(e) \\ & \left(2(a_0 - e) - (1-2e) + \frac{(1-2e)(a_0 - e - 2(1-2e))}{e} \right) \\ & + c_1^2 \left(H(a_0) - H(e) + H'(e)(2(1-2e) - (a_0 - e)) + 2H''(e)(1-2e)^2 \right) \\ & + O(\Delta\tau^3) \\ = {} & H(e) - 2c_1(1-2e)H'(e) \\ & + c_1^2 \left(H(a_0) - H(e) + H'(e) \left(3(a_0 - e) + \frac{2(1-2e)}{e}(a_0 + 3e - 2) \right. \right. \\ & \left. \left. + 2H''(e)(1-2e)^2 \right) \right) \\ & + O(\Delta\tau^3) \\ = {} & H(e) - \frac{2\Delta\tau}{3e(1-2e)} H'(e) \\ & + \frac{\Delta\tau^2}{9e^2(1-2e)^4} \left(H(a_0) - H(e) + H'(e) \left(3(a_0 - e) + \frac{2(1-2e)}{e}(a_0 + 3e - 2) \right) \right) \\ & + \frac{2\Delta\tau^2 H''(e)}{9e^2(1-2e)^2} + O(\Delta\tau^3). \end{aligned}$$

The linear coefficient is of course negative, since $H'(e)$ is positive for $e < 1/2$ and is negative for $e > 1/2$. The quadratic coefficient is more complicated but is also negative, diverging as $e^{-3}\ln(e)$ as $e \to 0$ and as $(e-1)^{-1}$ as $e \to 1$. See Fig. 3.

10.2 The region $\mathcal{O}_1$ with $\varepsilon = e, \tau < e^3, e > 1/2$

As before, we express the parameters (b, c, d) using the coordinates (a, μ). For convenience, we repeat the key identities:

$$\Delta b = \frac{c}{1-c} \left(\frac{c\Delta a}{1-c} - 2\mu \right),$$
$$\Delta d = \mu - \frac{c\Delta a}{1-c},$$
$$\frac{c\Delta a}{1-c} = -\delta + \frac{2\mu\delta^2 - \mu^2 e}{(2e-1)\delta} + O(\delta^4),$$

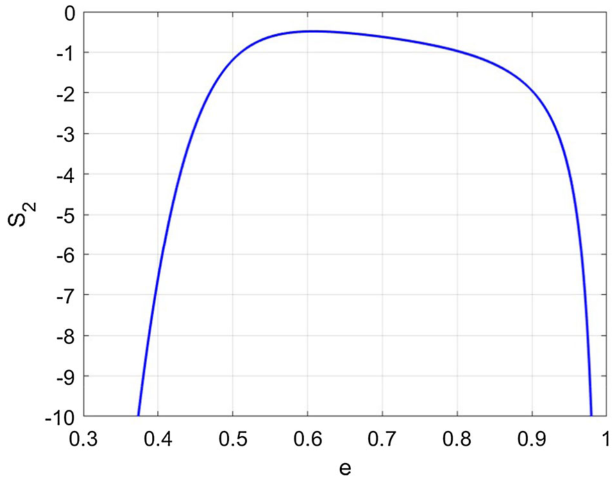

Fig. 3 The coefficient of $\Delta\tau^2$ as a function of e

$$c = \frac{\delta + \frac{\mu^2 e - 2\mu\delta^2}{(2e-1)\delta}}{2e - 1 + \delta - a_1} + O(\delta^4), \tag{50}$$

where $a = 1 - e + a_1$.

Since $S = H(b) + 2c(H(d) - H(b)) + c^2(H(a) + H(b) - 2H(d))$, we need to evaluate $H(b)$ through $O(\delta^4)$, $H(d)$ through $O(\delta^3)$ and $H(a)$ through $O(\delta^2)$:

$$\begin{aligned}
H(b) &= H(e) + H'(e)\Delta b + \frac{1}{2}H''(e)\Delta b^2 + O(\delta^6)\\
H(d) &= H(e) + H'(e)\Delta d + \frac{1}{2}H''(e)\Delta d^2 + \frac{1}{6}H'''(e)\Delta d^3 + O(\delta^4)\\
H(a) &= H(1-e) + H'(1-e)a_1 + \frac{1}{2}H''(1-e)a_1^2 + O(\delta^3)\\
&= H(e) - H'(e)a_1 + \frac{1}{2}H'(e)a_1^2 + O(\delta^3).
\end{aligned}$$

This makes

$$\begin{aligned}
S = H(e) &+ H'(e)\left((1-c)^2\Delta b + 2c(1-c)\Delta d - c^2 a_1\right)\\
&+ \frac{H''(e)}{2}\left((1-c)^2\Delta b^2 + 2c(1-c)\Delta d^2 + c^2 a_1^2\right)\\
&+ \frac{H'''(e)}{3}c(1-c)\Delta d^3 + O(\delta^5).
\end{aligned} \tag{51}$$

The coefficient of $H'(e)$ is

$$c(1-c)\left(\frac{c\Delta a}{1-c} - 2\mu\right) + 2c(1-c)\left(\mu - \frac{c\Delta a}{1-c}\right) - c^2 a_1 = c^2(2e - 1 - 2a_1).$$

Squaring and expanding the formula for c in (50) gives

$$
\begin{aligned}
c^2 &= \frac{\delta^2 + 2\frac{\mu^2 e - 2\mu\delta^2}{2e-1}}{(2e-1+\delta-a_1)^2} + O(\delta^5)\\
&= \frac{\delta^2}{(2e-1)^2} - \frac{2\delta^2(\delta-a_1)}{(2e-1)^3} + \frac{3\delta^2(\delta-a_1)^2}{(2e-1)^4} + \frac{2(\mu^2 e - 2\mu\delta^2)}{(2e-1)^3} + O(\delta^5).
\end{aligned}
$$

Multiplying by $(2e-1-2a_1)$, and using the fact that $a_1 = -\delta + O(\delta^2)$, the coefficient of $H'(e)$ becomes

$$
\frac{\delta^2}{(2e-1)} - \frac{2\delta^3}{(2e-1)^2} + \frac{4\delta^4}{(2e-1)^3} + \frac{2(\mu^2 e - 2\mu\delta^2)}{(2e-1)^2} + O(\delta^5).
$$

Next we consider the coefficient of $(1/2)H''(e)$, namely

$$
\begin{aligned}
&(1-c)\Delta b^2 + 2c(1-c)\Delta d^2 + c^2 a_1^2\\
&\quad = c^2\left(\frac{c\Delta a}{1-c} - 2\mu\right)^2 + 2c(1-c)\left(\mu - \frac{c\Delta a}{1-c}\right)^2 + c^2 a_1^2\\
&\quad = (c^2 + 2c(1-c))\left(\frac{c\Delta a}{1-c}\right)^2 - 4\mu c^2\Delta a + c^2 a_1^2.
\end{aligned}
$$

Using the third identity of (50), this reduces to

$$
(2c-c^2)\delta^2 + 4\mu c^2(2e-1) + c^2 a_1^2 + O(\delta^5).
$$

But $a_1 \approx -\delta$ and

$$
2c - c^2 = \frac{2\delta}{2e-1} - 5\frac{\delta^2}{(2e-1)^2} + O(\delta^3),
$$

so this becomes

$$
\frac{2\delta^3}{2e-1} - \frac{4\delta^4}{(2e-1)^2} + \frac{4\mu\delta^2}{2e-1} + O(\delta^5).
$$

Combining everything gives

$$
\begin{aligned}
S = H(e) &+ \frac{\delta^2}{2e-1}H'(e) + \delta^3\left(\frac{H''(e)}{2e-1} - \frac{2H'(e)}{(2e-1)^2}\right)\\
&+\delta^4\left(\frac{4H'(e)}{(2e-1)^3} - \frac{2H''(e)}{(2e-1)^2} + \frac{H'''(e)}{3(2e-1)}\right)\\
&+\mu\delta^2\left(\frac{2H''(e)}{2e-1} - \frac{4H'(e)}{(2e-1)^2}\right) + \frac{2\mu^2 e H'(e)}{(2e-1)^2} + O(\delta^5).
\end{aligned} \tag{52}
$$

This is a quadratic function of μ that is maximized at

$$\mu = \frac{\delta^2}{eH'(e)}\nu + O(\delta^3),$$

where

$$\nu = H'(e) - \left(e - \frac{1}{2}\right) H''(e),$$

in agreement with (32). Our final total, involving only the parameters e and δ, is then

$$S = H(e) + \frac{\delta^2}{2e-1}H'(e) - \frac{2\delta^3\nu}{(2e-1)^2} + \delta^4\left(\frac{H'''(e)}{3(2e-1)} + \frac{4\nu}{(2e-1)^3} - \frac{2\nu^2}{eH'(e)(2e-1)^2}\right) + O(\delta^5).$$

Although this formula for the entropy describes a smooth function of δ at $\delta = 0$, it is not a smooth function of $\tau = e^3 - \delta^3$. The first and second derivatives of $s(\varepsilon, \tau)$ with respect to τ are positive and diverge as δ^{-1} and δ^{-4}, respectively, as $\delta \to 0$.

Comparing the entropy functions in the regions $\mathcal{O}_1$ and $\mathcal{O}_2$, we see that the entropy is continuous in τ at $\tau = \varepsilon^3$, but the first derivative is not. The first derivative approaches a finite value as τ approaches ε^3 from above, as does the second derivative, but these quantities diverge as τ approaches ε^3 from below. This completes the proof of Theorem 3.

11 Longer cycles

Finally, we turn to the edge-k-cycle problem for odd $k > 3$ and prove Theorem 4. Before we discuss the proof we explain why we limit ourselves to odd k.

The density $\tau_k(g)$ of k-cycles for the graphon g is the trace of g^k, where we view g as an integral operator. That is, $\tau_k(g) = \sum_i \lambda_i^k$, where $\lambda_1, \lambda_2, \ldots$ are the eigenvalues of g. The leading eigenvalue λ_1 is always at least e. When k is odd, negative eigenvalues can make the sum smaller than e^k, but when k is even, all terms in the sum are non-negative and $\tau_k(g) \geq e^k$. In this paper we mostly study graphons with $\tau_k(g)$ slightly less than e^k, which only exist when k is odd.

The strategy of our proof for Theorem 4 is essentially the same as in the proof of Theorems 1–3.

The part of the proof concerning $\tau_k < e^k$ follows these steps:

(1) We show that, for any graphon, τ_k is bounded below by $e^k - \|\delta g\|_2^k$, with equality only when δg has rank 1 with eigenvalue $-\delta$ and the degree function is constant. Combined with our existing estimates of a graphon with $\|\delta g\|_2$ of fixed size, this gives upper bounds on the entropy of any graphon with the given values of e and δ. These bounds, as a function of e and δ, are identical to those obtained for triangles.

(2) A bipodal graphon with $a = 1-e$ and $d = 1+\delta$ comes within $O(\delta^3)$ of achieving that upper bound. This bounds the extent to which an optimal graphon can differ from our model graphon. These bounds are sharper than in the edge-triangle model. In the edge-triangle model, we had the a priori estimate $\int_0^1 \delta d(x)^2\,dx = O(\delta^4)$, but with k-cycles we have $\int_0^1 \delta d(x)^2\,dx = O(\delta^{k+1})$. When dealing with bipodal graphons, this means that the parameter μ is a priori $O(\delta^{k/2})$ instead of $O(\delta^{3/2})$. (The μ parameter will eventually turn out to be $O(\delta^{k-1})$.) We also obtain bounds on how far δg is (in an L^2 sense) from having rank one. These bounds are qualitatively similar to those derived for the edge-triangle model.

(3) The argument that g is approximately bipodal, taking on values that are close to e or $1-e$, proceeds exactly as before.

(4) The averaging argument is similar to before. If the optimal graphon weren't already bipodal, we could average its value over each quadrant to get a bipodal graphon with the same value of ε, and a quantifiable trade-off between τ_k and entropy. Moreover, a sharp trade-off would imply the same structural properties of g that we had in Sect. 6.

(5) If we knew that the optimal graphon were bipodal, we could maximize the entropy as in Sect. 10, expressing all parameters in terms of Δa and μ. There is a positive term proportional to $\delta^2\mu$, as in the edge-triangle model, and a negative term proportional to $\mu^2\delta^{3-k}$. The optimal value of μ is then $O(\delta^{k-1})$, implying that $d = e + \delta + O(\delta^{k-1})$ and the contribution of μ to the entropy is $O(\delta^{k+1})$. This shows that there is at most one possiblity for a bipodal optimizer, and gives good estimates for its parameters.

(6) Combining the last two steps, if the original graphon were not optimal, we could replace it with the averaged graphon of Step 3, then perturb the bipodal parameters using the estimates from Step 4 to construct a graphon with the same ε and τ_k, but larger entropy.

In the limit as $k \to \infty$, our problem reduces to finding a graphon of the form $g(x,y) = e-\delta v(x)v(y)$, where $\int_0^1 v(x)\,dx = 0$ and $\int_0^1 v(x)^2\,dx = 1$, that maximizes the entropy for fixed e and δ. This graphon is bipodal, with d and $-c\Delta a/(1-c)$ exactly equal to δ. Theorem 4 only gives $a = 1-e-\delta+O(\delta^2)$ to first order in δ, but there is no obstacle to computing higher coefficients. The parameters for the optimal graphon for any fixed k are all within $O(\delta^{k-1})$ of the parameters for this limiting graphon.

We now begin the proof.

Step 1 Writing $g(x,y) = e + \delta g(x,y)$, we have

$$\tau_k = \mathrm{Tr}(g^k) = \mathrm{Tr}((e+\delta g)^k).$$

Expanding this out gives $e^k + \mathrm{Tr}(\delta g^k)$ plus a number of cross terms. Some of the cross terms are of the form $e^{k-m}\langle 1, T_{\delta g^m} 1\rangle$, where 1 denotes the constant function in $L^2([0,1])$. Others are powers of e times the product of several $\langle 1, T_{\delta g^m} 1\rangle$ expressions.

The expression $\langle 1, T_{\delta g^m} 1\rangle$ is zero if $m = 1$, is $\int_0^1 \delta d(x)^2\,dx$ if $m = 2$, and is bounded by $\|\delta g\|_{op}^{m-2} \int_0^1 \delta d(x)^2\,dx$ if $m > 2$. Since δg (viewed as an integral operator) is self-adjoint, its operator norm is the size of its largest eigenvalue, which is bounded by its L^2 norm, which is small compared to 1. Of course, the product of two or more expressions of the form $\langle 1, T_{\delta g^m} 1\rangle$ is smaller still.

The upshot is that

$$\begin{aligned}\tau_k &= e^k + \mathrm{Tr}(\delta g^k) + ke^{k-2}\int_0^1 \delta d(x)^2\,dx + \\ &\quad + \text{terms small compared to } \int_0^1 \delta d(x)^2\,dx \\ &\geq e^k + \mathrm{Tr}(\delta g^k) + \frac{ke^{k-2}}{2}\int_0^1 \delta d(x)^2\,dx.\end{aligned}$$

The last term is of course positive-definite. This implies that $\|\delta g\|_2 \geq \delta$, with equality (if and) only if δg is rank-1 with eigenvalue $-\delta$ and $\int_0^1 \delta d(x)^2\,dx = 0$.

We have previously shown that any graphon with $\|\delta g\| \geq \delta$ has entropy at most

$$H(e) + \frac{\delta^2 H'(e)}{2e-1} + O(\delta^3),$$

so this is also an upper bound on the entropy of any graphon with $\tau_k = e^k - \delta^k$.

Step 2 Consider the bipodal graphon with $a = 1-e$, $b = e - \delta^2/(2e-1)$, $c = \delta/(2e-1+\delta)$ and $d = e+\delta$. This graphon has edge density e, k-cycle density $e^k - \delta^k$, and entropy $H(e) + \delta^2 H'(e)/(2e-1) + O(\delta^3)$. That is, it comes within $O(\delta^3)$ of saturating the upper bound.

This means that the errors in step 1, both those involving $\int_0^1 \delta d(x)^2\,dx$ being nonzero and those from comparing $\mathrm{Tr}(\delta g^k)$ to $\|\delta g\|_2^k$, cannot cost us more than $O(\delta^3)$ in entropy, or equivalently cannot cause $\|\delta g\|_2^2$ to be greater than $\delta^2 + O(\delta^3)$.

In particular, since an $O(\delta^3)$ change in $\|\delta g\|_2^2$ corresponds to an $O(\delta^{k+1})$ change in $\|\delta g\|_2^k$, we must have

$$\int_0^1 \delta d(x)^2\,dx = O(\delta^{k+1}).$$

Furthermore, the contribution of the small eigenvalues of δg to $\|\delta g\|_2^2$ is at most $O(\delta^3)$. That is, δg is within $O(\delta^{3/2})$, in an L^2-sense, of being rank-1.

Step 3 Approximate bipodality (i.e., Proposition 6) proceeds exactly as in the $k = 3$ case, because the only ingredients in the proof of Proposition 6 were the estimates from Step 1. The error terms in Proposition 6 are the same order in δ for every k.

Step 4 Let h be the graphon obtained by averaging g on podes, and write $g = h + \Delta h$. We need to estimate the changes to entropy and τ_k in going from g to h, as was done in Sect. 6 for $k = 3$. The change in entropy follows exactly as in Lemma 20:

$$S(g) \leq S(h) - C(e)\|\Delta h\|_2^2,$$

and the inequality is sharp if and only if the non-zero values of Δh are close to $1-2e$ on $C_1 \times C_1$ and $2e-1$ elsewhere.

To bound the change in τ_k, note that h (being bipodal) has rank two, and that its eigenvalues are approximately $\lambda_1(h) \approx e$ and $\lambda_2(h) \approx -\delta$. Since Δh integrates to zero on each pode, $\langle w, T_{\Delta h} w\rangle = 0$ for every non-trivial eigenfunction w of h. It follows that the eigenvalues of $g = h + \Delta h$ are

$$\begin{aligned}
\lambda_1(g) &= \lambda_1(h) + \frac{|\langle u_1, T_{\Delta h} u_2\rangle|^2}{e}(1+O(\delta)) \geq \lambda_1(h)\\
\lambda_2(g) &= \lambda_2(h) - \frac{|\langle u_1, T_{\Delta h} u_2\rangle|^2}{\delta}(1+O(\delta)) \geq \lambda_2(h) - \frac{\|\Delta h\|_2^2}{2\delta}(1+O(\delta)),
\end{aligned}$$

where u_1 and u_2 are the non-trivial (normalized) eigenfunctions of h. Moreover, the second inequality is sharp if and only if $\Delta h(x,y) = u_2(x) T_{\Delta h} u_2(y) + u_2(y) T_{\Delta h} u_2(x) + w(x,y)$ for some remainder w with $\|w\|_2 = o(\|\Delta h\|_2)$. All other eigenvalues of h are zero, and so all other eigenvalues of g are bounded in absolute value by $\|\Delta h\|_2 = O(\delta^{3/2})$, and also the sum of their squares is at most $\|\Delta h\|_2^2$. In particular, $\sum_{i\geq 3} |\lambda_i(g)|^k \leq \|\Delta h\|_2^{k-2} \sum_i |\lambda_i(g)|^2 \leq \|\Delta h\|_2^k$. Since $\tau_k(h) = \lambda_1^k(h) + \lambda_2^k(h)$ and $\tau_k(h) = \sum_i \lambda_i^k(g)$, we have

$$\begin{aligned}
\tau_k(h) &\leq \lambda_1^k(g) + \left(\lambda_2(g) + \frac{\|\Delta h\|_2^2}{2\delta}(1+O(\delta))\right)^k + \sum_{i\geq 3} |\lambda_i(g)|^k\\
&\leq \lambda_1^k(g) + \lambda_2^k(g) - k\frac{\delta^{k-2}\|\Delta h\|_2^2}{2}(1+O(\delta)) + \sum_{i\geq 3} |\lambda_i(g)|^k\\
&\leq \tau_k(g) - k\frac{\delta^{k-2}\|\Delta h\|_2^2}{2}(1+O(\delta)),
\end{aligned}$$

and the bound is sharp if and only if $\Delta h \approx u_2(x) T_{\Delta h} u_2(y) + u_2(y) T_{\Delta h} u_2(x)$. This is the analogue of Lemma 22 for general k.

Step 5 When considering bipodal graphons, we can compute the two eigenvalues of g from the trace and the Hilbert-Schmidt norm:

$$\begin{aligned}
\lambda_1 + \lambda_2 &= e + c\Delta a + (1-c)\Delta b\\
&= e + \frac{d\Delta a}{1-c} - 2\mu c\\
\lambda_1^2 + \lambda_2^2 &= c^2 a^2 + 2c(1-c)d^2 + (1-c)^2 b^2\\
&= e^2 + c^2 \Delta a^2 + 2c(1-c)\Delta d^2 + (1-c)^2 \Delta b^2\\
&= e^2 + \left(\frac{c\Delta a}{1-c}\right)^2 - 4c\mu\left(\frac{c\Delta a}{1-c}\right) + 2\mu^2(c+c^2).
\end{aligned}$$

From this we compute the two eigenvalues

$$\lambda_{1,2} = \frac{1}{2}\left(e + \frac{c\Delta a}{1-c} - 2\mu c \pm \left(e - \frac{c\Delta a}{1-c} + 2\mu c\right)\sqrt{1 + \frac{4\mu^2 c(1-c)}{\left(e - \frac{c\Delta a}{1-c} + 2\mu c\right)^2}}\right).$$

These work out to

$$\lambda_1 = e + \frac{\mu^2 c(1-c)}{e + 2\mu c - \frac{c\Delta a}{1-c}} + O(\mu^4\delta^2) \tag{53}$$

$$\lambda_2 = \frac{c\Delta a}{1-c} - 2\mu c - \frac{\mu^2 c(1-c)}{e + 2\mu c - \frac{c\Delta a}{1-c}} + O(\mu^4\delta^2) \tag{54}$$

Note that if $\mu = 0$, then our eigenvalues are exactly e and $\frac{c\Delta a}{1-c}$, so we have $e^k - \delta^k = \lambda_1^k + \lambda_2^k = e^k + \left(\frac{c\Delta a}{1-c}\right)^k$, so $\frac{c\Delta a}{1-c} = -\delta$. We are quantifying how much they differ from these values when μ is nonzero. However, we know that $\mu = O(\delta^{k/2})$, so these corrections are small.

Setting the full expressions for $\lambda_1^k + \lambda_2^k$ equal to $e^k - \delta^k$, and using the fact that $c = \delta/(2e-1) + O(\delta^2)$, gives

$$\left(\frac{c\Delta a}{1-c}\right)^k = -\delta^k - \frac{ke^{k-2}\mu^2\delta}{2e-1} + \frac{2k\mu\delta^k}{2e-1} + O(\mu^2\delta^2, \mu\delta^{k+1}).$$

Taking kth roots, we then have

$$\frac{c\Delta a}{1-c} = -\delta + \frac{2\mu\delta}{2e-1} - \frac{\mu^2 e^{k-2}}{(2e-1)\delta^{k-2}} + O(\mu\delta^2, \mu^2\delta^{3-k}), \tag{55}$$

which in turn implies that

$$c = \frac{\delta + \frac{\mu^2 e^{k-2}\delta^{2-k} - 2\mu\delta}{2e-1}}{2e - 1 + \delta - a_1} + O(\mu\delta^2, \mu^2\delta^{3-k}). \tag{56}$$

We now expand the entropy, using Taylor series around e and $1-e$, exactly as in Sect. 10. As before, the coefficient of $H'(e)$ is $c^2(2e-1-2a_1)$. Substituting for c using (56), the leading order contribution of μ to our $H'(e)$ term becomes

$$\frac{2\mu^2 e^{k-2}\delta^{3-k} - 4\mu\delta^2}{(2e-1)^2} H'(e).$$

The coefficient of $(1/2)H''(e)$ is

$$(2c - c^2)\left(\frac{c\Delta a}{1-c}\right)^2 - 4\mu c^2 \Delta a + c^2 a_1^2,$$

exactly as before. The leading contribution of μ is the second term. Finally, all contributions of μ to the coefficients of higher derivatives of H are higher order. Thus, to leading order, the contribution of μ to the entropy is

$$\frac{2\mu\delta^2 H''(e)}{2e-1} - \frac{4\mu\delta^2 H'(e)}{(2e-1)^2} + \frac{2\mu^2 e^{k-2} H'(e)}{\delta^{k-3}(2e-1)^2} = \frac{4\mu e^{k-2}\delta^{3-k} H'(e) - \delta^2 \nu}{(2e-1)^2}. \quad (57)$$

Setting the derivative with respect to μ equal to zero then gives

$$\mu = \frac{\nu \delta^{k-1}}{e^{k-2} H'(e)}.$$

That is, $\mu = O(\delta^{k-1})$ and only contributes to the entropy at order δ^{k+1}. For all computations at lower order, and in particular for computing a, b, c, and d through order δ^{k-2}, we can simply set $\mu = 0$. That is, to this order the computations for the edge-k-cycle model are identical to the limiting problem where Δd and $-c\Delta a/(1-c)$ are set exactly equal to δ. In that limiting problem, $a = 1 - e - \delta + O(\delta^2)$. This then determines c and b. The entropy then follows from (51), or equivalently from (52) with μ set equal to zero.

Finally, we consider the Hessian of the entropy. The coefficient of μ^2 in expression (57) is large and negative, much larger than when $k = 3$. However, the linear term is the same as when $k = 3$, as are the contributions to S that don't involve μ. The upshot is that $\partial^2_{\mu\mu} S$ is more negative than when $k = 3$, while $\partial^2_{\mu a} S$ and $\partial^2_{aa} S$ are essentially the same. Thus the Hessian is negative definite and any optimizing bipodal graphon is unique.

Step 6 Starting from the bipodal graphon obtained from averaging in Step 4, we can parametrize it in terms of c, Δa, and μ as in Step 5. Then we construct a perturbation by increasing c while holding Δa and μ constant. From (53) to (54), we see that

$$\frac{d\lambda_1}{dc} = O(\mu^2), \quad (58)$$

$$\frac{d\lambda_2}{dc} = \Delta a + O(\delta). \quad (59)$$

It follows that

$$\frac{d\tau_k}{dc} = k\lambda_1^{k-1}\frac{d\lambda_1}{dc} + k\lambda_2^{k-1}\frac{d\lambda_2}{dc} = -k\delta^{k-1}(2e-1) + o(\delta^{k-1}),$$

where the last equality follows from $\Delta a \approx 1 - 2e$, $\lambda_2 \approx -\delta$, and $\mu^2 = O(\delta^k)$.

On the other hand, the change in entropy is $dS/dc = [2\delta/(2e-1)]D(1-e)(1+o(1))$, exactly as in Sect. 7. That is, starting from the averaged graphon h we can trade off entropy for k-cycles at the rate

$$\frac{dS}{d\tau_k} = -(1+o(1))\frac{2D(1-e)}{k\delta^{k-2}(2e-1)^2} = -(1+o(1))\frac{2C(e)}{k\delta^{k-2}}.$$

The rest of the proof of Theorem 4 for $\tau_k < e^k$ follows exactly as in the $k = 3$ case: if the perturbation of the averaged graphon h does not contradict the optimality of g, both the k-cycle change and entropy change inequalities of Step 4 must be sharp. It follows that Δh approximately takes the prescribed values and that $\Delta h \approx u_2(x)T_{\Delta h}u_2(y) + u_2(y)T_{\Delta h}u_2(x)$; but as in Sect. 7 these two properties contradict the definitions of C_1 and C_2 in Proposition 6. The rest of the proof of Theorem 4 for $\tau_k < e^k$ continues as in the case $k = 3$.

We now turn to $\tau_k > e^k$. A graph is said to be 2-*star-like* if all of its vertices have degree 1 or 2. In particular, all k-cycles are 2-star-like. Theorem 1.1 of [15] then states that the optimizing graphon for τ_k slightly above e^3 is bipodal with the parameters (a, b, c, d) taking the approximate forms indicated in Theorem 4, only with errors that are $o(1)$ or $o(\Delta\tau)$ instead of the $O(\Delta\tau)$ or $O(\Delta\tau^2)$ errors claimed in Theorem 4.

All that remains is to sharpen the estimates and compute the entropy. We follow the same procedure as in Sect. 9, only with the expressions (47) and (48) for $\Delta\tau$ replaced with

$$\begin{aligned}\Delta\tau_k = &\left(\frac{c}{1-c}\right)ke^{k-2}\mu^2 + ke^{k-3}c^2\mu^2(\Delta a + \Delta b - 2\Delta d)\\ &+ke^{k-4}c^2\mu^2\left(c(\Delta a - \Delta d)^2 + (1-c)(\Delta d - \Delta b)^2\right)\\ &+\frac{k(k-5)e^{k-4}c^2\mu^4}{2(1-c)^2} + O(c^3).\end{aligned}$$

The first three terms are the multiples of $\langle 1, T_{\delta g^2} 1\rangle$, $\langle 1, T_{\delta g^3} 1\rangle$ and $\langle 1, T_{\delta g^4} 1\rangle$, respectively, discussed in Step 1 of the proof for $\tau < e^k$. The last term involves the product of two factors of $\langle 1, T_{\delta g^2} 1\rangle$ and only occurs when $k \geq 7$. All other terms in the expansion of $\Delta\tau_k$, including $\mathrm{Tr}(\delta g^k)$, are of higher order.

The upshot is that the calculation is the same as for triangles, only with a coefficient of ke^{k-2} instead of $3e$ in the leading term, and with different $O(c^2)$ terms. Adjusting the $O(c^2)$ terms does not affect the computation of (a, b, c, d) to the order specified in Theorem 4. The change from $3e$ to ke^{k-2} in the leading term does change the $O(\Delta\tau)$ terms in the expansions of c and b, but does not affect the leading expressions for d or a, or the fact that the errors are indeed $O(\Delta\tau)^2)$ for b and c and are $O(\Delta\tau)$ for a and d. Plugging these values of (a, b, c, d) into the formula for the entropy then yields the estimate (11).

Data availibility Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

References

1. Augeri, F.: Nonlinear large deviation bounds with applications to Wigner matrices and sparse Erdős-Rényi graphs. Ann. Prob. **48**(5), 2404–2448 (2020)
2. Bhattacharya, B.B., Ganguly, S., Lubetzky, E., Zhao, Y.: Upper tails and independence polynomials in random graphs. Adv. Math. **319**, 313–347 (2017)
3. Borgs, C., Chayes, J., Lovász, L., Sós, V.T., Vesztergombi, K.: Convergent graph sequences I: subgraph frequencies, metric properties, and testing. Adv. Math. **219**, 1801–1851 (2008)
4. Borgs, C., Chayes, J., Lovász, L.: Moments of two-variable functions and the uniqueness of graph limits. Geom. Funct. Anal. **19**, 1597–1619 (2010)
5. Chatterjee, S., Dembo, A.: Nonlinear large deviations. Adv. Math. **299**, 396–450 (2016)
6. Chatterjee, S., Diaconis, P.: Estimating and understanding exponential random graph models. Ann. Statist. **41**, 2428–2461 (2013)
7. Chatterjee, S., Varadhan, S.R.S.: The large deviation principle for the Erdős-Rényi random graph. Eur. J. Comb. **32**, 1000–1017 (2011)
8. Cook, N., Dembo, A.: Large deviations of subgraph counts for sparse Erdős-Rényi graphs. Adv. Math. **373**, 107289 (2020)
9. Eldan, R.: Gaussian-width gradient complexity, reverse log-Sobolev inequalities and nonlinear large deviations. Geom. Funct. Anal. **28**(6), 1548–1596 (2018)
10. Ellis, R.: Entropy, Large Deviations, and Statistical Mechanics. Springer-Verlag, Berlin (2006)
11. Hales, T.C.: Dense Sphere Packings. A Blueprint for Formal Proofs, Cambridge University Press, Cambridge (2012)
12. Harel, M., Mousset, F., Samotij, W.: Upper tails via high moments and entropic stability, arXiv preprint arXiv:1904.08212 (2019)
13. Kenyon, R., Radin, C., Ren, K., Sadun, L.: Multipodal structures and phase transitions in large constrained graphs. J. Stat. Phys. **168**, 233–258 (2017)
14. Kenyon, R., Radin, C., Ren, K., Sadun, L.: The phases of large networks with edge and triangle constraints. J. Phys. A: Math. Theor. **50**, 435001 (2017)
15. Kenyon, R., Radin, C., Ren, K., Sadun, L.: Bipodal structure in oversaturated random graphs. Int. Math. Res. Notices **2016**, 1009–1044 (2018)
16. Lanford, O.E.: Entropy and equilibrium states in classical statistical mechanics. In: Lenard, A. (eds.) Statistical Mechanics and Mathematical Problems, Battelle Seattle 1971 Rencontres, vol. 20 of Springer Lecture Notes in Physics, pp. 1–113. Springer-Verlag, Berlin/Heidelberg/New York (1973)
17. Lovász, L.: Large Networks and Graph Limits. American Mathematical Society, Providence (2012)
18. Lovász, L., Szegedy, B.: Limits of dense graph sequences. J. Combin. Theory Ser. B **98**, 933–957 (2006)
19. Lovász, L., Szegedy, B.: Szemerédi's lemma for the analyst. GAFA **17**, 252–270 (2007)
20. Lovász, L., Szegedy, B.: Finitely forcible graphons. J. Combin. Theory Ser. B **101**, 269–301 (2011)
21. Löwen, H.: Fun with hard spheres, pages 295-331 in Statistical physics and spatial statistics : the art of analyzing and modeling spatial structures and pattern formation. In: Mecke, K.R., Stoyen, D. (eds.) Lecture Notes in Physics No. 554, Springer-Verlag, Berlin, (2000)
22. Neeman, J., Radin, C., Sadun, L.: Moderate deviations in triangle count, arXiv:2101.08249 (2021)
23. Pikhurko, O., Razborov, A.: Asymptotic structure of graphs with the minimum number of triangles. Combin. Probab. Comput. **26**, 138–160 (2017)
24. Radin, C., Sadun, L.: Phase transitions in a complex network. J. Phys. A: Math. Theor. **46**, 305002 (2013)
25. Radin, C., Sadun, L.: Singularities in the entropy of asymptotically large simple graphs. J. Stat. Phys. **158**, 853–865 (2015)
26. Razborov, A.: On the minimal density of triangles in graphs. Combin. Probab. Comput. **17**, 603–618 (2008)
27. Turán, P.: On an extremal problem in graph theory (in Hungarian). Matematikai é s Fizikai Lapok **48**, 436–452 (1941)