Combinatorics, Probability and Computing (2017) 26, 301–320.
doi:10.1017/S0963548316000262

On the Lower Tail Variational Problem for Random Graphs

YUFEI ZHAO[†]

Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK
(e-mail: yufei.zhao@maths.ox.ac.uk)

Received 11 March 2015; revised 25 April 2016; first published online 16 August 2016

We study the lower tail large deviation problem for subgraph counts in a random graph. Let X_H denote the number of copies of H in an Erdős–Rényi random graph $\mathcal{G}(n,p)$. We are interested in estimating the lower tail probability $\mathbb{P}(X_H \leqslant (1-\delta)\mathbb{E}X_H)$ for fixed $0 < \delta < 1$.

Thanks to the results of Chatterjee, Dembo and Varadhan, this large deviation problem has been reduced to a natural variational problem over graphons, at least for $p \geqslant n^{-\alpha_H}$ (and conjecturally for a larger range of p). We study this variational problem and provide a partial characterization of the so-called 'replica symmetric' phase. Informally, our main result says that for every H, and $0 < \delta < \delta_H$ for some $\delta_H > 0$, as $p \to 0$ slowly, the main contribution to the lower tail probability comes from Erdős–Rényi random graphs with a uniformly tilted edge density. On the other hand, this is false for non-bipartite H and δ close to 1.

2010 *Mathematics subject classification*: Primary 05C80
Secondary 05C35, 60F10

1. Background

We consider large deviations of subgraph counts in Erdős–Rényi random graphs. Fix a graph H, and let X_H denote the number of copies of H in an Erdős–Rényi random graph $\mathcal{G}(n,p)$. For a fixed $\delta > 0$, we consider the problem of estimating the probabilities

$$\text{(upper tail)} \quad \mathbb{P}(X_H \geqslant (1+\delta)\mathbb{E}X_H) \quad \text{and}$$
$$\text{(lower tail)} \quad \mathbb{P}(X_H \leqslant (1-\delta)\mathbb{E}X_H).$$

This problem has a long history (see [7] and its references). For the order of the logarithm of the tail probability, the upper tail problem is considered more difficult and it was resolved only fairly recently [7, 13], whereas the corresponding lower tail problem had been solved earlier [16, 17]. We are now interested in the finer question of determining

[†] The author was supported by a Microsoft Research PhD Fellowship.

the large deviation rate, or equivalently the first-order asymptotics of the logarithm of the tail probability.

Chatterjee and Varadhan [10] (the dense setting, with p constant) and more recently Chatterjee and Dembo [8] (the sparse setting, with $p \to 0$ and $p \geqslant n^{-\alpha_H}$ for some $\alpha_H > 0$) showed that this large deviation problem reduces to a natural variational problems in the space of graphons, which are a certain type of graph limit. We begin by reviewing this connection, and then we shift our attention to analysing the variational problem.

The language of graph limits is used throughout our discussion, so let us review some terminology. We refer the readers to the beautifully written monograph by Lovász [25] or the original sources (*e.g.* [5, 6, 26, 27]) for more on the subject. A *graphon* is a symmetric measurable function $W : [0,1]^2 \to [0,1]$ (here symmetric means $W(x,y) = W(y,x)$). We write $V(H)$ and $E(H)$ to mean the vertex and edge set of a graph H, respectively, and $v(H) = |V(H)|$ and $e(H) = |E(H)|$ to denote their cardinalities. For any graphs H and G, we write $\hom(H,G)$ to denote the number of graph homomorphisms from H to G. We let

$$t(H,G) := \hom(H,G)/v(G)^{v(H)}$$

denote the H-*density* in G. The H-density of a graphon W is defined by

$$t(H,W) := \int_{[0,1]^{V(H)}} \prod_{ij \in E(H)} W(x_i,x_j) \prod_{i \in V(H)} dx_i$$

(here W could be $\mathbb{R}$-valued). As usual, K_t denotes the complete graph on t vertices. As an example, we have

$$t(K_3,W) = \int_{[0,1]^3} W(x,y)W(x,z)W(y,z)\,dx\,dy\,dz.$$

The notion of *cut distance* is mentioned a few times in this paper, but it is not used in a substantial way, so we refer the readers to [25, Chapter 8] for details.

We write

$$I_p(x) := x \log \frac{x}{p} + (1-x) \log \frac{1-x}{1-p}$$

for the relative entropy function. For any function f, we write

$$\mathbb{E}[f(W)] := \int_{[0,1]^2} f(W(x,y))\,dx\,dy.$$

We begin with a review of what is known for upper tails. In the dense case, for fixed $0 < p \leqslant q < 1$, it was shown in [10] that as $n \to \infty$,

$$\log \mathbb{P}(t(H,\mathcal{G}(n,p)) \geqslant q^{e(H)}) = -(1+o(1))\frac{n^2}{2}\mathsf{UT}_p(H,q), \tag{1.1}$$

where $\mathsf{UT}_p(H,q)$, for any graph H, is given by the *upper tail variational problem*:

$$\mathsf{UT}_p(H,q) := \begin{cases} \text{minimize } \mathbb{E}[I_p(W)] \\ \text{subject to } t(H,W) \geqslant q^{e(H)}. \end{cases} \tag{1.2}$$

https://doi.org/10.1017/S0963548316000262 Published online by Cambridge University Press

Here W is taken over all graphons. We shall use $\mathsf{UT}_p(H,q)$ to refer to the variational problem as well as its value, that is,

$$\min\{\mathbb{E}[I_p(W)] : t(H,W) \geqslant q^{e(H)}\}$$

(it is known that the minimum is always attained by some W: see Lemma 5.1 below).

Furthermore, as shown in [10, Theorem 3.1], the set of minimizing W in $\mathsf{UT}_p(H,q)$ represents the most likely models for $\mathcal{G}(n,p)$ conditioned on the rare event of $t(H,\mathcal{G}(n,p)) \geqslant q^{e(H)}$, in the sense that the random graph conditioned on this rare event is exponentially more likely to be close (in cut distance) to the minimizing set of W. This motivates the study of $\mathsf{UT}_p(H,q)$ and related variational problems.

We currently have few tools for solving variational problems of the type (1.2). Note that $W \equiv q$ always satisfies the constraint in (1.2). We focus on the basic question: Does the constant graphon $W \equiv q$ minimize $\mathsf{UT}_p(H,q)$? The answer depends on the graph H and parameters (p,q). For a fixed H, we wish to determine for each (p,q) whether $\mathsf{UT}_p(H,q) = I_p(q)$ or $\mathsf{UT}_p(H,q) < I_p(q)$, and in the former case, whether the constant function $W \equiv q$ is the unique minimizer.[1] The separation of these two cases can be illustrated via a phase diagram, as in Figure 1, by plotting the phases in the (p,q)-plane according to the behaviour of $\mathsf{UT}_p(H,q)$.

The constant graphon $W \equiv q$ is the limit of random graphs $\mathcal{G}(n,q)$ as $n \to \infty$, so if it were the unique minimizer of $\mathsf{UT}_p(H,q)$ then $\mathcal{G}(n,p)$, conditioned on having H-density at least $q^{e(H)}$, approaches the typical $\mathcal{G}(n,q)$ in cut distance; this is not the case when $W \equiv q$ is not a minimizer. Borrowing language from statistical physics, informally, when $W \equiv q$ is a minimizer we say that there is *replica symmetry*,[2] and otherwise there is *symmetry breaking*.

In a previous paper with Lubetzky [28], we completely identified the upper tail replica symmetric phase whenever H is a d-regular graph. The phase diagram depends only on d. The diagram for $H = K_3$ is shown in Figure 1 in the upper portion (*i.e.* $q > p$) of the diagram.[3] The lower portion of the diagram illustrates new results in paper concerning the lower tail problem.

[1] We identify graphons differing on a measure zero set, as well as up to a measure-preserving transformation on $[0,1]$, that is, denoting $W^\sigma(x,y) := W(\sigma(x),\sigma(y))$, graphons W and U are identified if W^σ and U^τ agree up to a measure zero set for some measure-preserving maps $\sigma,\tau\colon [0,1] \to [0,1]$ [4].

[2] There is a subtle issue of uniqueness of minimizer. When the constant graphon $W \equiv q$ is the unique minimizer, $\mathcal{G}(n,q)$ represents the most likely model for the conditioned random graph (in terms of cut metric). However, it may be the case that $W \equiv q$ is a non-unique minimizer (which is provably not the case for $\mathsf{UT}_p(K_3,q)$, but I suspect that it is the case for the corresponding lower tail problem $\mathsf{LT}_p(K_3,q)$). When there are multiple distinct minimizers to the variational problem, all minimizers give rise to the same exponential rate, but one minimizer might still dominate by a lower-order $\exp(o(n^2))$ factor, which I do not know how to discern purely from the variational problem.

[3] The boundary curve for the upper tail phase diagram for K_3 is given by the equation $(1 + (q^{-1} - 1)^{1/(1-2q)})p = 1$.

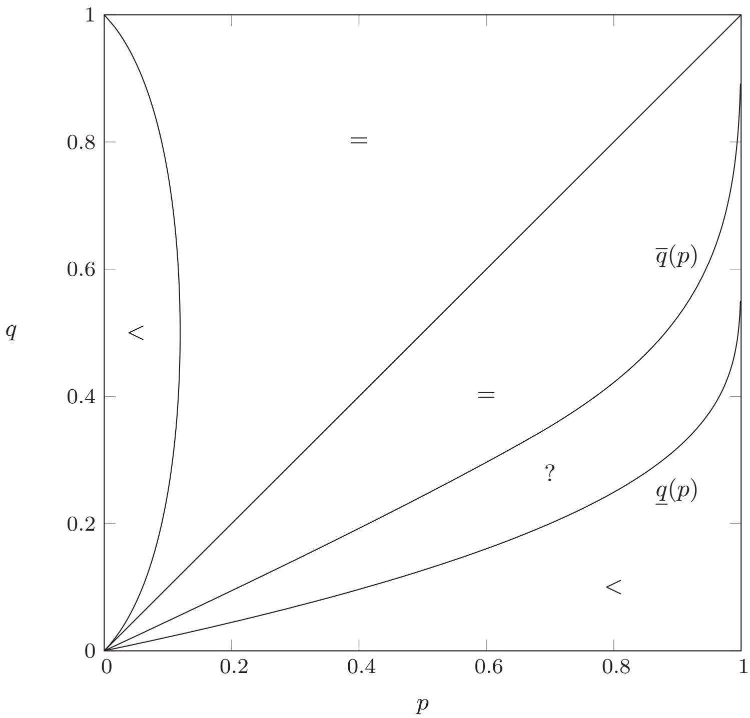


Figure 1. The phase diagram for triangle density upper tail variational problem $\mathsf{UT}_p(K_3, q)$ (when $q > p$) and lower tail variational problem $\mathsf{LT}_p(K_3, q)$ (when $q < p$). In regions marked '=', the constant graphon $W \equiv q$ is the unique minimizer to the variational problem. In regions marked '<', the constant graphon does not minimize the variational problem. The region marked '?' is unresolved. The results on lower tail are new. See Theorem 2.1.

In this paper we study the corresponding lower tail variational problem. For $0 < q \leqslant p < 1$, let

$$\mathsf{LT}_p(H, q) := \begin{cases} \text{minimize } \mathbb{E}[I_p(W)] \\ \text{subject to } t(H, W) \leqslant q^{e(H)}. \end{cases} \tag{1.3}$$

The connections between the large deviation problem and the variational problem discussed earlier hold for the lower tail just as they do for the upper tail. For example, as in (1.1), for fixed $0 \leqslant q \leqslant p \leqslant 1$, we have

$$\log \mathbb{P}(t(H, \mathcal{G}(n, p)) \leqslant q^{e(H)}) = -(1 + o(1)) \frac{n^2}{2} \mathsf{LT}_p(H, q). \tag{1.4}$$

As observed in [28], if H is a bipartite graph satisfying Sidorenko's conjecture [37], then $W \equiv q$ is the unique minimizer of $\mathsf{LT}_p(H, q)$. Recall that Sidorenko's conjecture asserts that for every bipartite graph H, we have $t(H, W) \geqslant \mathbb{E}[W]^{e(H)}$ for all graphons W. For any given H, if $t(H, W) \geqslant \mathbb{E}[W]^{e(H)}$ holds for all graphons W, then from the constraint $t(H, W) \leqslant q^{e(H)}$ of $\mathsf{LT}_p(H, q)$ we deduce $\mathbb{E}[W] \leqslant q$. Since $I_p(\cdot)$ is a convex function, from $\mathbb{E}[W] \leqslant q$ it follows that $W \equiv q$ is the unique minimizer of $\mathsf{LT}_p(H, q)$.

https://doi.org/10.1017/S0963548316000262 Published online by Cambridge University Press

Sidorenko's conjecture remains open,[4] though it has been proved for certain families of bipartite graphs H such as trees, cycles, hypercubes, and bipartite graphs containing one vertex adjacent to all vertices on the opposite side [11, 15, 22, 24, 38]. Even if Sidorenko's conjecture were false, it could still be true that $W \equiv q$ minimizes $\mathsf{LT}_p(H,q)$ for every bipartite H.

For the first non-bipartite case, namely K_3, new results in this paper partially characterize the lower tail phase diagram, as depicted in Figure 1. The region marked '?' remains unresolved. For other non-bipartite graph H, it is possible to draw similar partially identified phase diagrams using techniques in this paper. We will pay special attention to the slopes of the boundary curves at the origin.

The lower tail variational problem seems to be harder than the corresponding upper tail problem. By analogy, for the classical extremal graph theory problem of determining the range of possible triangle densities in a graph of fixed edge density, the maximization problem (analogous to the upper tail) follows as a corollary of the classic Kruskal–Katona theorem [19, 23],[5] whereas the corresponding minimization problem (analogous to the lower tail) was solved only relatively recently by Razborov [35] using his flag algebra machinery (also later solved for K_4 by Nikiforov [30] and all K_t by Reiher [36]). Furthermore, the qualitative nature of the phase transition seems to be different for the upper tail and the lower tail. It seems likely that the optimizing graphon W changes continuously as (p,q) crosses the upper tail phase boundary, but discontinuously for the lower tail.

The sparse setting, with $p = p_n \to 0$ and q/p kept constant, is more difficult. Using powerful new methods, Chatterjee and Dembo [8] showed that the large deviation problem in sparse random graphs also reduces to the natural variational problem,[6] provided that $p \geqslant n^{-\alpha_H}$ for some explicit $\alpha_H > 0$. A similar conclusion can be made about the lower tail variational problem using their techniques. With Lubetzky [29] we obtained the following asymptotic solution to the corresponding variational problem: for every fixed $\delta > 0$,

$$\lim_{p\to 0} \frac{\mathsf{UT}_p(K_3,(1+\delta)^{1/3}p)}{p^2\log(1/p)} = \min\left\{\delta^{2/3}, \frac{2}{3}\delta\right\}, \tag{1.5}$$

and as a corollary, as long as $p = p_n \to 0$ with $p_n \geqslant n^{-1/42}\log n$, we have

$$\mathbb{P}(t(K_3,\mathcal{G}(n,p)) \geqslant (1+\delta)p^3) = \exp\left(-(1-o(1))\min\left\{\frac{\delta^{2/3}}{2}, \frac{\delta}{3}\right\}n^2p^2\log\frac{1}{p}\right).$$

In this paper, we also study the lower tail variational problem as $p \to 0$. A nice feature of the lower tail problem in the sparse limit is that instead of being concerned with the

[4] The first unsettled case of Sidorenko's conjecture is for the graph H being $K_{5,5}$ with a Hamiltonian cycle removed (this H is sometimes called a 'Möbius strip'). There is some sentiment in the community that Sidorenko's conjecture may be false for this graph.

[5] The proof of the triangle upper tail result in [28] actually uses a strengthening of the Kruskal–Katona theorem, as we explain in Section 3.

[6] Some minor modifications need to made to the formulation variational problem $\mathsf{LT}_p(H,q)$ in order to match the statements in [8], namely that we only consider graphons that correspond to weighted graphs on n vertices. This difference is minor and does not affect the rest of this paper.

entire phase boundary curve, we can focus on its slope at the origin. In contrast, for the upper tail boundary curve, the slope at the origin is always 1.

The lower tail problem was also recently analysed by Janson and Warnke [18] using different methods (not relating to the variational problem). In the triangle case, for $n^{-1/2} \ll p \to 0$, they were able to determine the large deviation rate of $\mathbb{P}(t(K_3, \mathcal{G}(n,p)) \leqslant (1-\delta)p^3)$ for the two extremes $\delta = o(1)$ and $\delta = 1 - o(1)$. They left as an open question what happens for fixed $\delta \in (0,1)$, which is the subject of this paper.

There are other variants of the variational problem being studied in literature. For exponential random graphs, see [1, 3, 9, 21, 28, 34, 40, 41, 42, 43, 44, 45]. For the variational problem where several subgraph densities are simultaneously constrained (*e.g.* edge and triangle densities both fixed), see [2, 20, 31, 32, 33].

Section 2 contains statements of the results. Section 3 reviews the techniques used in proof of the upper tail results from [28]. Section 4 concerns the upper tail problem for triangle densities. Section 5 concerns general H-densities. The methods in Sections 4 and 5 are different since the techniques for triangles seem to be quantitatively superior but do not extend to all graphs. Section 6 concludes with some open problems.

2. Results

2.1. Triangle density

Here is our main result concerning the lower tail variational problem $\mathsf{LT}_p(K_3, q)$ for triangle densities. See Figure 1. The functions $\underline{q}$ and $\overline{q}$ below arise from the proof method as opposed to the true nature of the phase diagram. It is likely that there is a single curve separating the two phases.

Theorem 2.1. *There exist functions* $\underline{q}, \overline{q} : (0,1) \to (0,1)$ *satisfying* $0 < \underline{q}(p) \leqslant \overline{q}(p) \leqslant p$ *for* $0 < p < 1$ *with the following properties. Whenever* $\overline{q}(p) < q < p$*, the constant graphon* $W \equiv q$ *is the unique minimizer for* $\mathsf{LT}_p(K_3, q)$*. Whenever* $0 < q < \underline{q}(p)$*, the constant graphon* $W \equiv q$ *does not minimize* $\mathsf{LT}_p(K_3, q)$*. Further,* $\lim_{p\to 0} \underline{q}(p)/p = 0.209\ldots$ *while* $\lim_{p \to 0} \overline{q}(p)/p = 0.466\ldots$.

The two curves $\underline{q}(p)$ and $\overline{q}(p)$ are drawn in Figure 1. The nature of $\mathsf{LT}_p(K_3, q)$ remains unresolved for (p, q) between these two curves.[7]

In Theorem 2.1 and elsewhere, $0.466\ldots$ denotes the unique $0 < r < 1$ satisfying $\frac{3}{2} r \log r - r + 1 = 0$, and $0.209\ldots$ is defined as the maximum value of $r < 1$ such that that the function $f_r(x)$ in (4.9) (see also Figure 4) has a zero in the open interval $(0, r)$.

[7] The lower curve $\underline{q}(p)$ is drawn by only considering graphons on two equal steps (see Section 4.2). For larger values of p (*i.e.* closer to 1), the symmetry breaking region in Figure 1 can be enlarged by considering graphons with more than two steps, though we choose not to carry out this analysis here in order to focus on the small p regime.

https://doi.org/10.1017/S0963548316000262 Published online by Cambridge University Press

2.2. General subgraph density

We extend Theorem 2.1 to general subgraph counts. No serious effort is made here to optimize the quantitative bounds.

Theorem 2.2. *Let H be a graph. There exists $\overline{q}:(0,1)\to(0,1)$ with $\lim_{p\to 0}\overline{q}(p)/p<1$ such that whenever $\overline{q}(p)<q\leqslant p$, the constant graphon $W\equiv q$ is the unique minimizer for $\mathsf{LT}_p(H,q)$.*

Furthermore, if H is not bipartite, then there exists $\underline{q}:(0,1)\to(0,1)$ with $\lim_{p\to 0}\underline{q}(p)/p>0$ such that whenever $0\leqslant q<\underline{q}(p)$, the constant graphon $W\equiv q$ does not minimize $\mathsf{LT}_p(K_3,q)$.

The proof of the triangle case, Theorem 2.1, makes use of Goodman's inequality [14]:

$$t(K_3,W)+t(K_3,1-W)\geqslant 1/4.$$

If H satisfies $t(H,W)+t(H,1-W)\geqslant 2^{-e(H)+1}$ for all graphons W (such a graph H is sometimes called 'common' in the context of Ramsey multiplicities: see *e.g.* [12, Section 2.6]), then the same method can be used to establish regions where $W\equiv q$ is a minimizer of $\mathsf{LT}_p(H,q)$ (though the actual regions will not be the same as in Figure 1 due to other technical reasons). However, $t(H,W)+t(H,1-W)\geqslant 2^{-e(H)+1}$ does not hold in general. For example, Thomason [39] showed that K_t is a counterexample for all $t\geqslant 4$. Consequently, the proof method of Theorem 2.1 does not seem to extend to all H. Theorem 2.2 for general H is proved using a different method, which seems quantitatively inferior to the method for triangles.

For bipartite H, I conjecture that there is no phase transition.

Conjecture 2.3. *Let H be a bipartite graph. Then the constant function $W\equiv q$ is always the unique minimizer of $\mathsf{LT}_p(H,q)$.*

As mentioned in the Introduction, the conjecture holds for any H for which Sidorenko's conjecture is true, that is, $t(H,W)\geqslant\mathbb{E}[W]^{e(H)}$ for all graphons W. Conjecture 2.3 may be true even if Sidorenko's conjecture were not true.

2.3. Sparse limit

Since I_p is decreasing in $[0,p]$ and increasing in $[p,1]$, any minimizing W for $\mathsf{LT}_p(H,q)$ satisfies $0\leqslant W\leqslant p$ almost everywhere. Define

$$h(x):=x\log x-x+1 \tag{2.1}$$

so that

$$\lim_{p\to 0}p^{-1}I_p(px)=h(x)$$

uniformly for $x\in[0,1]$. It follows that for every graph H and $0\leqslant r\leqslant 1$ we have

$$\lim_{p\to 0}p^{-1}\mathsf{LT}_p(H,pr)=\mathsf{LT}(H,r) \tag{2.2}$$

https://doi.org/10.1017/S0963548316000262 Published online by Cambridge University Press

where

$$\mathsf{LT}(H,r) := \begin{cases} \text{minimize } \mathbb{E}[h(W)] \\ \text{subject to } t(H,W) \leqslant r^{e(H)}. \end{cases} \tag{2.3}$$

It would be interesting to solve this variational problem. As before, a basic question is whether the constant function $W \equiv r$ is a minimizer. Here is the main conjecture.

Conjecture 2.4. *Let H be a non-bipartite graph and $0 \leqslant r \leqslant 1$. There exists a $0 < r_H^* < 1$ so that $W \equiv r$ minimizes $\mathsf{LT}(H,r)$ if and only if $r \geqslant r_H^*$. Furthermore, W is the unique minimizer for $\mathsf{LT}(H,r)$ if and only if $r > r_H^*$.*

It seems likely that $r_{K_3}^* = 0.209\ldots$. The conjecture remains open for any non-bipartite graph H. For the bipartite case we make the following conjecture.

Conjecture 2.5. *The constant graphon $W \equiv r$ is the unique minimizer for $\mathsf{LT}(H,r)$ for every bipartite graph H and every $0 \leqslant r \leqslant 1$.*

In proving Theorem 2.1 and Theorem 2.2, we obtain the following results in the direction of the above conjectures.

Theorem 2.6. *If $0.466\cdots < r \leqslant 1$, the constant graphon $W \equiv r$ uniquely minimizes $\mathsf{LT}(K_3,r)$. If $0 \leqslant r < 0.209\ldots$, the constant graphon $W \equiv r$ does not minimize $\mathsf{LT}(K_3,r)$.*

Theorem 2.7. *Let H be a graph. There exists $\overline{r}_H < 1$ such that $W \equiv r$ uniquely minimizes $\mathsf{LT}(H,r)$ whenever $\overline{r}_H \leqslant r \leqslant 1$. If H is non-bipartite, then there exists $\underline{r}_H > 0$ such that $W \equiv r$ does not minimize $\mathsf{LT}(H,r)$ for $0 \leqslant r < \underline{r}_H$.*

Combining these results with the framework of Chatterjee and Dembo [8], we obtain

Corollary 2.8. *Let H be a graph. There is some explicit $\alpha_H > 0$ so that for $p = p_n \to 0$ with $p \geqslant n^{-\alpha}$, the following large deviation results hold.*

There exists $\overline{r}_H < 1$ so that for any fixed $r \in (\overline{r}_H, 1)$,

$$\lim_{n\to\infty} \frac{2}{n^2 p} \log \mathbb{P}(t(H, \mathcal{G}(n,p)) \leqslant (rp)^{e(H)}) = -h(r).$$

If H is non-bipartite, then there exists $\underline{r}_H > 0$ so that for any fixed $r \in (0, \underline{r}_H)$,

$$\liminf_{n\to\infty} \frac{2}{n^2 p} \log \mathbb{P}(t(H, \mathcal{G}(n,p)) \leqslant (rp)^{e(H)}) > -h(r).$$

For $H = K_3$, we may take $\overline{r}_{K_3} = 0.466\ldots$ and $\underline{r}_{K_3} = 0.209\ldots$.

3. Review of the proof for triangle upper tails

We begin with a quick review of the proof of the upper tail result from [28], as some of the ideas are used in the proof of Theorem 2.1.

The following extension of Hölder's inequality is very useful. See [28, Corollary 3.2].

Proposition 3.1. *Let H be a graph with maximum degree Δ. For any symmetric measurable function $W : [0,1]^2 \to \mathbb{R}$, we have $t(H,W) \leqslant \mathbb{E}[|W|^\Delta]^{e(H)/\Delta}$. In particular, $t(K_3,W) \leqslant \mathbb{E}[W^2]^{3/2}$.*

The inequality can be proved via repeated applications of Hölder's inequality (when $H = K_3$, the proof applies the Cauchy–Schwarz inequality three times). Observe that the inequality $t(K_3,W) \leqslant \mathbb{E}[W^2]^{3/2}$ strengthens a corollary of the Kruskal–Katona theorem on the maximum possible triangle density in a graph of given edge density: $t(K_3,W) \leqslant \mathbb{E}[W]^{3/2}$.

The following result from [28] gives the full replica symmetric phase for $\mathsf{UT}_p(K_3,q)$, the upper tail problem for triangle densities.

Theorem 3.2. *Let $0 < p \leqslant q < 1$. If the point $(q^2, I_p(q))$ lies on the convex minorant of the function $x \mapsto I_p(\sqrt{x})$, then $W \equiv q$ is the unique minimizer of $\mathsf{UT}_p(K_3,q)$.*

The upper tail boundary curve in Figure 1 is characterized by the condition in Theorem 3.2. See [28, Lemma 3.1] for the proof of symmetry breaking, that is, $\mathsf{UT}_p(K_3,q) < I_p(q)$, to the left of the boundary curve.

Proof. By the convex minorant condition, the tangent line to the function $I_p(\sqrt{x})$ at $x = q^2$ lies below the function, so that

$$I_p(\sqrt{x}) \geqslant I_p(q) + \frac{I_p'(q)}{2q}(x - q^2), \quad \text{for all } x \in [0,1].$$

Replacing x by x^2, we get

$$I_p(x) \geqslant I_p(q) + \frac{I_p'(q)}{2q}(x^2 - q^2), \quad \text{for all } x \in [0,1]. \tag{3.1}$$

Note that $I_p'(q) > 0$ since $q > p$.

Suppose graphon W satisfies $t(K_3,W) \geqslant q^3$. By Proposition 3.1, we have

$$\mathbb{E}[W^2] \geqslant t(K_3,W)^{3/2} \geqslant q^2.$$

Thus (3.1) implies that

$$\mathbb{E}[I_p(W)] \geqslant I_p(q) + \frac{I_p'(q)}{2q}(\mathbb{E}[W^2] - q^2) \geqslant I_p(q).$$

This shows that $W \equiv q$ is a minimizer for $\mathsf{UT}_p(K_3,q)$, and furthermore it is not too hard to check equality conditions to verify that this is the unique minimizer. □

4. Triangle lower tails

In this section we prove Theorems 2.1 and 2.6.

4.1. Replica symmetry phase

We begin with a small modification of Goodman's theorem [14] (which is usually generally stated for $U + W \equiv 1$).

Lemma 4.1. *If U and W are graphon such that $U + W \geqslant 2q$ for some constant $q \geqslant 0$, then*

$$t(K_3, W) + t(K_3, U) \geqslant 2q^3.$$

Proof. By decreasing U and W (while remaining non-negative), we may assume that they are graphons satisfying $U + W \equiv 2q$. Let $U = q + X$ and $W = q - X$ for some symmetric measurable function $X : [0,1]^2 \to \mathbb{R}$. Then

$$\begin{aligned} t(K_3, W) + t(K_3, U) &= t(K_3, q + X) + t(K_3, q - X) \\ &= 2q^3 + 6q\, t(K_{1,2}, X) \\ &\geqslant 2q^3 + 6q(\mathbb{E}[X])^2 \geqslant 2q^3. \end{aligned}$$ □

For any $a \in \mathbb{R}$ we write $a_+ := \max\{a, 0\}$. In the proposition below, a_+^2 means $(a_+)^2$. The inequality (4.1) below is motivated by considering the tangent line to $x \mapsto I_p(2q - \sqrt{x})$ at $x = q^2$, as in the proof of Theorem 3.2.

Proposition 4.2. *Let $0 < q \leqslant p < 1$ be such that*

$$I_p(x) \geqslant I_p(q) + \frac{-I_p'(q)}{2q}((2q - x)_+^2 - q^2) \quad \text{for all } x \in [0, p]. \tag{4.1}$$

Then $W \equiv q$ is the unique minimizer of $\mathsf{LT}_p(K_3, q)$.

Proof. Suppose W satisfies $t(K_3, W) \leqslant q^3$. Apply Lemma 4.1 to W and $U := (2q - W)_+$ to obtain

$$t(K_3, (2q - W)_+) \geqslant 2q^3 - t(K_3, W) \geqslant q^3.$$

Next, apply Proposition 3.1 and we obtain

$$\mathbb{E}[(2q - W)_+^2] \geqslant t(K_3, (2q - W)_+)^{2/3} \geqslant q^2.$$

By (4.1) we have (note that $I_p'(q) \leqslant 0$ as $q \leqslant p$)

$$\mathbb{E}[I_p(W)] \geqslant I_p(q) + \frac{-I_p'(q)}{2q}(\mathbb{E}[(2q - W)_+^2] - q^2) \geqslant I_p(q).$$

It follows that $\mathsf{LT}_p(K_3, q) = I_p(q)$. To show that $W \equiv q$ is the unique minimizer, observe that in order for any other W to be a minimizer, equality must occur at every step above. In particular, if (4.1) has single point of equality, namely for $x = q$, then the uniqueness of W is clear. Otherwise, one can check (details omitted, but see Figure 2) that that (4.1) has at most two points of equality, with one being $x = q$, so that if W has any positive mass

https://doi.org/10.1017/S0963548316000262 Published online by Cambridge University Press

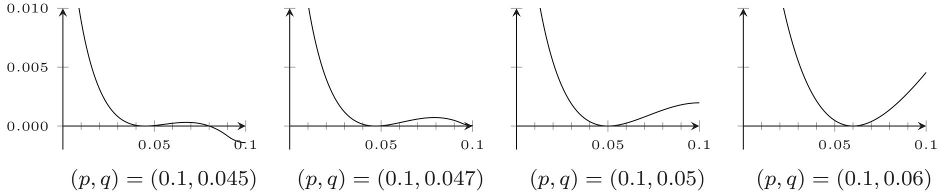


Figure 2. Plots of $f_{p,q}$ from (4.3) for $p = 0.1$ and various values of q.

with value being the other point of equality, then it would be impossible for $t(K_3, W) = q^3$ to hold. This shows that $W \equiv q$ is the unique minimizer. □

Using Lemma 4.3 below we obtain the $0 < p \leqslant 1/2$ portion of the curve $\overline{q}$ of Theorem 2.1, which is given implicitly by

$$I_p(q) + \frac{1}{2} q I_p'(q) = 0, \tag{4.2}$$

and shown in Figure 1. The rest of the curve (*i.e.* for $1/2 < p < 1$) in Figure 1 is produced by numerically checking (3.1). Taking the $p \to 0$ limit of (4.2), we see that the slope at the origin is equal to $\overline{r} = 0.466\ldots$, where $\overline{r}$ satisfies $h(\overline{r}) + \frac{1}{2}\overline{r}h'(\overline{r}) = 0$. This completes the proof of the replica symmetric phase in Theorem 2.1.

Lemma 4.3. *For* $0 < q \leqslant p \leqslant 1/2$, (4.1) *holds for all* $x \in [0, p]$ *if and only if it holds at* $x = p$.

Proof. Let

$$f(x) := f_{p,q} := I_p(x) - I_p(q) + \frac{I_p'(q)}{2q}((2q - x)_+^2 - q^2). \tag{4.3}$$

We plotted f for some representative values of (p, q) in Figure 2.

Suppose $f(p) \geqslant 0$. We have

$$f'(x) = I_p'(x) - \frac{I_p'(q)}{q}(2q - x)_+$$

and

$$f''(x) = I_p''(x) + \frac{I_p'(q)}{q} 1_{x<2q} = \frac{1}{x(1-x)} + \frac{I_p'(q)}{q} 1_{x<2q}.$$

Since $p \leqslant 1/2$, $f''(x)$ is decreasing for $0 < x < \min\{p, 2q\}$. Clearly f'' is positive near $x = 0$. We consider two cases.

Case I. $f''(x) > 0$ for all $0 < x < \min\{p, 2q\}$. Then f is convex on $(0, \min\{p, 2q\})$. We know that $f(q) = f'(q) = 0$. So $f(x) \geqslant 0$ for all $x \in [0, \min\{p, 2q\}]$. If $2q \geqslant p$, then we are done. Otherwise, note that

$$f(x) = I_p(x) - I_p(q) - qI_p'(q)/2 \quad \text{for } x \in [2q, p],$$

https://doi.org/10.1017/S0963548316000262 Published online by Cambridge University Press

and it is decreasing on this interval. Since we assumed that $f(p) \geqslant 0$, we obtain $f(x) \geqslant 0$ for all $x \in [0, p]$.

Case II. There is some $x_0 \in (0, \min\{p, 2q\})$ such that $f''(x_0) = 0$. So f is convex on $(0, x_0)$ and concave on $(x_0, \min\{p, 2q\})$. We assumed that $f(p) \geqslant 0$, so $f(\min\{p, 2q\}) \geqslant 0$ since if $2q < p$ then f is decreasing on $(2q, p)$. Since $f(q) = f'(q) = 0$, an analysis of the convexity of f shows that it is non-negative on $[0, p]$. □

For the sparse limit $p \to 0$, the proof of the first half of Theorem 2.6 is nearly identical. It follows from the next two propositions, whose proofs we omit.

Proposition 4.4. *Let $0 \leqslant r \leqslant 1$ be such that*

$$h(x) \geqslant h(r) + \frac{-h'(r)}{2r}((2r - x)_+^2 - r^2), \quad \text{for all } x \in [0, 1]. \tag{4.4}$$

Then $W \equiv r$ is the unique minimizer of $\mathsf{LT}(K_3, r)$.

Lemma 4.5. *The inequality* (4.4) *holds for all $x \in [0, 1]$ if and only if holds for $x = 1$, which holds if and only if $r \geqslant \bar{r} = 0.466\ldots$, where $\bar{r}$ satisfies $h(\bar{r}) + \frac{1}{2}\bar{r}h'(\bar{r}) = 0$.*

4.2. Symmetry breaking phase

Now we explain the lower curve $\underline{q}$ in Figure 1. It is obtained by by restricting the variational problem $\mathsf{LT}_p(K_3, q)$ to graphons W of the form $\mathrm{BIP}_{a,b}$, where $\mathrm{BIP}_{a,b}$, for $0 \leqslant a, b \leqslant 1$, is defined by

$$\mathrm{BIP}_{a,b}(x, y) := \begin{cases} a & \text{if } (x, y) \in [0, 1/2]^2 \cup (1/2, 1]^2, \\ b & \text{if } (x, y) \in [0, 1/2] \times (1/2, 1] \cup (1/2, 1] \times [0, 1/2]. \end{cases} \tag{4.5}$$

There is symmetry breaking if we can find $0 \leqslant a, b \leqslant p$ satisfying

$$\mathbb{E}[I_p(\mathrm{BIP}_{a,b})] = \frac{1}{2}I_p(a) + \frac{1}{2}I_p(b) < I_p(q) \tag{4.6}$$

and

$$t(K_3, \mathrm{BIP}_{a,b}) = \frac{1}{4}a^3 + \frac{3}{4}ab^2 \leqslant q^3. \tag{4.7}$$

We can assume that $0 \leqslant a \leqslant q \leqslant b \leqslant p$, since otherwise swapping a and b reduces $t(K_3, W)$ (observe that $t(K_3, \mathrm{BIP}_{a,b}) - t(K_3, \mathrm{BIP}_{b,a}) = \frac{1}{4}(a - b)^3$) while keeping $\mathbb{E}[I_p(W)]$ constant.

Set

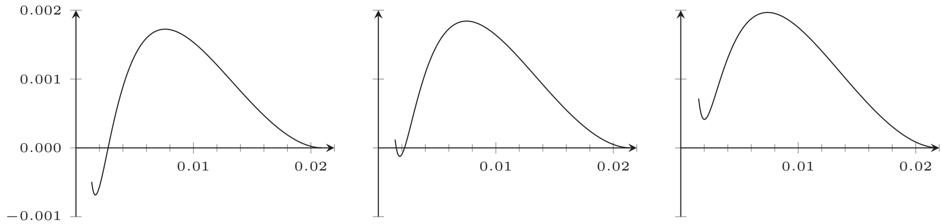
$$b = \sqrt{(4q^3 - a^3)/(3a)}$$

so that $t(K_3, \mathrm{BIP}_{a,b}) = q^3$. There is symmetry breaking if

$$f(x) := f_{p,q}(x) := \frac{1}{2}I_p(x) + \frac{1}{2}I_p\left(\sqrt{\frac{4q^3 - x^3}{3x}}\right) - I_p(q) \tag{4.8}$$

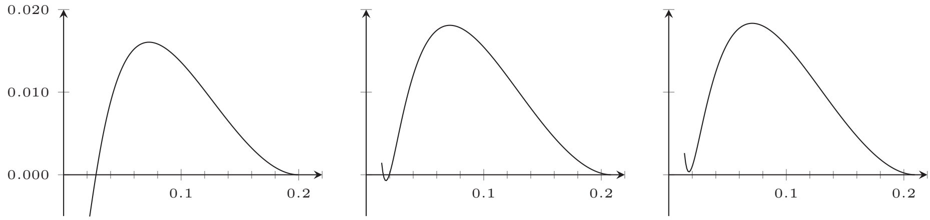
is negative for some $0 \leqslant x \leqslant q$, where f is only defined for

$$\sqrt{(4q^3 - x^3)/(3x)} \leqslant p.$$



$(p, q) = (0.1, 0.021)$ $(p, q) = (0.1, 0.0215)$ $(p, q) = (0.1, 0.022)$

Figure 3. The plot of $f_{p,q}$ from (4.8) for $p = 0.1$ and various values of q.



$r = 0.2$ $r = 0.209$ $r = 0.21$

Figure 4. The plot of f_r from (4.9) for various values of r.

Some representative examples of f are plotted in Figure 3. For every p, and sufficiently small q, $f(x)$ becomes negative in a region away from $x = q$.

Now we prove the claims in Theorem 2.1 more rigorously. For every $p > 0$, if q is sufficiently small that $\frac{1}{2}I_p(0) < I_p(q)$, then $W = \mathrm{BIP}_{0,p}$ satisfies $t(K_3, W) = 0$ while

$$\mathbb{E}[I_p(W)] = \frac{1}{2}I_p(0) < I_p(q),$$

so that $\mathsf{LT}_p(K_3, q) < I_p(q)$.

The argument in the previous paragraph does not give the optimal $\underline{q}$ in Theorem 2.1. To prove that $\underline{q}$ can be chosen so that $\lim_{p\to 0} \underline{q}(p)/p = 0.209\ldots$, it suffices, by (2.2), to prove the second half of Theorem 2.6, that $\mathsf{LT}(K_3, r) < h(r)$ for all $r < r_1 = 0.209\ldots$. As before, we seek $0 \leqslant a \leqslant r \leqslant b \leqslant 1$ with

$$\frac{1}{2}h(a) + \frac{1}{2}h(b) < h(r)$$

and

$$\frac{1}{4}a^3 + \frac{3}{4}ab^2 \leqslant r^3.$$

Let

$$f(x) := f_r(x) := \frac{1}{2}h(x) + \frac{1}{2}h\left(\sqrt{\frac{4r^3 - x^3}{3x}}\right) - h(r). \tag{4.9}$$

See Figure 4 for some examples of plots of f_r (as before, plotted for values of $x \leqslant r$ satisfying $\sqrt{(4r^3 - x^3)/(3x)} \leqslant 1$). At the critical $r = r_1 = 0.209\ldots$, there exists

https://doi.org/10.1017/S0963548316000262 Published online by Cambridge University Press

$0 < a_1 < r_1 < b_1 < 1$ such that

$$\frac{1}{4}a_1^3 + \frac{3}{4}a_1 b_1^2 = r_1^3 \quad \text{and} \quad \frac{1}{2}h(a_1) + \frac{1}{2}h(b_1) = h(r_1).$$

Now for any $0 \leqslant r < r_1$, let $s = r/r_1$, so that $(a,b) = (sa_1, sb_1)$ satisfies

$$\frac{1}{4}a^3 + \frac{3}{4}ab^2 = r^3.$$

Note that

$$\begin{aligned} h(sx) &= sx\log(sx) - sx + 1 \\ &= s(x\log x - x + 1) + (s\log s)x - s + 1 \\ &= sh(x) + (s\log s)x - s + 1. \end{aligned}$$

We have

$$\begin{aligned} \frac{1}{2}h(a) + \frac{1}{2}h(b) - h(r) &= \frac{1}{2}h(sa_1) + \frac{1}{2}h(sb_1) - h(sr_1) \\ &= s\left(\frac{1}{2}h(a_1) + \frac{1}{2}h(b_1) - h(r_1)\right) + (s\log s)\left(\frac{1}{2}a_1 + \frac{1}{2}b_1 - r_1\right) \\ &< 0 \end{aligned}$$

since

$$\frac{1}{2}h(a_1) + \frac{1}{2}h(b_1) = h(c_1),$$

and we know that

$$\frac{1}{2}a_1 + \frac{1}{2}b_1 > r_1$$

from

$$\begin{aligned} \left(\frac{1}{2}a_1 + \frac{1}{2}b_1\right)^3 - r_1^3 &= \left(\frac{1}{2}a_1 + \frac{1}{2}b_1\right)^3 - \frac{1}{4}a_1^3 - \frac{3}{4}a_1 b_1^2 \\ &= \left(\frac{1}{2}b_1 - \frac{1}{2}a_1\right)^3 > 0. \end{aligned}$$

It follows that $\mathsf{LT}(K_3, r) < I_p(r)$ for all $0 < r < r_1 = 0.209\ldots$.

5. General subgraph lower tails

In this section we prove Theorems 2.2 and 2.7. I will give the details only for Theorem 2.7 concerning the sparse limit $\mathsf{LT}(H,r)$ as it is somewhat cleaner and contains all the ideas. Theorem 2.2 regarding $\mathsf{LT}_p(H,q)$ can be proved analogously by considering sufficiently small but fixed values of p.

5.1. Replica symmetry

For any graph H and graphon W, we define the functional derivative $t'(H,W)$ to be the symmetric measurable function given by

$$t'(H,W) = \sum_{e\in E(H)} t_e(H,W), \tag{5.1}$$

where for each $ab \in E(H)$ we define the graphon

$$t_{ab}(H,W)(x_a,x_b) := \int_{[0,1]^{V(H)\setminus\{a,b\}}} \prod_{ij\in E(H)\setminus\{ab\}} W(x_i,x_j) \prod_{i\in V(H)\setminus\{a,b\}} dx_i. \tag{5.2}$$

For example,

$$t'(K_3,W)(x,y) = 3\int_{[0,1]} W(x,z)W(y,z)\,dz.$$

For any symmetric measurable $U : [0,1]^2 \to [-1,1]$, and $\delta \to 0$, we have

$$t(H, W+\delta U) = t(H,W) + \delta\,\mathbb{E}[t'(H,W)U] + O(\delta^2),$$

which justifies calling $t'(H,W)$ the functional derivative.

Lemma 5.1. *Let H be a graph and $0 < r < 1$. The variational problem $\mathsf{LT}(H,r)$ attains its minimum for some graphon W, and any such W satisfies the following Lagrange multiplier condition: for some $\lambda \geqslant 0$, we have*

$$h'(W(x,y)) + \lambda t'(H,W)(x,y) = 0, \quad a.e.\text{-}(x,y) \in [0,1]^2.$$

Proof. That the minimum of $\mathsf{LT}(H,r)$ is always attained follows from the compactness of the space of graphons with respect to the cut distance and the convexity of h, as was already observed in [10].[8]

Suppose W minimizes $\mathsf{LT}(H,r)$. To prove the lemma, it suffices to prove the following claim: for any symmetric measurable function $U : [0,1]^2 \to [-1,1]$ such that

$$0 \leqslant W + U \leqslant 1 \quad \text{and} \quad \mathbb{E}[t'(H,W)U] < 0,$$

we have $\mathbb{E}[h'(W)U] \geqslant 0$.

[8] We sketch here an alternative proof that the minimum is always attained. Let W_n be a sequence of graphons with $t(H,W_n) \geqslant r^{e(H)}$ and $\mathbb{E}[h(W_n)] \to \mathsf{LT}(H,r)$. By compactness of the space of graphons [27], there exists a subsequential limit W so that $\delta_\square(W_n,W) \to 0$ along some subsequence. Restrict to this subsequence. We have $t(H,W_n) \to t(H,W)$, so that $t(H,W) \geqslant r^{e(H)}$. It remains to show that $\mathbb{E}[h(W)] \leqslant \lim \mathbb{E}[h(W_n)] = \mathsf{LT}(H,r)$. We do not lose anything by assuming that $\|W_n - W\|_\square \to 0$. Let $\mathcal{P}_m$ denote the partition of the unit interval $[0,1]$ into m equal-length intervals. Let $W_{\mathcal{P}_m}$ denote W with its value inside each $I_i \times I_j$ replaced by its average, for every $I_i, I_j \in \mathcal{P}_m$. Define $(W_n)_{\mathcal{P}_m}$ similarly. Then $\|W_n - W\|_\square \to 0$ implies that $(W_n)_{\mathcal{P}_m} \to W_{\mathcal{P}_m}$ pointwise a.e. as $n \to \infty$. By convexity, we have

$$\lim_{n\to\infty} \mathbb{E}[h(W_n)] \geqslant \liminf_{n\to\infty} \mathbb{E}[h((W_n)_{\mathcal{P}_m})] = \mathbb{E}[h(W_{\mathcal{P}_m})].$$

Furthermore, $W_{\mathcal{P}_m} \to W$ pointwise a.e. by the Lebesgue density theorem, so $\lim_{m\to\infty} \mathbb{E}[h(W_{\mathcal{P}_m})] = \mathbb{E}[h(W)]$. It follows that $\mathbb{E}[h(W)] \leqslant \lim \mathbb{E}[h(W_n)] = \mathsf{LT}(H,r)$, as desired.

https://doi.org/10.1017/S0963548316000262 Published online by Cambridge University Press

Consider the graphon $W+\delta U$ for $\delta \searrow 0$. We have

$$t(H, W+\delta U) - t(H,W) = \delta \mathbb{E}[t'(H,W)U] + O(\delta^2).$$

Therefore $t(H, W+\delta U) < t(H,W) \leqslant r^{e(H)}$ for sufficiently small $\delta > 0$, and hence $\mathbb{E}[h(W+\delta U)] \geqslant \mathbb{E}[h(W)]$ since W minimizes $\mathsf{LT}(H,r)$. On the other hand,

$$\lim_{\delta\to 0} \frac{\mathbb{E}[h(W+\delta U) - h(W)]}{\delta} = \mathbb{E}[h'(W)U],$$

so that $\mathbb{E}[h'(W)U] \geqslant 0$ as claimed. The interchange of limit and expectation above can be justified by writing $U = U_+ - U_-$, where $U_+ = \max\{U, 0\}$ and $U_- = \max\{-U, 0\}$. Since h is convex, $(h(W+\delta U_+) - h(W))/\delta$ is pointwise monotonically decreasing as $\delta \searrow 0$, and likewise $(h(W-\delta U_-) - h(W))/\delta$ is pointwise monotonically increasing. So the interchange of limit and expectation is justified by the monotone convergence theorem. □

Lemma 5.2. *Let H be a graph with m edges, and $0 < r \leqslant 1$. Let W minimize $\mathsf{LT}(H,r)$. Then $W \geqslant r^{mr^{-m}}$ almost everywhere.*

Proof. Let $c = r^{mr^{-m}}$. Suppose on the contrary that $W < c$ on a set of positive measure. Let λ be the Lagrange multiplier in Lemma 5.1. From (5.1) we have $t'(H,W) \leqslant m$ everywhere. By considering a positive-measure set of (x,y) such that $W(x,y) < c$, we find

$$0 = h'(W(x,y)) + \lambda t'(H,W)(x,y) < h'(c) + m\lambda.$$

So that

$$\lambda > \frac{-h'(c)}{m} = \frac{\log(1/c)}{m}.$$

Therefore, up to a set of measure zero, for every (x,y) with $W(x,y) \geqslant r^m$, we have

$$t'(H,W)(x,y) = \frac{-h'(W(x,y))}{\lambda} < \frac{-mh'(r^m)}{\log(1/c)} = \frac{m\log(r^{-m})}{\log(r^{-mr^{-m}})} = mr^m.$$

On the other hand, for every (x,y) with $W(x,y) < r^m$, we have $t'(H,W)(x,y)W(x,y) < mr^m$. Thus $t'(H,W)W < mr^m$ almost everywhere. By (5.1) and (5.2), we have

$$t(H,W) = \frac{1}{m}\mathbb{E}[t'(H,W)W] < r^m.$$

However, any W with $t(H,W) < r^m$ cannot minimize $\mathsf{LT}(H,r)$. This gives the desired contradiction. □

Lemma 5.3. *If $t(H,W) \leqslant r^{e(H)}$, then $\mathbb{E}[\log W] \leqslant \log r$.*

Proof. The lemma follows from Jensen's inequality:

$$\begin{aligned} m\,\mathbb{E}[\log W] &= \int_{[0,1]^{V(H)}} \log\Bigg(\prod_{ij\in E(H)} W(x_i,x_j)\Bigg) \prod_{i\in V(H)} dx_i \\ &\leqslant \log\Bigg(\int_{[0,1]^{V(H)}} \prod_{ij\in E(H)} W(x_i,x_j) \prod_{i\in V(H)} dx_i\Bigg) = \log t(H,W) \leqslant m\log r. \end{aligned}$$ □

https://doi.org/10.1017/S0963548316000262 Published online by Cambridge University Press

Table 1. Some values of r_m from Proposition 5.4.

m	3	4	5	6	7	8	9	10	20	100
r_m	0.686	0.735	0.770	0.795	0.815	0.831	0.844	0.855	0.911	0.973

Proposition 5.4. *Let H be a graph with $m \geqslant 3$ edges. Let $r = r_m$ (see Table 1) be the unique solution in the interval $(0,1)$ to the equation*

$$h(r^{mr^{-m}}) = h(r) + r'h(r)(\log(r^{mr^{-m}}) - \log r).$$

Then $\mathsf{LT}(H,r)$ is uniquely minimized by the constant function $W \equiv r$ for all $r \geqslant r_m$.

Note that any graph H with at most two edges always satisfies $t(H,W) \geqslant (\mathbb{E}W)^{e(H)}$, so it follows (by the argument in the paragraph following (1.4)) that $W \equiv r$ is the unique minimizer. Thus it suffices to consider $m \geqslant 3$.

Proof. Let $r \geqslant r_m$. Let W be a minimizer for $\mathsf{LT}(H,r)$. By Lemma 5.2, $W \geqslant r^{mr^{-m}}$ almost everywhere. Thus it follows by Lemma 5.5 below (and it can be checked that $r_m \geqslant 1/e$) that

$$h(W) \geqslant h(r) + r\log r(\log W - \log r) \quad \text{a.e.} \tag{5.3}$$

Taking expectation of both sides and using $\mathbb{E}[\log W] \leqslant \log r$ from Lemma 5.3 (note that $\log r \leqslant 0$), we obtain $\mathbb{E}[h(W)] \geqslant h(r)$, as desired. To see that $W \equiv r$ is unique, suppose W is another minimizer of $\mathsf{LT}(H,r)$. Equality must hold everywhere in the argument. In particular, (5.3) must hold almost everywhere, which easily implies that $W \equiv r$ (for the critical case $r = r_m$, W might also take the value $r^{mr^{-m}}$, but only on a set of measure zero since $\mathbb{E}[h(W)] = h(r)$). □

Lemma 5.5. *If*

$$h(x) \geqslant h(r) + rh'(r)(\log x - \log r) \tag{5.4}$$

holds for some $(x,r) = (x_0, r_0)$, with $0 \leqslant x_0 \leqslant r_0 \leqslant 1$ and $r_0 \in [1/e, 1]$, then it holds for all $(x,r) \in [x_0, 1] \times [r_0, 1]$.

Proof. The partial derivative of the right-hand side of (5.4) with respect to r is

$$-(1 + \log r)(\log r - \log x),$$

which is at most zero as long as $x \leqslant r$ and $r \geqslant 1/e$. This shows that if (5.4) holds for some $(x,r) = (x_0, r_0)$ then it automatically holds for $(x,r) = (x_0, r)$ for all $r \in [r_0, 1]$.

Let us now fix r. Let

$$f(x) := f_r(x) := h(x) - h(r) - rh'(r)(\log x - \log r). \tag{5.5}$$

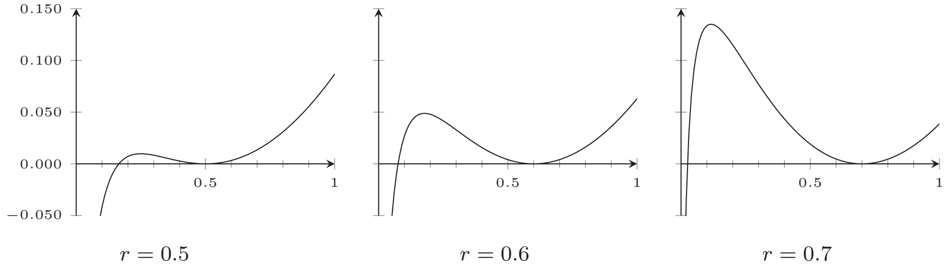


Figure 5. Plot of f_r from (5.5) for various values of r.

Some examples of f_r are plotted in Figure 5. We have

$$f'(x) = \log x - \frac{r \log r}{x} \quad \text{and} \quad f''(x) = \frac{x + r \log r}{x^2}.$$

So $f''(x) < 0$ for $x < -r \log r$ and $f''(x) > 0$ for $x > -r \log r$. Note also that $f(r) = f'(r) = 0$, and $-r \log r \leqslant r$ as long as $r \geqslant 1/e$. By analysing the convexity of f (see Figure 5), we see that $f(x_0) \geqslant 0$ implies $f(x) \geqslant 0$ for all $x \in [x_0, 1]$. □

5.2. Symmetry breaking

The proof of the second part of Theorem 2.7 is easy. One could fine-tune the bounds as in Section 4.2, though we omit the analysis here.

Proposition 5.6. *Let H be a non-bipartite graph. Then* $\mathsf{LT}(H,r) < h(r)$ *for all* $r < 0.186$.

Proof. The graphon $W = \mathrm{BIP}_{0,1}$ satisfies $t(H,W) = 0$, and $\mathbb{E}[h(W)] = \frac{1}{2}h(0)$, which is strictly less than $h(r)$ for all $r < 0.186$. □

6. Open problems

We conclude with some open problems concerning the variational problem for upper and lower tails.

- *Upper tail phase diagram.* Determine the upper tail replica symmetry phase diagram for non-regular H.
- *Lower tail phase diagram.* Determine the lower tail replica symmetry phase diagram for K_3, and more generally for any non-bipartite graph H. In particular, determine r_H^* from Conjecture 2.4. For a bipartite graph H, determine whether there is replica symmetry everywhere (Conjecture 2.3).
- *Solution in the symmetry breaking phase.* Solve the variational problem UT or LT at any non-trivial point where the constant graphon is not a minimizer.

Acknowledgements

I thank Eyal Lubetzky for helpful discussions.

References

[1] Aristoff, D. and Radin, C. (2013) Emergent structures in large networks. *J. Appl. Probab.* **50** 883–888.

[2] Aristoff, D. and Zhu, L. (2015) Asymptotic structure and singularities in constrained directed graphs. *Stochastic Process. Appl.* **125** 4154–4177.

[3] Aristoff, D. and Zhu, L. On the phase transition curve in a directed exponential random graph model. `arXiv:1404.6514`

[4] Borgs, C., Chayes, J. and Lovász, L. (2010) Moments of two-variable functions and the uniqueness of graph limits. *Geom. Funct. Anal.* **19** 1597–1619.

[5] Borgs, C., Chayes, J. T., Lovász, L., Sós, V. T. and Vesztergombi, K. (2008) Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing. *Adv. Math.* **219** 1801–1851.

[6] Borgs, C., Chayes, J. T., Lovász, L., Sós, V. T. and Vesztergombi, K. (2012) Convergent sequences of dense graphs II: Multiway cuts and statistical physics. *Ann. of Math.* (2) **176** 151–219.

[7] Chatterjee, S. (2012) The missing log in large deviations for triangle counts. *Random Struct. Alg.* **40** 437–451.

[8] Chatterjee, S. and Dembo, A. (2016) Nonlinear large deviations. *Adv. Math.* **299** 396–450.

[9] Chatterjee, S. and Diaconis, P. (2013) Estimating and understanding exponential random graph models. *Ann. Statist.* **41** 2428–2461.

[10] Chatterjee, S. and Varadhan, S. R. S. (2011) The large deviation principle for the Erdős–Rényi random graph. *European J. Combin.* **32** 1000–1017.

[11] Conlon, D., Fox, J. and Sudakov, B. (2010) An approximate version of Sidorenko's conjecture. *Geom. Funct. Anal.* **20** 1354–1366.

[12] Conlon, D., Fox, J. and Sudakov, B. (2015) Recent developments in graph Ramsey theory. In *Surveys in Combinatorics*, pp. 49-118.

[13] DeMarco, B. and Kahn, J. (2012) Upper tails for triangles. *Random Struct. Alg.* **40** 452–459.

[14] Goodman, A. W. (1959 On sets of acquaintances and strangers at any party. *Amer. Math. Monthly* **66** 778–783.

[15] Hatami, H. (2010) Graph norms and Sidorenko's conjecture. *Israel J. Math.* **175** 125–150.

[16] Janson, S. (1990) Poisson approximation for large deviations. *Random Struct. Alg.* **1** 221–229.

[17] Janson, S., Łuczak, T. and Ruciński, A. (2000) *Random Graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience.

[18] Janson, S. and Warnke, L. (2016) The lower tail: Poisson approximation revisited. *Random Struct. Alg.* **48** 219–246.

[19] Katona, G. (1968) A theorem of finite sets. In *Theory of Graphs: Proc. Colloq. Tihany 1966*, Academic Press, pp. 187–207.

[20] Kenyon, R., Radin, C., Ren, K. and Sadun, L. Multipodal structure and phase transitions in large constrained graphs. `arXiv:1405.0599`

[21] Kenyon, R. and Yin, M. On the asymptotics of constrained exponential random graphs. `arXiv:1406.3662`

[22] Kim, J., Lee, C. and Lee, J. (2016) Two approaches to Sidorenko's conjecture. *Trans. Amer. Math. Soc.* **368** 5057–5074.

[23] Kruskal, J. B. (1963) The number of simplices in a complex. In *Mathematical Optimization Techniques* (R. Bellman, ed.), University of California Press, pp. 251–278.

[24] Li, J. L. X. and Szegedy, B. On the logarithmic calculus and Sidorenko's conjecture. *Combinatorica*, to appear.

[25] Lovász, L. (2012) *Large Networks and Graph Limits*, Vol. 60 of *American Mathematical Society Colloquium Publications*, AMS.

[26] Lovász, L. and Szegedy, B. (2006) Limits of dense graph sequences. *J. Combin. Theory Ser. B* **96** 933–957.

[27] Lovász, L. and Szegedy, B. (2007) Szemerédi's lemma for the analyst. *Geom. Funct. Anal.* **17** 252–270.
[28] Lubetzky, E. and Zhao, Y. (2015) On replica symmetry of large deviations in random graphs. *Random Struct. Alg.* **47** 109–146.
[29] Lubetzky, E. and Zhao, Y. On the variational problem for upper tails in sparse random graphs. *Random Struct. Alg.*, to appear
[30] Nikiforov, V. (2011) The number of cliques in graphs of given order and size. *Trans. Amer. Math. Soc.* **363** 1599–1618.
[31] Radin, C., Ren, K. and Sadun, L. (2014) The asymptotics of large constrained graphs. *J. Phys. A* **47** 175001.
[32] Radin, C. and Sadun, L. (2013) Phase transitions in a complex network. *J. Phys. A* **46** 305002.
[33] Radin, C. and Sadun, L. (2015) Singularities in the entropy of asymptotically large simple graphs. *J. Statist. Phys.* **158** 853–865.
[34] Radin, C. and Yin, M. (2013) Phase transitions in exponential random graphs. *Ann. Appl. Probab.* **23** 2458–2471.
[35] Razborov, A. A. (2008) On the minimal density of triangles in graphs. *Combin. Probab. Comput.* **17** 603–618.
[36] Reiher, C. The clique density theorem. *Ann. Math.*, to appear
[37] Sidorenko, A. (1993) A correlation inequality for bipartite graphs. *Graphs Combin.* **9** 201–204.
[38] Szegedy, B. An information theoretic approach to Sidorenko's conjecture. `arXiv:1406.6738`
[39] Thomason, A. (1989) A disproof of a conjecture of Erdős in Ramsey theory. *J. London Math. Soc.* (2) **39** 246–255.
[40] Yin, M. (2013) Critical phenomena in exponential random graphs. *J. Statist. Phys.* **153** 1008–1021.
[41] Yin, M. (2015) Large deviations and exact asymptotics for constrained exponential random graphs. *Electron. Commun. Probab.* **20**, 14 pp.
[42] Yin, M., Rinaldo, A. and Fadnavis, S. Asymptotic quantization of exponential random graphs. *Ann. Appl. Probab.*, to appear
[43] Yin, M. and Zhu, L. Asymptotics for sparse exponential random graph models. *Braz. J. Prob. Stat.*, to appear
[44] Yin, M. and Zhu, L. (2016) Reciprocity in directed networks. *Phys. A* **447** 71–84.
[45] Zhu, L. Asymptotic structure of constrained exponential random graph models. `arXiv:1408.1536`