

Universality in Prelimiting Tail Behavior for Regular Subgraph Counts in the Poisson Regime

Mriganka Basu Roy Chowdhury^a

Submitted: Sep 27, 2024; Accepted: May 28, 2025; Published: Sep 19, 2025

© The author. Released under the CC BY-ND license (International 4.0).

Abstract

Let N be the number of copies of a small subgraph H in an Erdős-Rényi graph $G \sim \mathcal{G}(n, p_n)$ where $p_n \rightarrow 0$ is chosen so that $\mathbb{E}N = c$, a constant. Results of Bollobás [4] show that for regular graphs H , the count N weakly converges to a Poisson random variable. For large but finite n , and for the specific case of the triangle, investigations of the upper tail $\mathbb{P}(N \geq k_n)$ by Ganguly, Hiesmayr and Nam [Combinatorica 44, 2024] revealed that there is a phase transition in the tail behavior and the associated mechanism. Smaller values of k_n correspond to disjoint occurrences of H , leading to Poisson tails, with a different behavior emerging when k_n is large, guided by the appearance of an almost clique. We show that a similar phase transition also occurs for the tail probabilities when H is any regular graph, at the point where $k_n^{1-2/q} \log k_n = \log n$ (q is the number of vertices in H). This establishes universality of this transition, previously known only for the case of the triangle.

Mathematics Subject Classifications: 05C80, 60F10

1 Introduction

Let $\mathcal{G}(n, p)$ be the Erdős-Rényi random graph where each edge is present independently with probability p . Over the last few decades, various properties of these graphs have been investigated, resulting in a very fine understanding of this model. For a fixed graph H , let $Q_H(G)$ be the number of copies of H in $G \sim \mathcal{G}(n, p)$. Starting with the seminal paper of Erdős and Rényi [12], the distribution of $Q_H(G)$, both for fixed p and for p varying with n at different rates, has been the subject of intense research. For example, the probability of $Q_H(G) > 0$ was studied by [22] where H was a complete graph. This work was later extended to other subgraphs by Bollobás [4]. In addition to this, in many cases we also know asymptotic distributions of $Q_H(G)$. A criterion for asymptotic normality of $Q_H(G)$ was obtained by Ruciński [21]. This however must exclude the regime of p_n where we expect $\Theta(1)$ copies of the subgraph in expectation where asymptotic normality cannot

^a Department of Statistics, University of California, Berkeley, U.S.A. (mriganka.brc@berkeley.edu).

hold. A natural guess for the limiting distribution in this case is Poisson, but it turns out that it is not true in general. However, for the large class of *strictly balanced graphs* (which includes all regular graphs), asymptotic Poisson behavior was established by Bollobás [4] (see the discussion at the beginning of Section 3 for the definition of strictly balanced graphs).

A particularly active direction of research has been the study of the upper tail probability $\mathbb{P}(Q_H(G) \geq k_n)$, the so-called “infamous upper tail”. The seminal paper of Chatterjee and Varadhan [8] achieved asymptotically sharp estimates on this for the case of fixed $p_n = p > 0$, using the theory of dense graph limits and graphons (see [18] or [6]) as well as the celebrated Szemerédi regularity lemma [23], by reducing the computation of the upper tail probability to a natural “mean-field” variational problem. However, due to poor quantitative estimates in the regularity lemma, arguments in [8] could only be extended to p_n decaying as a negative power of $\log n$, see [20]. The most widely studied case is of $H = K_3$, the triangle. To allow p_n vanishing polynomially in n , Chatterjee and Dembo [7] developed the powerful nonlinear large deviations machinery, which, in the case of triangles, established the validity of the mean-field problem for p_n to decay slightly slower than $n^{-1/42}$ (a subsequent work of Eldan [11] improved to $p_n \gg n^{-1/18}$). However, the mean-field variational formulation is expected to hold as long as $p_n \gg \log n/n$. Major progress towards this was achieved by Augeri [1] and Cook and Dembo [9], extending the range of valid p_n to $\gg n^{-1/2}$, overcoming a significant barrier in this problem. Finally, a breakthrough by Harel, Mousset and Samotij [15] essentially solved the upper tail problem in the regime $p_n \gg 1/n$, using methods inspired by the classical moment arguments of Janson, Oleszkiewicz and Ruciński [16], leaving the interesting case of $p_n = \Theta(1/n)$ still open. Sharp understanding of the upper tail (for the case of K_3) in this sparse regime was finally achieved by Ganguly, Hiesmayr and Nam [14] and Chakraborty, van der Hofstad and den Hollander [5], thereby completing our understanding of the upper tail in the case of the triangle.

Let us now turn our attention to the upper tail problem for general graphs, results for which go back to [16]. The results established in [7] and [11] certified the validity of the mean-field variational problem for general graphs, sharp estimates for which were obtained by [3]. In this context, regular graphs turn out to be easier to analyze, and consequently the results obtained are sharper in this case. For instance, refining the techniques of [15], [2] proved an upper-tail result for general regular graphs H whenever $np_n^{\Delta/2} \gg (\log n)^{1/(q-2)}$, where Δ is the degree of H and q is the number of vertices of H . In contrast to this, our primary focus in this paper is to consider the case when the expected count is $\Theta(1)$ instead. Note that the expected number of copies of H in an Erdős-Rényi graph G is simply

$$\mathbb{E}Q_H(G) = \binom{n}{q} p_n^{q\Delta/2} = \Theta\left((np_n^{\Delta/2})^q\right), \quad \text{as } n \rightarrow \infty.$$

which is $\Theta(1)$ when $np_n^{\Delta/2} = \Theta(1)$.

As discussed earlier, the results of [4] show that $Q_H(G)$ is asymptotically Poisson in this case (hence the term “Poisson regime”). However, the arguments there do not yield

effective quantitative prelimiting tail estimates, which are quantities of the form

$$\mathbb{P}(Q_H(G) \geq k_n),$$

for an arbitrary sequence $k_n \rightarrow \infty$. (A point worth noting here is that since, in our regime, the subgraph count is Poisson with constant mean, probabilities of the form

$$\mathbb{P}(Q_H(G) \geq (1 + \delta)\mathbb{E}Q_H(G)),$$

for δ fixed, is given by the Poisson tail. Thus, our primary interest is in k_n growing as a function of n .) In the case of $H = K_3$, the triangle, this prelimiting upper tail was investigated only recently in [14], where it was shown that the behavior of this upper tail probability undergoes a transition when k_n is around $O(\log^3 n)$ (that the behavior for k_n larger than this persists till the maximum of $k_n = \Theta(n^3)$ was proved in [5]). The structural reason for the transition is the following: when k_n is smaller than the threshold, the dominant mechanism is the occurrence of many disjoint copies, but when k_n is above the threshold, the mechanism changes to having an (almost) clique.

In this paper we show that the tail probabilities exhibit a similar phase transition also happens when H is any regular graph. As above, we will fix H to be a connected q -vertex Δ -regular graph (with $q \geq 3$) throughout this paper. Our main result is the following.

Theorem 1. *Let $k_n \geq 2$ and $p_n = n^{-2/\Delta}$. If $G \sim \mathcal{G}(n, p_n)$, the probability of $Q_H(G) \geq k_n$ satisfies*

$$C_1 \exp(-C_2 L_n) \leq \mathbb{P}(Q_H(G) \geq k_n) \leq C_3 \exp(-C_4 L_n), \quad \text{where,}$$

$$L_n \stackrel{\text{def}}{=} \min(k_n \log k_n, k_n^{2/q} \log n).$$

Here the constants $C_1, C_2, C_3, C_4 > 0$ only depend on H . (This choice of $p_n = n^{-2/\Delta}$ ensures that $\mathbb{E}Q_H(G) = \Theta(1)$.)

It is easy to extend this result for $p_n = dn^{-2/\Delta}$ for any constant $d > 0$, but we choose $d = 1$ for notational simplicity. Theorem 1 almost (up to the constant preceding the exponent) achieves the sharp phase transition for the upper-tail probability at k_n satisfying $k_n^{1-2/q} \log k_n = \log n$, as obtained in [14] for the special case of $H = K_3$. Note that although our result is a significant first step, it does not indicate the precise constant in the exponent, or produce a structure theorem analogous to [14]. These questions are of great relevance, and we hope to address them in future work.

Remark: Throughout this paper, a *copy* of a graph H in G is counted only up to automorphisms. Equivalently, it is a subset of edges of G forming a graph isomorphic to H .

2 Idea of proof

We begin by setting up the notation. For any graph G , we will use $V(G), E(G)$ to denote the vertex set and the edge set respectively, while $v(G), e(G)$ will be the respective sizes

of these sets. We will write d_v to denote the degree of a vertex v . $C = C(H)$ will always denote a constant that depends only on H .

There are two main subparts to the proof. The first part deals with the case when k_n is small, and the second part is for the case of large k_n . It turns out that arguments in the first part works for all $k_n \leq n^c$ for some small c that depends on H , and the second part works for all k_n that grow faster than everything that is poly log n . It is then clear that these two proofs cover all possible values of k_n . The phase transition will occur in the former case, while the purpose of the latter is to show that the behavior shown in the first part persists after the phase transition all the way up to the maximum possible value of k_n .

The proof for the case of small k_n is via a reduction to the analysis of the structure of “spanned graphs”. This is inspired by [14] which treats the same problem in the case of the triangle. However, informally, that paper uses the fact that since the critical probability for a triangle is $p_n = O(1/n)$, the probability that a given subgraph with v vertices and e edges occurs as an isomorphic copy is at most

$$\binom{n}{v} p_n^{-e} \leq n^{-(e-v)}.$$

Thus this probability is directly related to the quantity $e - v + 1$ which can be interpreted as the number of “excess edges” in the subgraph after a spanning tree has been chosen. Such an idea fails in our case because for a general Δ -regular graph the critical probability is instead $p_n = n^{-2/\Delta}$, so that the above bound is now $n^{-(2e/\Delta - v)}$, for which a similar interpretation is not available. To circumvent this issue, we prove that even in this general case, any “spanned graph” G' must satisfy $2e(G')/\Delta - v \geq C e(G')$ for some constant C , see Lemma 12. Armed with this inequality, we bound the probability that the graph contains a certain spanned graph G' which has many copies of H . Due to the last inequality, this is now related to the minimal number of edges in a candidate G' containing, say, ℓ copies of H , which is at least $\ell^{2/q}$ up to constants (due to Lemma 7). It turns out that the concavity of this map $\ell \mapsto \ell^{2/q}$ implies that it is either optimal to have all the copies be in a large spanned component, or occur disjointly, depending on how large k_n is. As a consequence of this dichotomy, we obtain our phase transition. However, a formal proof requires one to also control the entropy stemming from the number of possible ways to distribute k_n copies of H among spanned graphs. This is done using a dyadic decomposition argument where we classify spanned components into logarithmic many collections depending on the number of copies they contain. See Lemma 14 for more details.

In the case of large k_n , the main idea is that if a graph has too many copies of H , then there must occur certain subgraphs, such that the expected count of H conditioned on their occurrence in the random graph is near optimal (see Lemma 16 for more details). This idea first appeared in [15] (who called such subgraphs “cores”), and was adopted by [5] to prove a result similar to ours but only for the triangle. However a naïve implementation of the approaches found in these papers fail in our case. As Lemma 16 will show, the upper tail bound essentially boils down to upper bounding *the occurrence probability of*

a core. This involves an efficient union bound over all cores. To illustrate the difficulties involved, let us see why a naïve direct count approach already fails for the case of $H = C_4$, the cycle on 4 vertices. In the case $k_n \sim n$, it turns out that all the complete bipartite graphs $K_{2,m}$ (i.e. 2 vertices on one side) with $m = O(\sqrt{n})$ qualify as cores. Note that their count is of the order of $\binom{n}{m} = \exp(O(\sqrt{n} \log n))$ whose exponent exactly matches the probability bound of $C' \exp(-C\sqrt{n} \log n)$ as claimed by Theorem 1 in this case. This reduces the direct union bound approach to a delicate game of constants.

Our proof avoids these issues, and instead crucially uses two estimates:

1. an upper bound on the number of copies of H in another graph G in terms of $e(G)$,
2. and, a lower bound on the product of degrees $d_u d_v$ for all edges (u, v) in a core.

It turns out that for simple graphs H , these estimates can follow from direct combinatorial arguments that exploit the specific structure of H . For example, in the case of cycles, a standard spectral argument is sufficient to deduce an (essentially optimal) upper bound on the number of copies of it in a graph with e edges. Further, the inequality involving the product of degrees can also be established via direct combinatorial methods. We expect other simple classes of graphs to also yield to similar arguments. Note that this is similar in spirit to [5] whose analysis also crucially relied on the geometry of the triangle. But since we wish to deal with all regular graphs at once, we must employ alternate means to deduce information about the structure of these cores. For this reason, we develop a novel strategy invoking *Finner's inequality* [13]. This is a form of Hölder's inequality which has proved to be particularly useful in the context of graphons, as demonstrated, for example, in [19] (see section 4). A straightforward application of this inequality yields the desired bound on the number of copies of H in a graph with e edges, as proved in Lemma 7. But this alone is not sufficient for the proof. A much more involved analysis also allows us to also prove bounds on the expected count of H in a graph containing a specific edge (Lemma 9). These results allow us to not only show that cores themselves contain many copies of H (Lemma 19), but also that each edge in the core has the interesting property of having a large product of degrees (Lemma 20), as asserted in point (2) above. All these facts are finally combined in the dyadic decomposition argument of Lemma 21, which essentially exploits the fact that since the product of degrees is high, very low degree vertices can only have edges to vertices with very high degree. Since there are few vertices with very high degree, this allows us to cheaply account for the neighbor sets of these low degree vertices, which are the main sources of trouble in the proof.

We end this section with a word about asymptotic notation. In addition to the usual $O(\cdot)$, $\Omega(\cdot)$, $\Theta(\cdot)$, we also use $\tilde{O}(g(n)) \ni f(n)$ to mean that $f(n)/\text{poly log } n = O(g(n))$, and $\tilde{\Omega}(g(n)) \ni f(n)$ to mean that $f(n)\text{poly log } n = \tilde{\Omega}(g(n))$. Finally, $\tilde{o}(g(n)) \ni f(n)$ if and only if $f(n)h(n) = o(g(n))$ for every $h(n)$ that is poly log n . So, for example, $n^{0.9} = \tilde{o}(n)$ but $n/\log^{100} n \neq \tilde{o}(n)$. In fact $n/\log^{100} n = \tilde{\Omega}(n)$.

3 Lower bounds

This short section collects the necessary results allowing us to assert the lower bound in our main theorem.

Define the *density* of a graph H to be the ratio $d(H) \stackrel{\text{def}}{=} e(H)/v(H)$ (see [17, Chapter 3]), and call a connected graph *strictly balanced* if for any proper subgraph $H' \subsetneq H$, $d(H') < d(H)$. It is easy to see that Δ -regular graphs are strictly balanced, because the entire graph has density $\Delta/2$, but any proper subgraph must have some vertex with degree $< \Delta$, so that the density is strictly smaller than $\Delta/2$. By the Poisson limit theorem for subgraph counts at the threshold¹ due to Bollobás [4], the asymptotic distribution of $Q_H(\mathcal{G}(n, p_n))$ is therefore Poisson. This fact is also used in the proof of the lemma below, which asserts a lower bound on the probability of having disjoint copies of H .

Lemma 2 (Disjoint occurrence). *Let D_s denote the event that there are s disjoint copies of H in $G \sim \mathcal{G}(n, p = p_n)$. Fix $\varepsilon > 0$. If $s \leq n^{1-\varepsilon}$,*

$$\mathbb{P}(D_s) \geq C \exp(-C' s \log s)$$

for all sufficiently large n depending on ε . Here C, C' are constants depending on H .

The matching upper bound is stated in Lemma 11.

Proof. Divide the vertex set $[n]$ of K_n into s (almost) equal sized groups of size either $\lfloor n/s \rfloor$ or $\lceil n/s \rceil$. Set $m \stackrel{\text{def}}{=} \lfloor n/s \rfloor$. Then,

$$\mathbb{P}(s \text{ disjoint copies of } H) \geq \mathbb{P}(\mathcal{G}(m, p_n) \text{ has a copy of } H)^s \quad (1)$$

To lower bound the probability of the latter, let $q = m^{-2/\Delta}$, and observe that $p_n = n^{-2/\Delta} \geq (2ms)^{-2/\Delta}$, as long as

$$2ms \geq n \iff 2\lfloor n/s \rfloor \geq n/s,$$

which happens for all sufficiently large n depending on s (and uniformly in $s \leq n^{1-\varepsilon}$ depending on ε). We assume this holds below. Then

$$\mathbb{P}(\mathcal{G}(m, p_n) \text{ has a copy of } H) \geq \mathbb{P}(\mathcal{G}(m, (2ms)^{-2/\Delta}) \text{ has a copy of } H). \quad (2)$$

We now employ a two stage sampling procedure as follows: to sample a $\mathcal{G}(m, (2ms)^{-2/\Delta})$, we first sample a $\mathcal{G}(m, m^{-2/\Delta})$, and then keep each of its edges independently with probability $(2s)^{-2/\Delta}$. It is clear that the resulting graph is a sample from $\mathcal{G}(m, (2ms)^{-2/\Delta})$. But crucially, since the expected number of copies of H in $\mathcal{G}(m, m^{-2/\Delta})$ is $\Theta(1)$, we can employ the classical fact that the distribution of subgraph counts at the threshold is Poisson for strictly balanced H (see, for instance, [4] or [17, Theorem 3.19]), to conclude that

$$\mathbb{P}(\mathcal{G}(m, m^{-2/\Delta}) \text{ has a copy of } H) \geq c,$$

¹This terminology is often used in the literature to mean a choice of p_n for which $\mathbb{E}Q_H(\mathcal{G}(n, p_n)) = \Theta(1)$.

for a constant c depending on H for all sufficiently large m . A given copy of H in $\mathcal{G}(m, m^{-2/\Delta})$ survives the second round of the procedure only if all its edges survive, which happens with probability

$$\left((2s)^{-2/\Delta}\right)^{e(H)} = (2s)^{-\frac{2}{\Delta} \frac{q\Delta}{2}} = (2s)^{-q}.$$

Putting these together,

$$\mathbb{P}(\mathcal{G}(m, (2ms)^{-2/\Delta}) \text{ has a copy of } H) \geq c(2s)^{-q},$$

so that we can invoke (1) and (2) to obtain

$$\mathbb{P}(s \text{ disjoint copies of } H) \geq (c(2s)^{-q})^s \geq C \exp(-C' s \log s),$$

for sufficiently large n depending on ε . □

The next lemma is a lower bound on clique occurrence probabilities.

Lemma 3 (Clique). *The probability that there is a clique of size $s \geq 2$ in $G \sim \mathcal{G}(n, p = p_n)$ is at least $\exp(-cs^2 \log n)$ where c depends only on Δ .*

Proof. The probability of having a clique of size s is at least that of the event that any fixed set of s vertices have all edges in between themselves. The probability of this latter event is exactly:

$$p_n^{\binom{s}{2}} \geq \exp\left(-\frac{1}{2}(s-1)^2 \log(1/p_n)\right) = \exp(-cs^2 \log n),$$

where $c = c(\Delta)$. This proves the result. □

These two results together imply

Lemma 4 (Lower bound of Theorem 1). *In the notation above,*

$$\mathbb{P}(Q_H(G) \geq k_n) \geq C \exp(-C' L_n) = C \exp(-C' \min(k_n \log k_n, k_n^{2/q} \log n))$$

for constants C, C' depending on H .

Proof. Given the lemmas above, the only remaining part is that the smallest clique containing at least k_n copies of H is of size $Ck_n^{2/q}$, but this is immediate from a counting argument. □

4 Finner's inequality

As described earlier, one of the most important results we require throughout the sequel is the following variant of Hölder's inequality. As far as we could tell, applications of this inequality in the context of large deviations on random graphs first appeared in [19], where they used it to analyze the variational problem associated with the upper tail deviations for $\mathcal{G}(n, p)$ with fixed p .

Lemma 5 (Finner's inequality, [13]). *Let $\mu_1, \mu_2, \dots, \mu_n$ be probability measures on $\Omega_1, \Omega_2, \dots, \Omega_n$ respectively and define $\Omega = \prod_{i=1}^n \Omega_i$ and $\mu = \prod_{i=1}^n \mu_i$. Also let A_1, A_2, \dots, A_m be nonempty subsets of $[n] = \{1, \dots, n\}$, and for any subset $A \subseteq [n]$, denote $\mu_A = \prod_{i \in A} \mu_i$ and $\Omega_A = \prod_{i \in A} \Omega_i$. Suppose $f_i \in L^{p_i}(\Omega_{A_i}, \mu_{A_i})$ for each $i \in [m]$, such that for all $x \in [n]$,*

$$\sum_{i: x \in A_i} p_i^{-1} \leq 1.$$

Then we have the inequality

$$\int \prod_{i=1}^m |f_i| d\mu \leq \prod_{i=1}^m \left(\int |f_i|^{p_i} d\mu_{A_i} \right)^{1/p_i}.$$

Here, on the left-hand side, one should think of each f_i extended to a function on Ω which depends only on the coordinates in A_i .

Our primary application of this will be to count homomorphisms from H to G , using the *graphon representation* of G .

Definition 6 (Graphon representation of a finite graph \mathbf{G}). Let G be a finite graph on n vertices, and without loss of generality, suppose its vertices are v_0, \dots, v_{n-1} . The graphon (representation) of G is defined as a function $f_G : [0, 1]^2 \rightarrow \{0, 1\}$ as follows.

$$f_G(x, y) = 1 \iff \exists 0 \leq i, j < n : \frac{i}{n} \leq x < \frac{i+1}{n}, \frac{j}{n} \leq y < \frac{j+1}{n}, (v_i, v_j) \in E(G)$$

else it is zero.

It is a classical result following from direct computation that the number of homomorphisms from H to G is given by (see, for instance, [18]*Chapter 7, or [6]*Chapter 3),

$$n^q \int_{[0,1]^q} \prod_{(u,v) \in E(H)} f(t_u, t_v) \prod_{u \in V(H)} dt_u.$$

This would enable us to bound the number of homomorphisms from H to G , and therefore the number of copies of H in G .

Later, we will need a slightly more general result relaxing the regularity requirement for H . While it may also be deduced from [19]*Corollary 3.2, we provide a proof for completeness.

Lemma 7. Let H be q -vertex e -edge (not necessarily connected) graph with each degree bounded above by Δ . Let G be a graph with graphon f_G . Write $n = |V(G)|$, and $m = |E(G)|$. Then the number of homomorphisms of H in G , $N_H(G)$ is at most

$$N_H(G) \leq C n^{q-2e/\Delta} m^{e/\Delta}.$$

In particular for a regular H , $2e/\Delta = q$, so we have $N_H(G) \leq C m^{q/2}$.

Proof. We apply Finner's inequality with $\Omega_i = [0, 1]$ for all $i \in [n]$, and μ_i being the uniform measure on $[0, 1]$. For each edge $e = (u, v)$ of H we choose $A_e = \{u, v\}$ and $p_e = \Delta$. Then the conditions are satisfied, and we get

$$\begin{aligned} \int_{[0,1]^q} \prod_{(u,v) \in E(H)} f_G(t_u, t_v) \prod_{u \in V(H)} dt_u &\leq \prod_{(u,v) \in E(H)} \left(\int_{[0,1]^2} f_G(t_u, t_v)^\Delta dt_u dt_v \right)^{1/\Delta} \\ &= \left(\frac{2m}{n^2} \right)^{e/\Delta} \\ &\leq C m^{e/\Delta} n^{-2e/\Delta}, \end{aligned}$$

where we use the fact that f_G is zero-one valued, so, $f_G^\Delta = f_G$. This finishes the proof. \square

The graphon representation and the formula for homomorphism density is also useful in bounding the *expected number of copies* of a particular graph H in a random graph G with independent edges. For a given collection of probabilities $0 \leq p_{uv} \leq 1$ for $u < v \in [n]$, consider the random graph model on the vertex set $[n]$ where the edge (u, v) is present with probability p_{uv} , independently of other edges (undirected edges). For such a model we define:

Definition 8 (Graphon representation of a finite random graph G with independent edges). For the model described above, the associated graphon is defined as a function $f_G : [0, 1]^2 \rightarrow [0, 1]$ satisfying

$$f_G(x, y) = p_{ij} \iff \exists 0 \leq i, j < n : \frac{i}{n} \leq x < \frac{i+1}{n}, \frac{j}{n} \leq y < \frac{j+1}{n}, (i, j) \in E(G)$$

else it is zero.

A particular application of Finner's inequality also produces the following result about the number of copies of H in G containing a given edge $e \in G$, which, as alluded to in section 2, will be crucial in our analysis of the core, see Lemma 20. While similar results have appeared in the literature, this precise form appears to be new.

Lemma 9. Let G^* be a subgraph of K_n (with vertex set $[n]$). Consider the random graph model G on $[n]$ where the edges in G^* are present with probability 1 and every other edge is present with probability p independently. Suppose G^* has a distinguished edge (a, b) ,

and let H be a Δ -regular graph with q vertices. Then, the expected number of copies of H in G containing (a, b) is at most

$$\mathbb{E}Q_H(G) \leq C(H) (d_a + np^\Delta)^{\frac{\Delta-1}{\Delta}} (d_b + np^\Delta)^{\frac{\Delta-1}{\Delta}} (e + n^2 p^\Delta)^{q/2-2+1/\Delta}$$

where $C(H)$ is a constant depending only on H , $e = e(G^*)$ and d_a, d_b are the degrees of a, b respectively in G^* .

In addition, if $Q'_H(G)$ is the number of expected number of copies of H in G containing (a, b) where some edge must come from outside G^* , then

$$\mathbb{E}Q'_H(G) \leq C(H) \max(B_1, B_2, B_3)$$

where

$$\begin{aligned} B_1 &\stackrel{\text{def}}{=} (d_a + np^\Delta)^{\frac{\Delta-2}{\Delta}} (d_b + np^\Delta)^{\frac{\Delta-1}{\Delta}} (e + n^2 p^\Delta)^{q/2-2+1/\Delta} pn^{1/\Delta}, \\ B_2 &\stackrel{\text{def}}{=} (d_a + np^\Delta)^{\frac{\Delta-1}{\Delta}} (d_b + np^\Delta)^{\frac{\Delta-2}{\Delta}} (e + n^2 p^\Delta)^{q/2-2+1/\Delta} pn^{1/\Delta}, \\ B_3 &\stackrel{\text{def}}{=} (d_a + np^\Delta)^{\frac{\Delta-1}{\Delta}} (d_b + np^\Delta)^{\frac{\Delta-1}{\Delta}} (e + n^2 p^\Delta)^{q/2-2} pn^{2/\Delta}. \end{aligned}$$

The terms B_1, B_2, B_3 correspond to the cases when the edge coming from outside G^* is adjacent to a, b or none, respectively.

Proof. We begin by proving the first claim. Observe that

$$Q_H(G) \leq \sum_{u \sim v} N(u, v),$$

where

$$N(u, v) \stackrel{\text{def}}{=} \sum_{\substack{\phi: V(H) \rightarrow [n] \\ \phi(u)=a, \phi(v)=b, \\ \phi \text{ injective}}} \prod_{x \sim u} \mathbf{1}_{(\phi(u), \phi(x)) \in E(G)} \prod_{y \sim v} \mathbf{1}_{(\phi(v), \phi(y)) \in E(G)} \prod_{\substack{w \sim z \\ w, z \neq u, v}} \mathbf{1}_{(\phi(w), \phi(z)) \in E(G)}. \quad (3)$$

$N(u, v)$ counts all (injective) mappings $\phi: H \rightarrow [n]$ preserving edges in H , such that the edge (u, v) maps to (a, b) . Here by $x \sim y$ we mean $(x, y) \in E(H)$ (for brevity, in this proof we also adopt the convention that for u , if we write $x \sim u$, we assume that $x \neq v$, and similarly for v). Therefore, it is sufficient to fix $u \sim v$ and prove the result, because summing over all distinct $u \sim v$ will only change the constant $C(H)$. For fixed $u \sim v$ we have

$$\mathbb{E}N(u, v) \leq \sum_{\phi} \mathbb{E} \left[\prod_{x \sim u} \mathbf{1}_{(\phi(u), \phi(x)) \in E(G)} \prod_{y \sim v} \mathbf{1}_{(\phi(v), \phi(y)) \in E(G)} \prod_{\substack{w \sim z \\ w, z \neq u, v}} \mathbf{1}_{(\phi(w), \phi(z)) \in E(G)} \right],$$

where ϕ is understood to satisfy the conditions in (3). Since ϕ is injective, all the edge random variables in the expectation above are independent, so we can write:

$$= \sum_{\phi} \prod_{x \sim u} \mathbb{E} [\mathbf{1}_{(\phi(u), \phi(x)) \in E(G)}] \prod_{y \sim v} \mathbb{E} [\mathbf{1}_{(\phi(v), \phi(y)) \in E(G)}] \prod_{\substack{w \sim z \\ w, z \neq u, v}} \mathbb{E} [\mathbf{1}_{(\phi(w), \phi(z)) \in E(G)}].$$

For any vertex $\alpha \in [n]$, define the block of α as $B_{\alpha} \stackrel{\text{def}}{=} [\alpha/n, (\alpha + 1)/n)$. Then note that, for instance, $\mathbb{E} [\mathbf{1}_{(\phi(x), \phi(y)) \in E(G)}] = f(t_x, t_y)$ for any $t_x \in B_{\phi(x)}$, $t_y \in B_{\phi(y)}$, where f is the graphon associated with the random graph model G . For any collection $\{t_x \in B_{\phi(x)} : x \in H\}$, we then have that:

$$= \sum_{\phi} \prod_{x \sim u} f(t_u, t_x) \prod_{y \sim v} f(t_v, t_y) \prod_{\substack{w \sim z \\ w, z \neq u, v}} f(t_w, t_z).$$

Fix $t_u \in B_a, t_v \in B_b$. Since the summand is a constant over all choices of $(t_x)_{x \neq u, v}$ as long as $(t_x)_{x \neq u, v} \in \prod_{x \neq u, v} B_{\phi(x)}$, the above sum is the same as

$$= n^{q-2} \sum_{\phi} \int_{\prod_{x \neq u, v} B_{\phi(x)}} \prod_{x \sim u} f(t_u, t_x) \prod_{y \sim v} f(t_v, t_y) \prod_{\substack{w \sim z \\ w, z \neq u, v}} f(t_w, t_z) \prod_{x \neq u, v} dt_x,$$

where the factor of n^{q-2} is because the volume of each block $B_{\phi(x)}$ is n^{-1} . Finally observe that across various choices of ϕ , the domains $\prod_{x \neq u, v} B_{\phi(x)}$ are disjoint, so that the above is

$$\leq n^{q-2} \int_{[0,1]^{q-2}} \prod_{x \sim u} f(t_u, t_x) \prod_{y \sim v} f(t_v, t_y) \prod_{\substack{w \sim z \\ w, z \neq u, v}} f(t_w, t_z) \prod_{x \neq u, v} dt_x.$$

At this point we apply Finner's inequality with the sets given by $\{x\}$ for all $x \sim u$, $\{y\}$ for all $y \sim v$, and $\{w, z\}$ for all $w \sim z$ with $w, z \neq u, v$. We use the weights $p_i \equiv \Delta$ for all these sets. It is easy to verify that the conditions in Lemma 5 hold, and therefore:

$$\leq n^{q-2} \prod_{x \sim u} \left(\int f(t_u, t_x)^{\Delta} dt_x \right)^{1/\Delta} \prod_{y \sim v} \left(\int f(t_v, t_y)^{\Delta} dt_y \right)^{1/\Delta} \prod_{\substack{w \sim z \\ w, z \neq u, v}} \left(\int f(t_w, t_z)^{\Delta} dt_w dt_z \right)^{1/\Delta}.$$

In these products, the values for x, y, w, z do not matter because the integrals are always over $[0, 1]$ for each variable, so we can simplify to (we drop the domains $[0, 1]$ and $[0, 1]^2$ from the notation):

$$= n^{q-2} \left(\int f(t_u, t_x)^{\Delta} dt_x \right)^{\frac{\Delta-1}{\Delta}} \left(\int f(t_v, t_y)^{\Delta} dt_y \right)^{\frac{\Delta-1}{\Delta}} \left(\int f(t_w, t_z)^{\Delta} dt_w dt_z \right)^{\frac{q}{2}-2+\frac{1}{\Delta}},$$

where we use the fact that the number of edges with no endpoint among u, v is $q\Delta/2 - (2\Delta - 1) = q\Delta/2 - 2\Delta + 1$. As the last step, we use the fact that $f(t_u, t_x)$ is 1 when $t_x \in B_c$ with $c \sim a$ in G^* and otherwise it is p . Similar observations also hold for $f(t_v, t_y)$ and $f(t_w, t_z)$. Therefore the above is

$$\leq n^{q-2} \left(\frac{d_a}{n} + p^\Delta \right)^{\frac{\Delta-1}{\Delta}} \left(\frac{d_b}{n} + p^\Delta \right)^{\frac{\Delta-1}{\Delta}} \left(\frac{2e}{n^2} + p^\Delta \right)^{\frac{q}{2}-2+\frac{1}{\Delta}},$$

because $f(t_u, t_x) = 1$ if and only if $\phi(u) \sim \phi(x) = a$ in G^* (otherwise it is p). A similar argument holds for the other two integrals. The factor of 2 (in $2e$) in the last integral is because each edge in G^* is counted twice. Simplifying the above,

$$= \underbrace{n^{q-2} \cdot n^{-2\frac{\Delta-1}{\Delta}-2(\frac{q}{2}-2+\frac{1}{\Delta})}}_{=1} (d_a + np^\Delta)^{\frac{\Delta-1}{\Delta}} (d_b + np^\Delta)^{\frac{\Delta-1}{\Delta}} (2e + n^2 p^\Delta)^{\frac{q}{2}-2+\frac{1}{\Delta}},$$

which completes the proof by absorbing the factor of 2 in front of e into the constant $C(H)$.

We now modify this to prove the second claim. Similar to (3), define

$$N^{\alpha,\beta}(u, v) \stackrel{\text{def}}{=} \sum_{\substack{\phi: V(H) \rightarrow [n] \\ \phi(u)=a, \phi(v)=b, \\ \phi \text{ injective}}} \prod_{x \sim u} \mathbf{1}_{(\phi(u), \phi(x)) \in E(G)} \prod_{y \sim v} \mathbf{1}_{(\phi(v), \phi(y)) \in E(G)} \\ \prod_{\substack{w \sim z \\ w, z \neq u, v}} \mathbf{1}_{(\phi(w), \phi(z)) \in E(G)} \cdot \mathbf{1}_{(\phi(\alpha), \phi(\beta)) \notin E(G^*)}$$

for all $(\alpha, \beta) \in E(H)$ such that $(\alpha, \beta) \neq (u, v)$. This counts the same quantity as $N(u, v)$, except now, (α, β) maps to an edge outside G^* . Observe that

$$Q'_H(G) \leq \sum_{(\alpha, \beta) \in E(H)} \sum_{E(H) \ni (u, v) \neq (\alpha, \beta)} N^{\alpha, \beta}(u, v)$$

so that it is again sufficient to prove the bound in question for $N^{\alpha, \beta}(u, v)$ for fixed $\alpha, \beta, u, v \in H$ instead of $Q'_H(G)$. Using calculations similar to the above we then obtain that

$$\mathbb{E} N^{\alpha, \beta}(u, v) \leq n^{q-2} \int_{[0,1]^{q-2}} \prod_{x \sim u} f(t_u, t_x) \prod_{y \sim v} f(t_v, t_y) \prod_{\substack{w \sim z, \\ w, z \neq u, v}} f(t_w, t_z) \cdot \mathbf{1}_{f(t_\alpha, t_\beta)=p} \prod_x dt_x,$$

because $f(t_\alpha, t_\beta) = p$ if and only if $(\phi(\alpha), \phi(\beta)) \notin E(G^*)$. As indicated in the remark after the statement, now there are several cases depending on if $\alpha = u$, $\alpha = v$ or $\alpha, \beta \neq u, v$ (the cases for β being similar).

• Case $\alpha = u, \beta \neq v$: The integral is then bounded by Finner's inequality similar to the above to obtain

$$\mathbb{E}N^{\alpha,\beta}(u, v) \leq n^{q-2} \prod_{\beta \neq x \sim u} \left(\int f(t_u, t_x)^\Delta dt_x \right)^{1/\Delta} \prod_{y \sim v} \left(\int f(t_v, t_y)^\Delta dt_y \right)^{1/\Delta} \\ \prod_{\substack{w \sim z \\ w, z \neq u, v}} \left(\int f(t_w, t_z)^\Delta dt_w dt_z \right)^{1/\Delta} \left(\int f(t_u, t_\beta)^\Delta \mathbf{1}_{f(t_u, t_\beta)=p} dt_\beta \right)^{1/\Delta}.$$

Observe that the last factor is $\leq p$, so the above is bounded by

$$\leq n^{q-2} \left(\frac{d_a}{n} + p^\Delta \right)^{\frac{\Delta-2}{\Delta}} \left(\frac{d_b}{n} + p^\Delta \right)^{\frac{\Delta-1}{\Delta}} \left(\frac{2e}{n^2} + p^\Delta \right)^{\frac{q}{2}-2+\frac{1}{\Delta}} \cdot p \\ = (d_a + np^\Delta)^{\frac{\Delta-2}{\Delta}} (d_b + np^\Delta)^{\frac{\Delta-1}{\Delta}} (2e + n^2 p^\Delta)^{\frac{q}{2}-2+\frac{1}{\Delta}} pn^{1/\Delta}.$$

• Case $\alpha = v, \beta \neq u$: A calculation similar to the previous case will yield:

$$\mathbb{E}N^{\alpha,\beta}(u, v) \leq (d_a + np^\Delta)^{\frac{\Delta-1}{\Delta}} (d_b + np^\Delta)^{\frac{\Delta-2}{\Delta}} (2e + n^2 p^\Delta)^{\frac{q}{2}-2+\frac{1}{\Delta}} pn^{1/\Delta}.$$

• Case $\alpha, \beta \neq u, v$: Then,

$$\mathbb{E}N^{\alpha,\beta}(u, v) \leq n^{q-2} \prod_{x \sim u} \left(\int f(t_u, t_x)^\Delta dt_x \right)^{1/\Delta} \prod_{y \sim v} \left(\int f(t_v, t_y)^\Delta dt_y \right)^{1/\Delta} \\ \prod_{\substack{w \sim z \\ w, z \neq u, v \\ (w, z) \neq (\alpha, \beta)}} \left(\int f(t_w, t_z)^\Delta dt_w dt_z \right)^{1/\Delta} \left(\int f(t_\alpha, t_\beta)^\Delta \mathbf{1}_{f(t_\alpha, t_\beta)=p} dt_\alpha dt_\beta \right)^{1/\Delta}.$$

Again the last factor is at most p , so that the above is bounded by

$$\leq n^{q-2} \left(\frac{d_a}{n} + p^\Delta \right)^{\frac{\Delta-1}{\Delta}} \left(\frac{d_b}{n} + p^\Delta \right)^{\frac{\Delta-1}{\Delta}} \left(\frac{2e}{n^2} + p^\Delta \right)^{\frac{q}{2}-2} \cdot p \\ = (d_a + np^\Delta)^{\frac{\Delta-1}{\Delta}} (d_b + np^\Delta)^{\frac{\Delta-1}{\Delta}} (2e + n^2 p^\Delta)^{\frac{q}{2}-2} pn^{2/\Delta}.$$

Taking a maximum over these three cases and modifying $C = C(H)$, we obtain the desired result. \square

5 Few copies

In this section, we will prove our main theorem for “small” values of k_n . Henceforth we will abbreviate $Q = Q_H(G)$.

Definition 10 (Spanned Graph). A graph G is called an **H -spanned graph** (or “spanned graph” for short) if G is connected and each edge is contained in a copy of H in G . We say it is spanned by ℓ copies of H if there are ℓ copies of H in G such that

each edge of G is contained in one of these copies. We will often use $\ell_* = \ell_*(G)$ to indicate the smallest possible value of ℓ given G .

For brevity, we will often say that a graph is spanned by ℓ copies of H if it is a spanned graph and it is spanned by ℓ copies of H .

This is the analog of “triangle-induced subgraphs” as defined in [14]. Also note that $\ell_* = 1$ if and only if $G = H$ up to isomorphism.

Next we define some useful events. Informally, the purpose of these events is to decompose the occurrences of H in G into connected components (where two components are adjacent if they share a vertex).

- F_ℓ : there is a spanned subgraph of G which is spanned by ℓ copies of H .
- D_s : there are s vertex-disjoint copies of H in G (for brevity we will write “disjoint” to mean vertex-disjoint).
- $E_{\ell_1, \dots, \ell_m}^s$: there are (at least) $s + m$ disjoint spanned subgraphs of G , with s of them containing exactly one copy of H , and the remaining m of them each spanned by ℓ_i copies for $1 \leq i \leq m$. Since the order of the ℓ_i are immaterial, we will always assume that $2 \leq \ell_1 \leq \ell_2 \leq \dots \leq \ell_m$. When $s = 0$, we drop s , and simply write $E_{\ell_1, \dots, \ell_m}$. (Here s stands for “singleton” copies, and m stands for “multiple” copies)

Using \circ for the “disjoint occurrence” operator on events (see [24, Section 3] for the definition), it is easy to see that

$$E_{\ell_1, \dots, \ell_m}^s = D_s \circ F_{\ell_1} \circ \dots \circ F_{\ell_m},$$

and so

$$\mathbb{P}(E_{\ell_1, \dots, \ell_m}^s) \leq \mathbb{P}(D_s) \prod_{i=1}^m \mathbb{P}(F_{\ell_i}), \quad (4)$$

due to the BK-inequality, see [24]. To see why such a bound is useful, observe that if a graph G satisfies $Q = Q_H(G) \geq k_n$, one can drop edges in G not contained in any copy of H , and decompose the remaining graph into connected components. Clearly each remaining component is H -spanned, and no copies of H were deleted, so that there are numbers s, ℓ_1, \dots, ℓ_m such that $G \in E_{\ell_1, \dots, \ell_m}^s$. The rest of this section will be devoted to efficiently union bounding over various choices of these numbers. But first we bound the probabilities of D_s and F_ℓ individually.

Lemma 11. $\mathbb{P}(D_s) \leq C \exp(-s \log s)$ for some $C = C(H)$.

Proof. The number of ways we can pick s disjoint subsets of $[n]$ is at most

$$\frac{1}{s!} \binom{n}{q} \binom{n-q}{q} \dots \binom{n-(s-1)q}{q} = \frac{n!}{s!(q!)^s(n-q s)!} \leq \frac{n^{qs}}{s!(q!)^s}.$$

Given a fixed set of vertices S of size q , the probability of it having a copy of H is at most

$$Cp_n^{-e} = Cn^{-2e/\Delta} = Cn^{-q},$$

where the constant $C = C(H)$ is due to the number of isomorphic copies of H in K_q . Then by a union bound, the probability of D_s satisfies

$$\mathbb{P}(D_s) \leq C \exp(qs \log n - s \log(q!) - \log(s!) - qs \log n) = C \exp(-s \log(q!) - \log(s!)).$$

By Stirling's formula, we have $\log(s!) = s \log s - s + O(\log s)$. Therefore

$$s \log(q!) + s \log s - s + O(\log s) \geq s \log s, \quad \forall s \geq 1,$$

(using $q \geq 3$ and $\log 6 > 1$) finishing the proof. \square

The next lemma delivers the control on $\frac{2}{\Delta}e(S) - v(S)$, serving as a replacement for the “tree-excess edges” interpretation of $e(S) - v(S) + 1$ in [14] as described in the idea of proof section. This will be useful for furnishing a bound on $\mathbb{P}(F_\ell)$.

Lemma 12. *If S is a graph spanned by $\ell \geq 2$ copies of H then $\frac{2}{\Delta}e(S) - v(S) \geq Ce(S)$, where $C = C(H)$.*

Proof. Since each copy of H contains at most q^2 edges, the optimal number of copies $\ell_* = \ell_*(S)$ needed to span S satisfies $\ell_* \geq Ce(S)$ with $C = 1/q^2$. Order these copies of H in S as H_1, \dots, H_{ℓ_*} , such that for each i , $V(H_i) \cap V(S_{i-1}) \neq \emptyset$, where S_i is the subgraph of S spanned by H_1, \dots, H_i (this can be done because S is spanned and connected). Note that in general S_i is distinct from the subgraph of S induced by $V(S_i)$. Each edge incident on any $x \in V(H_i) - V(S_{i-1})$ comes from H_i , and hence the degree of any such x is exactly Δ in S_i . For each $i > 1$, also write $b_i \geq 1$ for the number of edges between $V(H_i) - V(S_{i-1})$ and $V(S_{i-1})$ (the number of boundary edges), and let $v_i = |V(H_i) - V(S_{i-1})|$ (the number of new vertices). See Figure 1 below for an illustration of these definitions.

We track the quantity $f_i \stackrel{\text{def}}{=} \frac{2}{\Delta}e(S_i) - v(S_i)$. Observe that $e(S_i) \geq e(S_{i-1}) + b_i + e(H_i - S_{i-1})$, where $H_i - S_{i-1}$ is the subgraph of S_i induced by $V(H_i) - V(S_{i-1})$ (the inequality is due to ignoring edges in H_i with both endpoints in $V(S_{i-1})$). Crucially, note that

$$2e(H_i - S_{i-1}) + b_i = \Delta v_i$$

because both sides count the edges within $V(H_i) - V(S_{i-1})$ twice. Therefore

$$e(S_i) \geq e(S_{i-1}) + b_i + \frac{\Delta v_i - b_i}{2} = e(S_{i-1}) + \frac{b_i}{2} + \frac{\Delta v_i}{2},$$

and so,

$$\begin{aligned} f_i &= \frac{2}{\Delta}e(S_i) - v(S_i) \geq \frac{2}{\Delta}e(S_{i-1}) + \frac{b_i}{\Delta} + v_i - v(S_i) \\ &= \frac{2}{\Delta}e(S_{i-1}) + \frac{b_i}{\Delta} + v_i - v(S_{i-1}) - v_i \end{aligned}$$

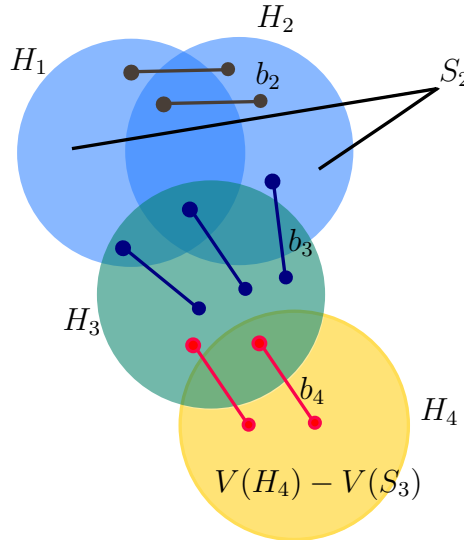


Figure 1: Variables in the proof of Lemma 12 with four copies of H .

(substituting $v(S_i) = v(S_{i-1}) + v_i$). Therefore, the above is

$$= f_{i-1} + \frac{b_i}{\Delta} \geq f_{i-1} + \frac{1}{\Delta} \quad (\text{since } b_i \geq 1).$$

Since $f_1 = 0$ we get $f_{\ell_*} \geq \frac{\ell_* - 1}{\Delta} \geq \frac{\ell_*}{2\Delta}$ as long as $\ell_* \geq 2$ which holds if $\ell \geq 2$ (this is because $\ell \geq 2$ implies that there are more than $e(H)$ edges in S). To conclude the proof, recall that $\ell_* \geq e(S)/q^2 = Ce(S)$. \square

Using this structural result, we can now provide an upper bound on $\mathbb{P}(F_\ell)$.

Lemma 13. $\mathbb{P}(F_\ell) \leq \exp(-C\ell^{2/q} \log n)$ for all $2 \leq \ell \leq n^c$ and all sufficiently large n , where $c = c(H)$, $C = C(H)$ are constants.

Proof. For a fixed graph S spanned by ℓ copies of H , the probability of its occurrence in $G \sim \mathcal{G}(n, p = n^{-2/\Delta})$ is at most

$$\binom{n}{v(S)} p^{e(S)} \leq n^{v(S) - \frac{2}{\Delta} e(S)},$$

and as we saw above in Lemma 12, for such a graph, $\frac{2}{\Delta} e(S) - v(S) \geq Ce(S)$. Therefore this probability is at most $\exp(-Ce(S) \log n)$.

Let $F_{\ell, v, e}$ be the event that there is a subgraph of G with v vertices, e edges, and spanned by ℓ copies of H (note that $v \leq q\ell$). The number of such S is at most $\binom{v^2}{e} \leq v^{2e}$, so by a direct union bound,

$$\begin{aligned} \mathbb{P}(F_{\ell, v, e}) &\leq \exp(-Ce \log n) v^{2e} \\ &\leq \exp(-Ce \log n) \exp(2e \log(q\ell)), \\ &\leq \exp(-Ce \log n), \end{aligned}$$

by a different choice of C , as long as $\ell \leq n^c$ for some small c depending on H . Finally, using the fact that $\ell \leq Ce^{q/2}$ (using Lemma 7) we see that the above is

$$\leq \exp(-C\ell^{2/q} \log n).$$

For a graph spanned by ℓ copies of H we have $v, e \leq C\ell$, so by union bounding over all such possibilities, we get

$$\mathbb{P}(F_\ell) \leq C\ell^2 \exp(-C\ell^{2/q} \log n) \leq \exp(-C\ell^{2/q} \log n),$$

(up to changing C) for sufficiently large n . \square

As alluded to earlier, we will now combine the estimates above using a union bound. The following definition will be useful.

Definition 1. For a tuple (c_1, c_2, \dots, c_t) , define $\tilde{E}_{c_1, \dots, c_t}$ to be the event that in the random graph $G \sim \mathcal{G}(n, p_n)$, there are $c_1 + \dots + c_t$ disjoint spanned graphs, with c_i of them spanned by 2^i copies of H , for all i .

Observe that $\tilde{E}_{c_1, \dots, c_t}$ is a special case of events of the form $E_{\ell_1, \dots, \ell_m}$ where $m = \sum_{i=1}^t c_i$ and the ℓ_i s are powers of 2. Also, if a graph G satisfies $E_{\ell_1, \dots, \ell_m}$, then there is a tuple (c_1, \dots, c_t) such that it also satisfies $\tilde{E}_{c_1, \dots, c_t}$, with

$$\sum_{i=1}^t c_i 2^i \geq \frac{1}{2} \sum_{j=1}^m \ell_j.$$

This is essentially done via “dropping” copies of H till the components are spanned by 2^i components for some i . More precisely, for each valid i , define

$$c_i \stackrel{\text{def}}{=} |\{j : 2^i \leq \ell_j < 2^{i+1}\}|. \quad (5)$$

We claim that $G \in \tilde{E}_{c_1, \dots, c_t}$. Let U_1, \dots, U_m be some choice of disjoint spanned subgraphs of G witnessing $E_{\ell_1, \dots, \ell_m}$. Choose j such that U_j is spanned by ℓ_j copies of H and let i be such that $2^i \leq \ell_j < 2^{i+1}$. Order these copies of H such that the subgraph induced by any number of copies from the beginning is connected (which can always be done because U_j is connected). Then take the union of the first 2^i copies. By construction this is a connected subgraph spanned by 2^i copies of H . Doing this for each j proves the claim.

As we will shortly see, $\mathbb{P}(E_{\ell_1, \dots, \ell_m})$ is comparable to $\mathbb{P}(\tilde{E}_{c_1, \dots, c_t})$, whereas the number of possible choices is a lot less for (c_1, \dots, c_t) than for (ℓ_1, \dots, ℓ_m) , facilitating an efficient union bound.

Lemma 14. For $2 \leq k_n \leq n^c$ (for a sufficiently small $c = c(H)$), we have

$$\mathbb{P}(Q \geq k_n) \leq C \exp(-C' \min(k_n \log k_n, k_n^{2/q} \log n)),$$

where $C = C(H)$, $C' = C'(H)$ are constants depending only on H .

As described earlier, this lemma demonstrates the transition in tail behavior promised in our main theorem. However, this result only holds for k_n at most a small polynomial of n , and one must show that there are *no other transitions*, i.e., this behavior persists all the way up to the maximum possible value of k_n . This is achieved in the next section after this proof.

Proof. We abbreviate $k = k_n$ throughout this proof. Recall from the discussion below (4) that the event $\{Q \geq k\}$ is equivalent to

$$\bigcup_{\substack{s, m \geq 0, \ell_i \geq 2 \\ s + \sum_{i \leq m} \ell_i \geq k}} E_{\ell_1, \dots, \ell_m}^s.$$

However any graph satisfying this also satisfies the event

$$\bigcup_{\substack{s, m \geq 0, \ell_i \geq 2 \\ s + \sum_{i \leq m} \ell_i = k}} E_{\ell_1, \dots, \ell_m}^s,$$

because a spanned graph spanned by ℓ copies of H has a subgraph spanned by ℓ' copies of H for any $\ell' \leq \ell$ by reasoning as below (5). We therefore focus on bounding the probability of the latter.

Using the BK inequality (4) and the bound on $\mathbb{P}(F_\ell)$ we obtain

$$\begin{aligned} \mathbb{P}(\tilde{E}_{c_1, \dots, c_t}) &\leq \prod_{i=1}^t \mathbb{P}(F_{2^i})^{c_i} \\ &\leq \exp\left(-C \log n \sum_{i=1}^t c_i 2^{2i/q}\right) \end{aligned}$$

We use this bound and the fact that for each collection ℓ_1, \dots, ℓ_m (with each $\ell_i \geq 2$), we can find a vector c_1, \dots, c_t with

$$\sum_{i \leq m} \ell_i \geq \sum_{j \leq t} c_j 2^j \geq \frac{1}{2} \sum_{i \leq m} \ell_i, \quad (6)$$

such that $E_{\ell_1, \dots, \ell_m} \subseteq \tilde{E}_{c_1, \dots, c_t}$. This allows us to conclude that for any k ,

$$\begin{aligned} \mathbb{P}\left(\bigcup_{\substack{m \geq 1, \ell_i \geq 2 \\ \sum_{i \leq m} \ell_i = k}} E_{\ell_1, \dots, \ell_m}\right) &\leq \mathbb{P}\left(\bigcup_{\substack{t \geq 1, c_i \geq 0 \\ k \geq \sum_i c_i 2^i \geq k/2}} \tilde{E}_{c_1, \dots, c_t}\right) \\ &\leq \exp((\log k)^2) \exp\left(-C \log n \cdot \min \sum_{i \geq 1} c_i 2^{2i/q}\right), \end{aligned}$$

where the min is over all possible c_i s satisfying the condition in (6). The $\exp((\log k)^2)$ term is a simple upper bound on the number of possible tuples (c_1, \dots, c_t) where $t \leq \log_2 k$ and $0 \leq c_i \leq k/2$. By an elementary inequality (see Lemma 22 in the Appendix) we see that the above is

$$\leq \exp(C(\log^2 k - k^{2/q} \log n)).$$

Now observe that $E_{\ell_1, \dots, \ell_m}^s = D_s \circ E_{\ell_1, \dots, \ell_m}$. Another application of the BK inequality and our estimate of $\mathbb{P}(D_s)$ from Lemma 11 yields

$$\begin{aligned} \mathbb{P} \left(\bigcup_{\substack{s, m \geq 0, \ell_i \geq 2 \\ s + \sum_{i \leq m} \ell_i = k}} E_{\ell_1, \dots, \ell_m}^s \right) &\leq \sum_{s \leq k} \mathbb{P}(D_s) \mathbb{P} \left(\bigcup_{\substack{m \geq 0, \ell_i \geq 2 \\ \sum_{i \leq m} \ell_i = k-s}} E_{\ell_1, \dots, \ell_m} \right) \\ &\leq C \exp(\log^2 k + \log k) \exp \left(-C \min_{1 \leq s \leq k} (s \log s + (k-s)^{2/q} \log n) \right), \end{aligned}$$

at which point we invoke Lemma 23 from the Appendix to obtain that the above is

$$\leq C \exp(2 \log^2 k) \exp(-C \min(k \log k, k^{2/q} \log n)).$$

Finally observe that since $2 \log^2 k \ll \min(k \log k, k^{2/q} \log n)$ as $k \rightarrow \infty$, by modifying constants we may drop the prefactor of $\exp(2 \log^2 k)$, yielding the required result. \square

6 Many copies

In this section we will consider the case of large k_n and demonstrate that the behavior described in Lemma 14 actually persists all the way up to the maximum possible value of k_n . However our results here will only hold for k_n such that $k_n \gg f(n)$ for all $f(n) = \text{poly log } n$ (we will assume this lower bound on k_n throughout this section). But this suffices to complete the proof of our main theorem since Lemma 14 holds even for k_n that are small polynomial powers of n .

Our arguments here will adapt a version of the techniques from [15] and hence is related to the ideas in [5]. The main idea is to consider a class of subgraphs (called “cores”) which are present with high probability in G when conditioned to have k_n triangles. Once such a subgraph is planted in G , the expectation of Q is almost as much as the required number of k_n . As in [15], before defining a “core” we define a relaxed version called a “seed”.

Let $w_n = 1/\log n$ in the sequel. In principle, w_n can be taken to be anything that is $1/\text{poly log } n$, but we fix it so for concreteness.

Definition 15 (Seed). Let $S \subseteq K_n$ be a (labeled) graph on the vertex set V with $|V| = n$. We call S a seed if and only if:

1. $\mathbb{E}_S Q \geq (1 - w_n)k_n$,

$$2. e(S) \leq C_s w_n^{-1} k_n^{2/q} \log(1/p_n),$$

where $e(S)$ is the number of edges in S . Here $\mathbb{E}_S Q = \mathbb{E}[Q \mid S \subseteq K_n]$ is the conditional expectation of the number of copies of H in G given that the edges in S are present in G , and C_s is a constant depending on H that will be fixed later.

Following the arguments of [15] one can deduce the following.

Lemma 16 (Must have a seed). *There is a sequence $\xi_n \rightarrow 0$ such that*

$$\mathbb{P}(Q \geq k_n) \leq (1 + \xi_n) \mathbb{P}(G \text{ has a seed})$$

where $G \sim \mathcal{G}(n, p_n)$.

Proof. For a subgraph $S \subseteq K_n$, set $Z_S \stackrel{\text{def}}{=} \mathbf{1}_{\{G \cap S \text{ has no seed}\}}$ (we will also let $Z \stackrel{\text{def}}{=} Z_{K_n}$) and for any H , a subset of edges of K_n forming an isomorphic copy of H , denote $Y_H \stackrel{\text{def}}{=} \mathbf{1}_{\{\text{all edges in } H \text{ are present in } G\}}$. Then observe that

$$\mathbb{P}(Q \geq k_n, G \text{ has no seed}) = \mathbb{P}(QZ \geq k_n) \leq \frac{\mathbb{E}[Q^\ell Z]}{k_n^\ell}, \quad (7)$$

for any $\ell \geq 1$ (we will choose ℓ later). Since $Q = \sum_H Y_H$ (sum over all possible edge subsets of K_n which are isomorphic to H).

$$\begin{aligned} \mathbb{E}[Q^\ell Z] &= \sum_{H_1, \dots, H_\ell} \mathbb{E}[Y_{H_1} \cdots Y_{H_\ell} \cdot Z] \\ &\leq \sum_{H_1, \dots, H_\ell} \mathbb{E}[Y_{H_1} \cdots Y_{H_\ell} \cdot Z_{H_1 \cup \dots \cup H_{\ell-1}}], \quad \text{because } Z_S \geq Z_T \text{ if } S \subseteq T, \\ &= \sum_{H_1, \dots, H_\ell} \mathbb{P}(H_1, \dots, H_\ell \text{ are present}, H_1 \cup \dots \cup H_{\ell-1} \text{ has no seed}), \end{aligned}$$

But $H_1 \cup \dots \cup H_{\ell-1}$ having (or not having) a seed is a deterministic fact, so we can reduce the sum over all collections of H_i where this holds. We can group the terms in the sum above by $H_1, \dots, H_{\ell-1}$ to obtain

$$\begin{aligned} &= \sum_{H_1, \dots, H_{\ell-1} \text{ with no seed}} \mathbb{P}(H_1, \dots, H_{\ell-1} \text{ are present}) \sum_{H_\ell} \mathbb{E}[Y_{H_\ell} \mid Y_{H_i} = 1, \forall i \leq \ell-1] \\ &= \sum_{H_1, \dots, H_{\ell-1} \text{ with no seed}} \mathbb{E}[Y_{H_1} \cdots Y_{H_{\ell-1}} Z_{H_1 \cup \dots \cup H_{\ell-1}}] \underbrace{\mathbb{E}_{H_1 \cup \dots \cup H_{\ell-1}} Q}_{\leq (1-w_n)k_n}, \end{aligned}$$

where we use the fact that $H_1 \cup \dots \cup H_{\ell-1}$ has no seed, and in particular, is not a seed itself. But we can only apply this reasoning if the total number of edges in $H_1 \cup \dots \cup H_{\ell-1}$ is at most $C w_n^{-1} k_n^{2/q} \log(1/p_n)$ (see Definition 15) which happens at least as long as

$$\ell \leq C_s q^{-2} w_n^{-1} k_n^{2/q} \log(1/p_n), \quad (8)$$

because $H_1 \cup \dots \cup H_{\ell-1}$ has at most $q^2 \ell$ edges (and thus the only reason it is not a seed is that the conditional expectation condition is violated). Observe that the bound obtained for $\mathbb{E}[Q^\ell Z]$ has a recursive structure, which we can iterate in ℓ to obtain

$$\mathbb{E}[Q^\ell Z] \leq (1 - w_n)^\ell k_n^\ell,$$

giving us the eventual upper bound via (7)

$$\mathbb{P}(Q \geq k_n, G \text{ has no seed}) \leq (1 - w_n)^\ell$$

for all ℓ satisfying the bound in (8). Now observe that a clique of size $ck_n^{1/q}$ has at least k_n copies of H (where $c = c(H)$ is a constant), and it has $Ck_n^{2/q}$ many edges. Therefore by choosing C_s in the definition of a seed large enough, we can make this clique a seed. The probability of its occurrence is

$$\mathbb{P}(ck_n^{1/q} \text{ clique in } G) \geq \exp(-Ck_n^{2/q} \log n)$$

(recall that various occurrences of C can mean different H -dependent constants). Choose $\ell = \lfloor C_s q^{-2} w_n^{-1} k_n^{2/q} \log(1/p_n) \rfloor$ to get:

$$\begin{aligned} \mathbb{P}(Q \geq k_n, G \text{ has no seed}) &\leq (1 - w_n)^\ell \\ &\leq \exp(-Cq^{-2} k_n^{2/q} \log(1/p_n)) \\ &\leq \xi_n \mathbb{P}(ck_n^{1/q} \text{ clique}) \end{aligned}$$

where $\xi_n \rightarrow 0$ as long as C_s/q^2 is large enough. But as we saw above, this is a seed, so we can replace the bound above by

$$\leq \xi_n \mathbb{P}(G_n \text{ has seed}),$$

which finishes the proof. \square

The previous lemma reduces our task to upper bounding $\mathbb{P}(G \text{ has a seed})$. This is essentially achieved via union-bounding over all possible seeds. But the number of possible seeds is extremely high, so we instead consider a more restrictive structure called “cores”, as defined below. Each seed contains a core, so we have a version of the preceding lemma with $\{G \text{ has a seed}\}$ replaced by $\{G \text{ has a core}\}$.

Definition 17 (Core). Let $G^* \subseteq K_n$ be a (labeled) graph on a vertex set V with $|V| = n$. We call it a core if it has no isolated vertices and,

1. $\mathbb{E}_{G^*} Q \geq (1 - 2w_n)k_n$,
2. $e(G^*) \leq C_s w_n^{-1} k_n^{2/q} \log(1/p_n)$,
3. $\mathbb{E}_{G^*} Q - \mathbb{E}_{G^* - f} Q \geq t_n \stackrel{\text{def}}{=} \frac{w_n^2 k_n^{(q-2)/q}}{C_s \log(1/p_n)} \left(= \tilde{\Omega}(k_n^{(q-2)/q}) \right)$, for all edges $f \in E(G^*)$,

that is, it is a (slightly weaker) seed, but now each edge has to contribute significantly.

As indicated earlier, every seed contains a core:

Lemma 18. *If G is a seed, then there is a subset $G^* \subseteq G$ which is a core.*

Proof. We iteratively delete edges violating (3). Removing such an edge does not reduce $\mathbb{E}_G Q$ by more than t_n . Therefore over all $e(G)$ edges, the maximum possible decrease is $t_n e(G) \leq w_n k_n$, so that the reduced graph still satisfies (1). If we cannot remove an edge, it is already a core. \square

The strict restrictions on cores allow us to prove a few crucial structural results about them.

Lemma 19 (Cores have many copies of \mathbf{H}). *For a core G^* , let $Q_H(G^*)$ be the number of copies of H in G^* . Then $Q_H(G^*) \geq k_n - o(k_n)$. In particular, via Lemma 7, there is a C such that $e(G^*) \geq Ck_n^{2/q}$.*

Proof. We denote by \mathbf{H} an isomorphic copy of H in K_n , i.e. a subset of edges which are isomorphic to H . Then we know that

$$(1 - 2w_n)k_n \leq \mathbb{E}_{G^*} Q = \sum_{\mathbf{H}} p_n^{|\mathbf{H} - G^*|} \leq Q_H(G^*) + \sum_{|\mathbf{H} - G^*| \geq 1} p_n^{|\mathbf{H} - G^*|},$$

where $\mathbf{H} - G^*$ is the collection of edges in \mathbf{H} that are not in G^* . For a given copy \mathbf{H} in K_n , fix an isomorphism $\phi_{\mathbf{H}}$ mapping $E(H) \rightarrow \mathbf{H}$. Let c be the number of edges in $\mathbf{H} - G^*$ with both endpoints in $V(G^*)$, b be the number of edges in $\mathbf{H} - G^*$ with one endpoint in $V(K_n) - V(G^*)$ and other in $V(G^*)$, and let x be the number of vertices in \mathbf{H} outside G^* . Also denote by $X \subseteq H$ the subgraph which is mapped (under $\phi_{\mathbf{H}}$) to outside G^* (so that $x = |V(X)|$), and let $Y \subseteq H$ be the subgraph of H whose vertices are mapped (under $\phi_{\mathbf{H}}$) to G^* and whose edges are mapped to $E(G^*)$ (so that $V(X) \cup V(Y) = V(H)$ but $E(X) \cup E(Y)$ is not $E(H)$; the edges counted in b and c are the ones in $E(H) - (E(X) \cup E(Y))$). Observe that Y may contain isolated vertices. Denote by $y = |V(Y)|$. See Figure 2 for an illustration of these definitions. We will now upper-bound the contribution of all copies \mathbf{H} which satisfy the configuration specified by (b, c, X, Y) . Observe that

$$\begin{aligned} x + y &= q, \\ |E(Y)| + |E(X)| + b + c &= E(H) = q\Delta/2, \\ 2|E(X)| + b &= \Delta|V(X)| = x\Delta. \end{aligned}$$

Solving these equations we get

$$\begin{aligned} |E(X)| &= \frac{x\Delta - b}{2}, \\ |E(Y)| &= \frac{q\Delta}{2} - b - c - \frac{x\Delta}{2} + \frac{b}{2} = \frac{q\Delta - x\Delta - b - 2c}{2}. \end{aligned}$$

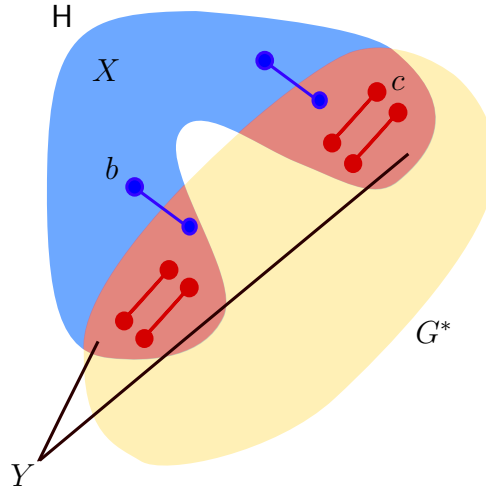


Figure 2: Variables in the proof of Lemma 19.

Since $b + |E(X)| + c$ edges come from outside G^* , for each such H , $p_n^{|H-G^*|}$ is at most

$$Cn^{-\frac{2}{\Delta} \cdot (b + |E(X)| + c)} = Cn^{-\frac{2}{\Delta} \cdot \frac{x\Delta + b + 2c}{2}} = Cn^{-x - \frac{b}{\Delta} - \frac{2c}{\Delta}}.$$

Now we need to count the number of such copies of H . Firstly, the image of X can be chosen in (at most) n^x ways, which can be included in the “cost” to obtain $n^{-\frac{b}{\Delta} - \frac{2c}{\Delta}}$. The final cost will be the product of this with the number of copies of Y in G^* (fixing it will fix all the vertices, and therefore all the edges counted in b and c).

To count the number of copies of Y in G^* , observe that Y being a subgraph of H has max-degree Δ . So we can apply Lemma 7 to obtain that the count is at most

$$Ce^{|E(Y)|/\Delta} v^{y-2|E(Y)|/\Delta}$$

where $v = v(G^*)$ and $e = e(G^*)$. Putting everything together, the total cost is at most

$$Ce^{\frac{q-x}{2} - \frac{b+2c}{2\Delta}} v^{q-x-2\frac{q-x}{2} + \frac{b+2c}{\Delta}} n^{-\frac{b}{\Delta} - \frac{2c}{\Delta}} = Ce^{\frac{q-x}{2} - \frac{b+2c}{2\Delta}} v^{\frac{b+2c}{\Delta}} n^{-\frac{b+2c}{\Delta}} \leq Ce^{\frac{q}{2} - \frac{x+b+2c}{2\Delta}}$$

as $v \leq n$ and $\Delta \geq 1$. If $b + 2c + x \geq 1$, then this is $\tilde{o}(k_n)$ because $e = \tilde{O}(k_n^{2/q})$ (and $k_n \gg \text{poly log } n$ as assumed at the beginning of this section). But this is indeed the case since H is connected and $|H - G^*| \geq 1$, implying that the contribution of the part of the sum corresponding to the configuration (b, c, X, Y) is $\tilde{o}(k_n)$.

Considering all possible distinct configurations of (b, c, X, Y) (of which there are $C = C(H)$ many), we see that the total contribution of the second term in the first display of the proof is $\tilde{o}(k_n)$. This concludes the proof. \square

As indicated in the idea of proof section, the following “product of degrees” lemma is a key combinatorial input in our proof, derived via an application of Finner’s inequality.

Lemma 20 (Product of degrees). *There is a function $g(n) = \tilde{\Omega}(k_n^{1/q})$, depending only on H , such that the following holds for all sufficiently large n . Suppose G^* is a core. For*

every edge $f = (a, b) \in E(G^*)$, we have $d_a d_b \geq g^2(n)$ and $d_a, d_b \geq \Delta$ (here d_a, d_b are the degrees of a, b respectively in G^*). Therefore every vertex in a core has degree at least Δ since none of them are isolated.

Every vertex which is a part of some copy of H in G^* must have degree at least Δ . But the lemma above does not claim that every vertex in G^* is a part of some copy of H , just the weaker statement that every vertex has degree at least Δ . Specifically, it does not rule out the case of a vertex which has a very large degree in G^* but is not part of any copy of H in G^* .

Proof. This proof is a direct application of Lemma 9. Observe that

$$\mathbb{E}_{G^*} Q - \mathbb{E}_{G^* - f} Q = \sum_{\mathbf{H}} p_n^{|\mathbf{H} - G^*|} - \sum_{\mathbf{H}} p_n^{|\mathbf{H} - (G^* - f)|}$$

where $|\mathbf{H} - G^*|$ is again a shorthand for the number of edges in \mathbf{H} not in G^* . Therefore the above is

$$= \sum_{\mathbf{H}} p_n^{|\mathbf{H} - G^*|} - p_n^{|\mathbf{H} - (G^* - f)|}.$$

Now observe that for any \mathbf{H} which does not contain f , the term is zero, and for any \mathbf{H} containing f , $|\mathbf{H} - (G^* - f)| = 1 + |\mathbf{H} - G^*|$. Therefore, the above is

$$= (1 - p_n) \sum_{\mathbf{H} \ni f} p_n^{|\mathbf{H} - G^*|},$$

where we sum over all \mathbf{H} containing f . This is bounded by

$$\leq \sum_{\mathbf{H} \ni f} p_n^{|\mathbf{H} - G^*|}.$$

This last expression is the expected number of copies of \mathbf{H} containing the edge f given that $G^* \subseteq G$. We may now invoke Lemma 9 to bound the above by:

$$\leq C(H)(d_a + np_n^\Delta)^{\frac{\Delta-1}{\Delta}}(d_b + np_n^\Delta)^{\frac{\Delta-1}{\Delta}}(e + n^2 p_n^\Delta)^{\frac{q}{2}-2+\frac{1}{\Delta}},$$

where $e = e(G^*)$. Recall that $p_n^\Delta = n^{-2}$, so that by modifying the constants, we can write

$$\leq C(H)(d_a d_b)^{\frac{\Delta-1}{\Delta}} e^{\frac{q}{2}-2+\frac{1}{\Delta}}.$$

Since G^* is a core, $e = \tilde{O}(k_n^{2/q})$, so that

$$\leq (d_a d_b)^{\frac{\Delta-1}{\Delta}} \tilde{O}\left(k_n^{1-\frac{4}{q}+\frac{2}{q\Delta}}\right).$$

Definition 17 and the preceding chain of inequalities then imply

$$t_n = \tilde{\Omega} \left(k_n^{1-\frac{2}{q}} \right) \leq \mathbb{E}_{G^*} Q - \mathbb{E}_{G^*-f} Q \leq (d_a d_b)^{\frac{\Delta-1}{\Delta}} \tilde{O} \left(k_n^{1-\frac{4}{q}+\frac{2}{q\Delta}} \right),$$

and therefore,

$$(d_a d_b)^{\frac{\Delta-1}{\Delta}} \geq \tilde{\Omega} \left(k_n^{\frac{2}{q}-\frac{2}{q\Delta}} \right) \iff d_a d_b \geq \tilde{\Omega}(k_n^{2/q}).$$

Observe that this lower bound is uniform over all cores, and just depends on H . We define this function to be $g^2(n)$, completing the proof of the first claim.

For the other part, we apply the second half of Lemma 9. Suppose $d_a < \Delta$. Then (given $G^* \subseteq G$) there is no copy of H in G^* containing f because H is Δ -regular. Then every copy of H in G containing f must have some other edge coming from outside G^* . The number of such copies is bounded above in expectation, using Lemma 9, by $C(H) \max(B_1, B_2, B_3)$ where B_1, B_2, B_3 are as defined in the lemma. Observe that $d_b \leq e(G^*) = \tilde{O}(k_n^{2/q})$, so that we can bound B_1, B_2, B_3 as follows:

$$\begin{aligned} B_1 &\stackrel{\text{def}}{=} (d_a + np^\Delta)^{\frac{\Delta-2}{\Delta}} (d_b + np^\Delta)^{\frac{\Delta-1}{\Delta}} (e + n^2 p^\Delta)^{q/2-2+1/\Delta} p n^{1/\Delta} \\ &\leq \tilde{O} \left((k_n^{2/q})^{\frac{\Delta-1}{\Delta}+\frac{q}{2}-2+\frac{1}{\Delta}} \right) n^{-1/\Delta} = \tilde{O}(k_n^{1-2/q}) n^{-1/\Delta} = \tilde{o}(k_n^{1-2/q}), \\ B_2 &\stackrel{\text{def}}{=} (d_a + np^\Delta)^{\frac{\Delta-1}{\Delta}} (d_b + np^\Delta)^{\frac{\Delta-2}{\Delta}} (e + n^2 p^\Delta)^{q/2-2+1/\Delta} p n^{1/\Delta} \\ &\leq \tilde{O} \left((k_n^{2/q})^{\frac{\Delta-2}{\Delta}+\frac{q}{2}-2+\frac{1}{\Delta}} \right) n^{-1/\Delta} = \tilde{O}(k_n^{1-2/q}) n^{-1/\Delta} = \tilde{o}(k_n^{1-2/q}), \\ B_3 &\stackrel{\text{def}}{=} (d_a + np^\Delta)^{\frac{\Delta-1}{\Delta}} (d_b + np^\Delta)^{\frac{\Delta-1}{\Delta}} (e + n^2 p^\Delta)^{q/2-2} p n^{2/\Delta} \\ &\leq \tilde{O} \left((k_n^{2/q})^{\frac{\Delta-1}{\Delta}+\frac{q}{2}-2} \right) = \tilde{O} \left(k_n^{1-\frac{2}{q}-\frac{2}{q\Delta}} \right) = \tilde{o}(k_n^{1-2/q}), \end{aligned}$$

because k_n grows superpolylogarithmically. Therefore the expectation is bounded by $\tilde{o}(k_n^{1-2/q})$. But, as in the proof of the first claim, this must be at least $t_n = \tilde{\Omega}(k_n^{1-2/q})$. This is a contradiction (for sufficiently large n), proving the claim. \square

The next lemma proves the theorem of this paper in the case of large k_n . As mentioned earlier, combined with Lemma 14, this will complete the proof of the main theorem.

The following is a high-level overview of the main steps involved. The proof proceeds via successively specializing the class of cores whose (collective) probability we have to bound. We bi-partition the vertices of the core into high degree and low degree vertices and using Lemma 20 ensure that edges from low degree vertices can only go to sufficiently high degree vertices. Since there are a relatively small number of high degree vertices, we can fix all of them in $V(K_n) = [n]$ (without incurring too high a cost from the union bound). At this point the case where at least half of the edges in the core comes from the edges within this high-degree set is dealt with using straightforward binomial tail probabilities,

leaving us with the more complicated case when at least half of the edges go between the high and low degree sets. The remainder of the proof assumes that this is the case. Now we attempt to fix all the low degree vertices as well. But the straightforward bound of $e = E(G^*)$ on their count is too lossy, thus necessitating further control. However, by a dyadic decomposition, we can fix all of them except the ones with very low degree. But vertices with extremely low degree can have edges only to vertices with very high degree, again by Lemma 20, whose number is very small allowing us to also fix all the neighbor sets of these low degree vertices (and therefore their degrees as well). However, since each vertex in the core has degree at least Δ (Lemma 20), the number of such low degree vertices cannot exceed $1/\Delta$ times the number of edges between the high and low degree sets. This improved bound by a factor of $1/\Delta$ indeed turns out to be sufficient to balance the cost of fixing the low degree vertices against the probability of these edges actually existing (recall that we have already fixed the neighbor sets for these vertices). A calculation then shows that this works out favorably, finishing the proof.

Lemma 21. *Assume k_n grows fast enough (as described at the beginning of this section). The probability that $\mathcal{G}(n, p_n)$ has a core is at most $C \exp(-C' k_n^{2/q} \log n)$ for constants C, C' depending on H . As a consequence, the probability of $Q \geq k_n$ also satisfies the same bound (up to changing C).*

Proof. Let $g(n) = \tilde{\Omega}(k_n^{1/q})$ be the function in Lemma 20, and suppose $G^* \subseteq K_n$ is a core. Let L be the set of *high-degree vertices* in G^* defined to be those with degree $\geq g(n)$ in G^* , and set $R = V(G^*) - L$. By Lemma 20 every edge in G^* must have an endpoint in L . Also we have $|L| \leq L_{\max}$ where $L_{\max} = \tilde{O}(k_n^{1/q})$ because $e(G^*) = \tilde{O}(k_n^{2/q})$. Note that the implicit poly $\log n$ factor in $\tilde{O}(k_n^{1/q})$ does not depend on G^* .

Now recall that a core has at least $C k_n^{2/q}$ edges as was proved in Lemma 19. For brevity let $e = e(G^*)$. Since all these edges have an endpoint in L , either at least $e/2$ of them have both their endpoints in L or at least $e/2$ of them have one endpoint in L and the other in R . The probability that at least $e/2$ edges come from within L is bounded by the probability that there is a set of size at most L_{\max} containing at least $C k_n^{2/q}$ edges (with a different C). For a fixed set of size at most L_{\max} , the probability it has so many edges is at most

$$\mathbb{P} \left(\text{Bin} \left(\binom{L_{\max}}{2}, p_n \right) \geq C k_n^{2/q} \right) \leq \mathbb{P} \left(\text{Bin}(L_{\max}^2/2, p_n) \geq C k_n^{2/q} \right)$$

which may be upper bounded using Lemma 24 from the Appendix (since $p_n \ll \frac{C k_n^{2/q}}{L_{\max}^2}$) to obtain

$$\leq \exp \left(-C' k_n^{2/q} \log \left(\frac{C k_n^{2/q}}{L_{\max}^2 p_n} \right) \right), \quad C, C' \text{ constants.}$$

Recalling that $L_{\max} = \tilde{O}(k_n^{1/q})$, we can conclude that the above is therefore at most

$$\leq \exp \left(-C k_n^{2/q} \log n \right), \quad \text{because } C k_n^{2/q} / L_{\max}^2 \geq 1/\text{poly } \log n,$$

for sufficiently large n . The total number of ways of picking such a subset of size at most L_{\max} from $V(K_n)$ is at most

$$n^{L_{\max}} \leq \exp(k_n^{1/q} \text{poly log } n),$$

so that the probability that there is a subset of vertices of size at most L_{\max} with at least $e/2$ edges inside it is at most

$$\exp(k_n^{1/q} \text{poly log } n - Ck_n^{2/q} \log n) \leq C' \exp(-Ck_n^{2/q} \log n). \quad (9)$$

Therefore for the class of cores satisfying this condition, we have the desired bound.

In the sequel we will prove a similar bound for cores which do not satisfy this condition and hence have fewer than $e/2$ edges coming from within L so that there are at least $e/2$ edges between L and R . Partition L into $m = O(\log n)$ pieces dyadically depending on the degree, i.e., define the (disjoint) sets

$$L_i \stackrel{\text{def}}{=} \{v \in L : g(n)2^{i-1} \leq d_v < g(n)2^i\}, \quad i = 1, 2, \dots, m$$

where d_v is the degree of v in G^* . Note that in fact, we can choose m such that

$$10\rho(n) < 2^m g(n) \leq 20\rho(n) \quad (10)$$

where $\rho(n)$ is the maximum number of edges in a core as in Definition 17, because this ensures that the sets L_{m+1}, \dots are empty by the bound $d_v \leq e$ for all $v \in G^*$ (observe that this definition does not depend on the choice of core, but only on n , H and k_n). Also define the nested sets:

$$R_i \stackrel{\text{def}}{=} \{v \in R : d_v \geq g(n)2^{-i}\}, \quad i = 1, 2, \dots, m$$

so that $R_1 \subseteq R_2 \subseteq \dots$ (see Figure 3). Invoking Lemma 20, observe that every edge (a, b) between L and R goes from a vertex in L_i to one in R_i , for some i , and that

$$|R_i| \leq \frac{2^{i+1}e}{g(n)}. \quad (11)$$

First we fix m, e and all the $\{e_{ij}\}_{i,j \leq m}$ defined as

$$e_{ij} \stackrel{\text{def}}{=} \text{number of edges between } L_i \text{ and } R_j - R_{j-1}, \quad R_0 \stackrel{\text{def}}{=} \emptyset.$$

The total number of such distinct sequences is at most $\exp(\text{poly log } n)$ because each of these is at most n^2 , and there are $O(\log^2 n)$ numbers to fix. This factor will be negligible while applying a union bound, and hence it suffices to consider the class of cores which satisfy these choices. In the sequel we also use the notation

$$\begin{aligned} e_i &\stackrel{\text{def}}{=} \text{number of edges between } L_i \text{ and } R_i \\ &= \sum_{j=1}^i e_{ij}. \end{aligned}$$

Now fix the sets L_1, \dots, L_m . Since the total number of distinct choices of each L_i is at most

$$n^{L_{\max}} \leq \exp(Ck_n^{1/q} \log n),$$

the total number of ways to choose L_1, \dots, L_m is only

$$\exp(Ck_n^{1/q} \log^2 n) = \exp(o(k_n^{2/q}))$$

so that we can fix these and bound the probability of the resulting subclass of cores, again using the fact that this factor will be negligible in a union bound.

Note that by definition (see (10))

$$2^m \leq C \frac{\rho(n)}{g(n)} \leq \frac{\tilde{O}(k_n^{2/q})}{\tilde{\Omega}(k_n^{1/q})} = \tilde{O}(k_n^{1/q}).$$

Then there is a function $h(n) = o(k_n^{2/q})$ depending only on H such that we can choose $s = \Theta(\log \log n)$ satisfying

$$\sum_{i=1}^{m-s} |R_i| \leq \sum_{i=1}^{m-s} 2^{i+1} \frac{e}{g(n)} \leq 2^{m-s+2} \frac{e}{g(n)} = \frac{\tilde{O}(k_n^{1/q}) \tilde{O}(k_n^{1/q})}{2^s} \leq h(n), \quad (12)$$

and also for all $i > m - s$,

$$|L_i| \leq \frac{Ce}{2^i g(n)} \leq C' \frac{e}{g(n)} \cdot \frac{g(n)}{\rho(n)} \cdot 2^s \leq C' \cdot 2^s \leq \text{poly log } n, \quad C, C' \text{ constants.} \quad (13)$$

Observe that s only depends on n and H , and is fixed beforehand. For this choice of s , we can similarly fix R_1, \dots, R_{m-s} by first fixing the sizes of all the R_i (which is a sequence of length $O(\log n)$ each of value at most n , so that total number of ways is $\exp(\text{poly log } n)$) satisfying the constraint that

$$\sum_{i=1}^{m-s} |R_i| \leq h(n), \quad (\text{same as in (12)})$$

and then choosing these sets as subsets of $[n]$ in

$$\exp\left(\sum_{i=1}^{m-s} |R_i| \log n\right) = \exp(o(k_n^{2/q} \log n)) \text{ ways.}$$

Denote by $L' \stackrel{\text{def}}{=} \cup_{i>m-s} L_i$, $R' \stackrel{\text{def}}{=} R - R_{m-s}$, and e' for the number of edges between L' and R' (see Figure 3). Observe that

$$e' = \sum_{m-s < i, j \leq m} e_{ij}.$$

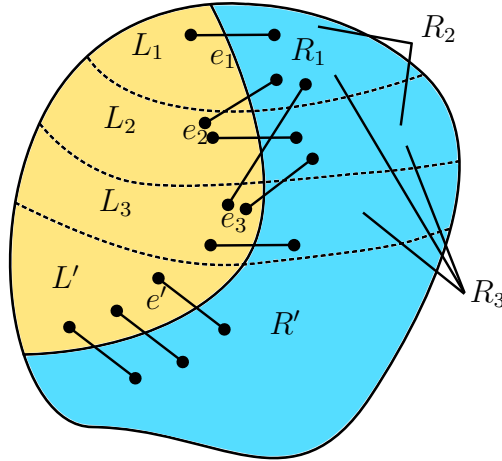


Figure 3: Variables in the proof of Lemma 21 where $m - s = 3$.

Since $|L'| \leq \text{poly log } n$ and $|R_{m-s}| \leq h(n)$, we may exclude the case of at least $e/4$ edges between L' and R_{m-s} by bounding the probability that there is any pair of sets $A, B \subseteq [n]$ of size at most $\text{poly log } n$ and $h(n) \leq o(k_n^{2/q})$ (respectively) such that they have at least $e/4 \geq Ck_n^{2/q}$ edges between them. We may choose the sizes of A and B and then A and B themselves in at most

$$\exp(\text{poly log } n) \binom{n}{\text{poly log } n} \binom{n}{h(n)} \leq \exp(o(k_n^{2/q}) \log n)$$

ways, and after fixing A and B , the probability of there being at least $Ck_n^{2/q}$ edges between them is at most

$$\begin{aligned} \mathbb{P}(\text{Bin}(\text{poly log } n \cdot h(n), p_n) \geq Ck_n^{2/q}) &\leq \exp\left(-Ck_n^{2/q} \log \frac{Ck_n^{2/q}}{\text{poly log } n \cdot h(n) \cdot n^{-2/\Delta}}\right) \\ &\leq \exp(-Ck_n^{2/q} \log n) \end{aligned}$$

using Lemma 24 and the bound $h(n) \leq o(k_n^{2/q})$. Therefore, by a union bound over all A, B , this probability is at most of the same order as our eventual bound, so we may restrict ourselves to the case when the number of edges between L' and R_{m-s} is at most $e/4$. Therefore, since

$$\begin{aligned} &\sum_{i=1}^{m-s} e_i + e' + (\text{number of edges between } L' \text{ and } R_{m-s}) \\ &= (\text{number of edges between } L \text{ and } R) > e/2, \end{aligned}$$

we can restrict ourselves to the case when $\sum_{i=1}^{m-s} e_i + e' > e/4$. We assume this in the sequel.

Note that due to our choices, L' and e' are fixed (since all the L_i and e_{ij} were fixed earlier), but R' depends on the core. However every edge between L and R with an

endpoint in R' must go to L' . Fix the size $r' = |R'|$ (in n ways) and then fix the degrees of the vertices in R' , say $d_1, \dots, d_{r'}$. Since we must ensure $d_1 + \dots + d_{r'} = e'$, and each degree $d_i \leq |L'| \leq \text{poly log } n$, the number of ways to choose these is at most

$$(\text{poly log } n)^{e'} \leq \exp(ce' \log \log n),$$

where c is an absolute constant. Being in a core, vertices in R' must have degree at least Δ (due to Lemma 20). Now we also fix the neighbor sets S_i (with $|S_i| = d_i$) of each these vertices, the number of ways of which is at most

$$|L'|^{\sum_{i=1}^{r'} d_i} \leq \exp(e' \log |L'|) \leq \exp(ce' \log \log n).$$

Therefore, we can fix $r', d_1, \dots, d_{r'}, S_1, \dots, S_{r'}$ in a total of

$$n \exp(ce' \log \log n)$$

ways. We will prove a final probability bound of $\exp(-ce \log n)$ on this class of cores, so we can ignore these factors (observe that this bound is sufficient because $e \geq Ck_n^{2/q}$ as claimed earlier).

For the final step of the proof, we first summarize the information we have now. For each distinct class of cores, we know

- $m = O(\log n)$, the number of dyadic pieces,
- e , the number of edges in the core,
- L_1, \dots, L_m , all the pieces in the high-degree set,
- R_1, \dots, R_{m-s} , all except the last s pieces in the low-degree set,
- e_1, \dots, e_m , the number of edges between every L_i and R_i (in fact we know all the e_{ij}),
- $\sum_{i=1}^{m-s} e_i + e' \geq e/4$. Define $e'' = \sum_{i=1}^{m-s} e_i$.
- r' , the size of R' .
- $d_1, \dots, d_{r'}$, the degrees of vertices in R' .
- and $S_1, \dots, S_{r'}$, their associated neighbor sets (which are known to be subsets of L').

with a total count of $\exp\left(o(k_n^{2/q} \log n)\right)$ many such classes. We now wish to show a total probability bound of $\exp\left(-Ck_n^{2/q} \log n\right)$ on each such class of cores. Note that this is sufficient to complete the proof. For the rest of this argument, we fix such a class and show this claimed upper bound.

Choose R' and a bijection $\psi : [r'] \rightarrow R'$ in at most $\exp(r' \log n)$ ways. But each degree is at least Δ , therefore $e' \geq r' \Delta \implies r' \leq e'/\Delta$, so that the total number of ways is at most $\exp(\frac{e'}{\Delta} \log n)$. Fixing R' and ψ also fixes all the edges between R' and L' via the information about $S_1, \dots, S_{r'}$ so that the probability of this class is then at most

$$\exp(-e' \log(1/p_n)) \prod_{i=1}^{m-s} \mathbb{P}(\text{Bin}(|L_i||R_i|, p_n) \geq e_i),$$

where the second factor accounts for e_i edges between L_i and R_i . Multiplying the number of ways to fix R' and ψ we get

$$\exp\left(\frac{e'}{\Delta} \log n - \frac{2}{\Delta} e' \log n\right) \prod_{i=1}^{m-s} \mathbb{P}(\text{Bin}(|L_i||R_i|, p_n) \geq e_i),$$

where we substitute $p_n = n^{-2/\Delta}$. If $e' \geq e/8$, we are done, because the first factor itself is at most $\exp(-ce \log n) \leq \exp(-Ck_n^{2/q} \log n)$. If not, we must have $e'' \stackrel{\text{def}}{=} \sum_{i=1}^{m-s} e_i \geq e/8$. In that case the first term is at most 1 and the second term is at most

$$\mathbb{P}\left(\text{Bin}\left(\sum_{i=1}^{m-s} |L_i||R_i|, p_n\right) \geq e/8\right) \leq \mathbb{P}\left(\text{Bin}\left(\tilde{O}(k_n^{2/q}), p_n\right) \geq e/8\right)$$

because from (11) and (13) we have $|L_i||R_i| \leq C \frac{e^2}{g(n)^2} = \tilde{O}(k_n^{2/q})$ and $m = O(\log n)$. Applying Lemma 24 we bound the above by

$$\leq \exp\left(-Ce \log\left(\frac{e}{8\tilde{O}(k_n^{2/q})n^{-2/\Delta}}\right)\right)$$

using the fact that $p_n = n^{-2/\Delta} \ll e/\tilde{O}(k_n^{2/q}) = \tilde{\Omega}(1)$. So the quantity inside the log is $\tilde{\Omega}(n^{2/\Delta})$ yielding the bound,

$$\leq \exp(-Ce \log n)$$

for a constant C , as required. \square

Combining the results of Lemma 14 and Lemma 21 finishes the proof of our main theorem, Theorem 1.

7 Appendix

This appendix collects useful lemmas that were required in the proofs above.

Lemma 22. *Let $p > 1$ be a real. Then for any sequence of nonnegative reals $(x_i)_{i=1}^n$ we have $\sum_i x_i^{1/p} \geq (\sum_i x_i)^{1/p}$.*

Proof. Let $s = \sum x_i$, and $y_i = x_i/s$ for $i \in [n]$. Then $y_i \leq 1$, so $y_i^{1/p} \geq y_i$. Therefore $\sum_i y_i^{1/p} \geq \sum_i y_i = 1$. Substituting back the values of y_i , we obtain the result. \square

Lemma 23. *For any $k \geq 2$ and $A > 0$,*

$$\min_{0 \leq s \leq k} (s \log s + A(k-s)^{2/q}) \geq c \min(k \log k, Ak^{2/q}),$$

where c is an absolute constant.

Proof. Let s_* be the optimal value of s . If $s_* \geq k/2$, then the above quantity is at least

$$\frac{1}{2}k \log(k/2) \geq \frac{1}{10}k \log k,$$

as long as $k \geq 3$. If not, $k - s_* \geq k/2$ in which case the result is clear. Finally if $k = 2$, the result can be verified manually. \square

Next we prove an estimate about the tails of the binomial distribution. Similar results are already present in the literature, for instance in [20, Lemma 3.3], but we include a short proof for completeness.

Lemma 24 (Binomial tails). *Suppose $N = N(n), M = M(n) \rightarrow \infty$ and $p_n = o(M/N)$. Then*

$$\mathbb{P}(\text{Bin}(N, p_n) \geq M) \leq \exp \left(-(1 - o(1))M \log \left(\frac{M}{Np_n} \right) \right).$$

Proof. It follows from usual Chernoff bounds that

$$\mathbb{P}(\text{Bin}(N, p_n) \geq M) \leq \exp(-NH_{p_n}(M/N)),$$

where $H_p(t) = D_{\text{KL}}(\text{Ber}(t) \parallel \text{Ber}(p))$, so it suffices to show that $H_{p_n} \left(\frac{M}{N} \right) \geq (1 - o(1)) \frac{M}{N} \log \left(\frac{M}{Np_n} \right)$, which can be reduced to proving that for $p \ll x < 1$, we have

$$H_p(x) \geq (1 - o(1))x \log(x/p).$$

It is indeed well-known that $H_p(x) \sim x \log(x/p)$ in this regime, see for instance [20, Lemma 3.3]. However, our weaker variant admits a simple argument, which we describe next. Observe that

$$H_p(x) = x \log\left(\frac{x}{p}\right) + \underbrace{(1-x) \log\left(\frac{1-x}{1-p}\right)}_{<0},$$

so it suffices to show that

$$(1-x) \log \frac{1-p}{1-x} = o(1) \cdot x \log \frac{x}{p} \iff \frac{1-x}{x} \frac{\log \frac{1-p}{1-x}}{\log \frac{x}{p}} \rightarrow 0 \iff \frac{1-x}{x} \frac{\log\left(1 + \frac{x-p}{1-x}\right)}{\log \frac{x}{p}} \rightarrow 0.$$

Now $\log(1+x) \leq x$, so this is implied by

$$\frac{1-x}{x} \frac{x-p}{1-x} \log^{-1}(x/p) \rightarrow 0$$

which is clear because $x-p \leq x$ and $x/p \rightarrow \infty$. □

Acknowledgements

I would like to thank my advisor Prof. Shirshendu Ganguly for suggesting the problem and for numerous discussions and helpful insights. I also had several interesting and informative discussions with Ella Hiesmayr during the initial stages of this project.

References

- [1] Fanny Augeri. Nonlinear large deviation bounds with applications to Wigner matrices and sparse Erdős-Rényi graphs. *The Annals Of Probability*. **48**, 2404-2448 (2020)
- [2] Anirban Basak and Riddhipratim Basu. Upper tail large deviations of regular subgraph counts in Erdős-Rényi graphs in the full localized regime. *Communications on Pure and Applied Mathematics*, 76(1):3–72, 2023.
- [3] Bhaswar B Bhattacharya, Shirshendu Ganguly, Eyal Lubetzky, and Yufei Zhao. Upper tails and independence polynomials in random graphs. *Advances in Mathematics*, 319:313–347, 2017.
- [4] Béla Bollobás. Threshold functions for small subgraphs. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 90, pages 197–206. Cambridge University Press, 1981.
- [5] Suman Chakraborty, Remco van der Hofstad, and Frank den Hollander. Sparse random graphs with many triangles. [arXiv:2112.06526](https://arxiv.org/abs/2112.06526), 2021.
- [6] Sourav Chatterjee. Large deviations for random graphs. *Lecture Notes in Mathematics*, 2197, 2017.

- [7] Sourav Chatterjee and Amir Dembo. Nonlinear large deviations. *Advances in Mathematics*, 299:396–450, 2016.
- [8] Sourav Chatterjee and SR Srinivasa Varadhan. The large deviation principle for the Erdős–Rényi random graph. *European Journal of Combinatorics*, 32(7):1000–1017, 2011.
- [9] Nicholas Cook and Amir Dembo. Large deviations of subgraph counts for sparse Erdős–Rényi graphs. *Advances in Mathematics*, 373:107289, 2020.
- [10] Nicholas Cook and Amir Dembo. Large deviations of subgraph counts for sparse Erdős–Rényi graphs. *Advances in Mathematics*, 373:107289, 2020.
- [11] Ronen Eldan. Gaussian-width gradient complexity, reverse log-Sobolev inequalities and nonlinear large deviations. *Geometric and Functional Analysis*, 28:1548–1596, 2018.
- [12] Paul Erdos and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci*, 5:17–61, 1960.
- [13] Helmut Finner. A generalization of holder’s inequality and some probability inequalities. *The Annals of probability*, pages 1893–1901, 1992.
- [14] Shirshendu Ganguly, Ella Hiesmayr, and Kyeongsik Nam. Upper tail behavior of the number of triangles in random graphs with constant average degree. *Combinatorica*. **44**, 699–740 (2024)
- [15] Matan Harel, Frank Mousset, and Wojciech Samotij. Upper tails via high moments and entropic stability. *Duke Mathematical Journal*, 1(1):1–104, 2022.
- [16] Svante Janson, Krzysztof Oleszkiewicz, and Andrzej Ruciński. Upper tails for subgraph counts in random graphs. *Israel Journal of Mathematics*, 142:61–92, 2004.
- [17] Svante Janson, Andrzej Rucinski, and Tomasz Luczak. *Random graphs*. John Wiley & Sons, 2011.
- [18] László Lovász. *Large networks and graph limits*, volume 60. American Mathematical Soc., 2012.
- [19] Eyal Lubetzky and Yufei Zhao. On replica symmetry of large deviations in random graphs. *Random Structures & Algorithms*, 47(1):109–146, 2015.
- [20] Eyal Lubetzky and Yufei Zhao. On the variational problem for upper tails in sparse random graphs. *Random Structures & Algorithms*, 50(3):420–436, 2017.
- [21] Andrzej Ruciński. When are small subgraphs of a random graph normally distributed? *Probability Theory and Related Fields*, 78(1):1–10, 1988.
- [22] Klaus Schürger. Limit theorems for complete subgraphs of random graphs. *Periodica Mathematica Hungarica*, 10:47–53, 1979.
- [23] Endre Szemerédi. Regular partitions of graphs. Technical report, Stanford Univ Calif Dept of Computer Science, 1975.
- [24] Jacob Van Den Berg and Harry Kesten. Inequalities with applications to percolation and reliability. *Journal of Applied Probability*, 22(3):556–569, 1985.