THE UPPER TAIL OF CYCLE DISTRIBUTIONS IN SPARSE ERDŐS-RÉNYI RANDOM GRAPHS

UROP+ FINAL PAPER, SUMMER 2022
HOLDEN MUI, TOMASZ ŚLUSARCZYK
MENTORS: ASHWIN SAH, MEHTAAB SAWHNEY
PROJECT SUGGESTED BY: ASHWIN SAH, MEHTAAB SAWHNEY

Abstract. The distribution of subgraph counts in Erdős-Réyni random graphs is of great interest in extremal combinatorics. In this paper, we analyze the upper tail of the k-cycle count in the sparse (constant-degree) regime for fixed k, in the limit as the number of vertices tends to infinity. We provide asymptotically sharp bounds for the probability the graph contains a large number of k-cycles via a coring argument; the logarithmic probability of containing t cycles is asymptotically equal to the logarithmic probability of containing t disjoint k-cycles or containing a clique with t cycles.

1. Introduction

Let $\mathbb{G}(n,p)$ be the Erdős–Rényi random graph on n vertices, where each edge is included independently with probability p. It is a long-standing question to understand various random variables and large deviations of random variables in this setting. For example, the random variable corresponding to the subgraph count of a specific graph H in $G \sim \mathbb{G}(n,p)$ was first studied in [10]. [22] studied this problem for small complete graphs $H = K_s$, and [4] extended this result to other graphs.

In the large subgraph count setting, [7] estimated probabilities of the form $\mathbb{P}[N(H,G) \geq (1+\varepsilon)\mathbb{E}N(H,G)]$ where $G \sim \mathbb{G}(n,p)$; here, $N(H,G)$ denotes the subgraph count of H in G. [19] also studied the structure of graphs conditional on large subgraph counts. [15] famously used concentration inequalities to compute this probability, but these estimates were not tight. Note that these results only dealt with regimes for which $n^{-1}\log n \ll p \ll 1$. However, [13] was able to obtain upper and lower bounds for the logarithmic probability that differed only by a factor of $\log n$. Finally, [12] and [20] were able to pinpoint the logarithmic probability to within $(1+\varepsilon)$ error. Work in this area has progressed for over two decades, and includes other work such as [24], [18], [13], [16], [6], [8], and [9]. Additionally, the lower tail – that is, studying $\mathbb{P}[N(H,G) \leq (1-\varepsilon)\mathbb{E}N(H,G)]$ – has been considered in works such as [25] and [17].

The *sparse regime* occurs when $p = \frac{\lambda}{n}$ for some fixed λ; this is the regime we will be interested in for the entire paper, and is known as the *constant-degree Erdős–Rényi random graph* because the expected degree of any vertex is λ. It is well-known that in this setting, the total number of edges is approximately $\frac{\lambda n}{2}$ and the graph is locally tree-like; for example, see [11]. The first paper to consider large deviations in sparse regimes is [5], which approximated the probability of containing a large number of triangles with the probability of containing a complete graph with the corresponding number of triangles. Our work generalizes their result to k-cycles for any positive integer $k \geq 3$. In this paper, the upper tail of this distribution will be studied in the sparse regime.

Intuitively, the probability that at least t k-cycles will appear in $\mathbb{G}(n,p)$ should be approximately equal to the probability that at least t vertex-disjoint k-cycles appear when t is low, and approximately equal to the probability that a complete graph containing t k-cycles will appear when t is

This work was funded by UROP+.

large. The main result of this paper will provide asymptotically sharp bounds on the probability that $G \sim \mathbb{G}(n,p)$ will contain t k-cycles in the sparse regime.

First, some additional graph-theoretic notation will be needed.

Definition 1.1. Fix an integer $k \geq 3$. Then a k-cycle C_k is a graph on k vertices for which there is a labeling $v_1, \ldots, v_k$ such that $v_i v_j \in E(C_k)$ if and only if $i - j \equiv \pm 1 \pmod{k}$.

Definition 1.2. Fix an integer $k \geq 3$ and let G be a graph. Then the *k-cycle count* $N(C_k, G)$ is the number of distinct unlabeled k-cycles in G.

Definition 1.3. Given a positive real number μ, let $\mathrm{Pois}(\mu)$ denote the Poisson distribution with mean μ; in particular,

$$\mathbb{P}_{X\sim \mathrm{Pois}(\mu)}[X = i] = \frac{e^{-\lambda}\lambda^i}{i!}$$

for all nonnegative integers i.

Theorem 1.4. *Fix an integer $k \geq 3$, a positive real number ε, and a positive real number $\delta \leq 1$. Then there exists a positive integer N, depending only on k, ε, and δ, for which the following statement holds for all $n \geq N$. For every positive real number $p \in (0,1) \cap [\delta/n, \delta^{-1}/n]$ and for any nonnegative integer $t \leq \frac{k!}{2k}\binom{n}{k}$,*

$$\mathbb{P}_{G\sim\mathbb{G}(n,p)}[N(C_k,G) \geq t] \geq \max\left\{(1-\varepsilon)\mathbb{P}_{X\sim \mathrm{Pois}(\lambda^k/(2k))}\left[X \geq t\right], p^{(1+\varepsilon)(2kt)^{2/k}/2}\right\}$$

and

$$\mathbb{P}_{G\sim\mathbb{G}(n,p)}[N(C_k,G) \geq t] \leq \max\left\{(1+\varepsilon)\mathbb{P}_{X\sim \mathrm{Pois}(\lambda^k/(2k))}\left[X \geq t\right], p^{(1-\varepsilon)(2kt)^{2/k}/2}\right\},$$

where λ is the positive real number satisfying $p = \frac{\lambda}{n}$.

Remark 1.5. Observe that the first expression corresponds to the probability of containing t disjoint triangles, and the second term corresponds to containing a complete graph with t k-cycles. In fact, one can characterize the value of t for which the transition from disjoint cycles to complete graph occurs. A short calculation shows that the two arguments in the maximum function are equal when c is approximately $\frac{1}{2k}(k-2)^{k/(k-2)}(\frac{\log n}{\log\log n})^{k/(k-2)}$.

The lower bound for the probability is easy to prove, as all one needs to do is compute the probability that a certain subgraph – either t disjoint k-cycles or a complete graph on $(2kt)^{1/k}$ vertices – will appear. The upper bound is more technical to prove. To summarize, it involves a technique known as *coring*, first introduced in [12]. To get this argument to work, a moments method is used to approximate the probability of containing t k-cycles with the probability a specific kind of subgraph, known as a *seed*, apppears. By coring, one can show that the seed contains an even more specific type of subgraph known as a *core*. Finally, one can compute the probability a core appears and show that this upper bound effectively matches the lower bound. Our work generalizes [12], which deals with 3-cycles (or triangles), to cycles of arbitrary size.

For odd values of k, the idea given in [12] generalizes easily. However, the situation is more difficult when k is even because a minimum degree bound on a vertex in a core cannot be established, so cores cannot be proved to resemble complete graphs (which is the case when k is odd). In fact, when k is even, the minimum degree bound can never work; indeed, $K_{2,\sqrt{2t}}$ (for example) is a core, so it is impossible to even prove a vertex bound of $(2kt)^{1/k}$. To remedy this issue, one can instead prove that any core must separate into connected components, each of which either resembles a complete graph or is bipartite. A cycle-count bound on a lemma similar to a result from [2] can then be used to bound the probability that a core appears.

One interesting observation is that the bipartite graph that maximizes the probability of its appearance matches the probability the complete graph on $(2kt)^{1/k}$ appears for $k = 4$ only. This ultimately does not affect the final bound, however, but does give complications when proving stability results.

1.1. **Notation.** Standard asymptotic notation will be used throughout, a summary of which is given below. For functions $f, g : \mathbb{Z}^+ \to \mathbb{R}$, $f = O(g)$ or $f \lesssim g$ means that there is a constant C such that $|f(n)| \le C|g(n)|$ for all sufficiently large n. Similarly, $f = \Omega(g)$ or $f \gtrsim g$ means that there is a constant $c > 0$ such that $f(n) \ge c|g(n)|$ for all sufficiently large n. Finally, $f \asymp g$ or $f = \Theta(g)$ means that $f \lesssim g$ and $g \lesssim f$, and $f = o(g)$ or $g = \omega(f)$ means that $f(n)/g(n) \to 0$ as $n \to \infty$. Subscripts in asymptotic notation indicate quantities that should be treated as constants.

Standard graph-theoretic notation will also be used. In particular, given a graph G, $V(G)$ is the set of vertices in G and $E(G)$ is the set of edges in G. Additionally, $v(G)$ denotes the number of vertices in G and $e(G)$ denotes the number of edges in G.

Lastly, the following notation for k-cycle counts will be used throughout.

Definition 1.6. Fix an integer $k \ge 3$, let G be a graph, and let uv be an edge of G. Then the *uv-containing k-cycle count* $N_{uv}(C_k, G)$ is the number of distinct unlabeled k-cycles in G containing uv.

Definition 1.7. Fix an integer $k \ge 3$, let G be a graph, and let v be a vertex of G. Then the *v-containing k-cycle count* $N_v(C_k, G)$ is the number of distinct unlabeled k-cycles in G containing v.

Organization. The remainder of this paper is structured as follows. In Section 2 we prove the lower bound in Theorem 1.4. In Section 3 we prove the upper bound.

2. Lower Bound of Theorem 1.4

The lower bound in Theorem 1.4 states that the asymptotic distribution of the number of C_k's in $G(n, p)$ is asymptotically Poisson. This was proven by Bollobás [3] via the method of moments; a version with a precise quantified dependence follows from applying [1, (2.9)] or the proof of [21, Theorem 1].

Lemma 2.1. *Fix a positive integer $k \ge 3$ and a positive real number $\delta \le 1$. Additionally, let $p \in (0, 1) \cap (\delta/n, \delta^{-1}/n)$ and let λ be the positive real number satisfying $p = \frac{\lambda}{n}$. Finally, let $G \sim \mathbb{G}(n, p)$ and $W = N(C_k, G)$ be random variables. Then*

$$d_{\mathrm{TV}}\big(W, \mathrm{Pois}(\lambda^k/2k)\big) \lesssim_{k,\delta} \frac{1}{n}.$$

For convenience, the following claim will be stated.

Claim 2.2. *Fix an integer $k \ge 3$. Then any proper nonempty induced subgraph of C_k has fewer edges than vertices.*

Proof. Any proper subgraph of a cycle is a disjoint union of paths and isolated vertices, and each such connected component contains fewer edges than vertices. □

The following claim asserts that Poisson random variables with similar parameters are close in total variation distance.

Claim 2.3. *Let μ_1 and μ_2 be positive real numbers. Then*

$$d_{\mathrm{TV}}(\mathrm{Pois}(\mu_1), \mathrm{Pois}(\mu_2)) \le |\mu_2 - \mu_1|.$$

Proof. Without loss of generality assume that $\mu_1 \le \mu_2$. Then

$$\begin{aligned} d_{\mathrm{TV}}\big(\mathrm{Pois}(\mu_1), \mathrm{Pois}(\mu_2)\big) &= d_{\mathrm{TV}}\big(\mathrm{Pois}(\mu_1), \mathrm{Pois}(\mu_2 - \mu_1) + \mathrm{Pois}(\mu_1)\big) \\ &\le d_{\mathrm{TV}}\big(\mathrm{Pois}(0), \mathrm{Pois}(\mu_2 - \mu_1)\big) = 1 - e^{-(\mu_2 - \mu_1)} \le \mu_2 - \mu_1. \end{aligned}$$ □

Now, Theorem 1.4 will be proved for $t \le \sqrt{\log n}$.

Claim 2.4. *Fix $k \ge 3$, $\varepsilon, \delta > 0$, and let $np = \lambda \in [\delta, \delta^{-1}]$ and $t \le \sqrt{\log n}$. Then for $n \gtrsim_{k,\delta,\varepsilon} 1$, we have*

$$\mathbb{P}_{G\sim\mathbb{G}(n,p)}\left[N(C_k, G) \ge t\right] = (1 \pm \varepsilon)\mathbb{P}_{X\sim\mathrm{Pois}(\lambda^{2k}/(2k))}\left[X \ge t\right].$$

Proof. Lemma 2.1 immediately gives the desired result noting that for $t \le \sqrt{\log n}$ and $n \gtrsim_{k,\delta,\varepsilon} 1$ we have

$$\mathbb{P}_{X\sim\mathrm{Pois}(\lambda^{2k}/(2k))}\left[X \ge t\right] > \mathbb{P}_{X\sim\mathrm{Pois}(\lambda^{2k}/(2k))}\left[X = \lfloor\sqrt{\log n}\rfloor\right] > n^{-1/2}. \qquad \square$$

Now let $t > \sqrt{\log n}$, $\varepsilon < 1$, and $n \gtrsim_{\varepsilon,\delta,k} 1$ and note that

$$\mathbb{P}_{X\sim\mathrm{Pois}(\lambda^{2k}/(2k))}\left[X = t+1\right]/\mathbb{P}_{X\sim\mathrm{Pois}(\lambda^{2k}/(2k))}\left[X = t\right] = (\lambda^k/(2k))/(t+1) \le \varepsilon/2$$

and therefore

$$\begin{aligned}\mathbb{P}_{X\sim\mathrm{Pois}(\lambda^{2k}/(2k))}\left[X = t\right] \le \mathbb{P}_{X\sim\mathrm{Pois}(\lambda^{2k}/(2k))}\left[X \ge t\right] &\le \mathbb{P}_{X\sim\mathrm{Pois}(\lambda^{2k}/(2k))}\left[X = t\right]\Big(\sum_{i=0}^{\infty}\left(\varepsilon/2\right)^i\Big)\\ &\le (1+\varepsilon)\mathbb{P}_{X\sim\mathrm{Pois}(\lambda^{2k}/(2k))}\left[X = t\right]. \qquad (2.1)\end{aligned}$$

Therefore to prove the lower bound for Theorem 1.4 it suffices to prove the following pair of results (noting the transition between the two expressions as noted in Remark 1.5).

Lemma 2.5. *Fix an integer $k \ge 3$, reals $\varepsilon, \delta > 0$ and suppose $n \gtrsim_{k,\varepsilon,\delta} 1$. Then for $np = \lambda \in [\delta, \delta^{-1}]$ and for any nonnegative integer $t \le n^{1/2k}$ we have*

$$\mathbb{P}_{G\sim\mathbb{G}(n,p)}[N(C_k, G) \ge t] \ge (1-\varepsilon)\mathbb{P}_{X\sim\mathrm{Pois}\left(\frac{\lambda^k}{2k}\right)}[X = t].$$

Lemma 2.6. *Fix an integer $k \ge 3$, reals $\varepsilon, \delta > 0$ and suppose $n \gtrsim_{k,\varepsilon,\delta} 1$. Then for $np = \lambda \in [\delta, \delta^{-1}]$ and for any nonnegative integer $\lceil\sqrt{\log n}\rceil \le t \le \frac{k!}{2k}\binom{n}{k}$ we have*

$$\mathbb{P}_{G\sim\mathbb{G}(n,p)}[N(C_k, G) \ge t] \ge p^{(1+\varepsilon)(2kt)^{2/k}/2}.$$

We first prove Lemma 2.6.

Proof of Lemma 2.6. Note that if there is a clique among the first $\min(\lceil(2kt)^{1/k}\rceil + k, n)$ vertices that there are at least t k-cycles. This occurs with probability at least $p^{(\min(\lceil(2kt)^{1/k}\rceil+k,n))^2/2} \ge p^{(1+\varepsilon)(2kt)^{2/k}/2}$ where we have used $n \gtrsim_{k,\varepsilon,\delta} 1$ in the final inequality. $\square$

The proof of Lemma 2.5 is slightly more involved as we aim to capture the behavior of the tail to within a factor of $1 + o(1)$ within this regime.

Proof of Lemma 2.5. Let $G \sim \mathbb{G}(n,p)$ and $D_{k,t}$ be the disjoint union of l k-cycles.

First we need to estimate the number $M = N(D_{k,t}, K_n)$ of copies of $D_{k,t}$ in K_n. Let us label these copies $S_1, \ldots, S_M$ and note that

$$M = \frac{1}{t!}\prod_{j=0}^{t-1}\left(\frac{k!}{2k}\binom{n-kj}{k}\right) \ge \frac{1}{t!}\left(\frac{(n-kt)^k}{2k}\right)^t = \frac{n^{kt}}{t!(2k)^t}(1-kn^{1/2k-1})^{n^{1/2k}} > (1-\varepsilon)\frac{n^{kt}}{t!(2k)^t} \quad (2.2)$$

when $n \gtrsim_{k,\varepsilon} 1$ and $t \le n^{1/2k}$. Also clearly $\mathbb{P}[S_i \subseteq G] = p^{kt}$ for all $1 \le i \le M$.

Note that the events $\{S_i \subseteq G\} \cap \{N(C_k, G) = t\}_{1\le i\le M}$ are disjoint and therefore it suffices (by symmetry) to prove that

$$\mathbb{P}[S_1 \subseteq G \cap N(C_k, G) = t] \ge (1-\varepsilon)e^{-\lambda^k/(2k)}(\lambda/n)^{kt}.$$

Let X_1 denote the number of k-cycles in G that are disjoint from S_1 and X_2 denote the number of k-cycles in G which are not contained in S_1 but not fully disjoint from S_1. Note that

$$\mathbb{P}[N(C_k, G) = t | S_1 \subseteq G] = \mathbb{P}[X_1 = X_2 = 0 | S_1 \subseteq G].$$

First note that by considering subsets J of S_1 we have

$$\mathbb{E}[X_2 | S_1 \subseteq G] \lesssim_k \sum_{\substack{J \subseteq S_1 \\ 0 < |J| < k}} n^{k - v(J)} (\lambda/n)^{k - e(J)} \lesssim_{k,\delta} n^{-1} |S_1|^k \leq n^{-1/3}$$

where we have used that $e(J) + 1 \leq v(J)$ by Claim 2.2 and $|S_1| \leq n^{1/(2k)}$. Therefore by Markov, we have $\mathbb{P}[X_2 \neq 0 | S_1 \subseteq G] \leq n^{-1/3}$.

Thus it suffices to prove that $\mathbb{P}[X_1 = 0 | S_1 \subseteq G] \geq (1 - \varepsilon) e^{-\lambda^k/(2k)}$. To see this note that

$$\mathbb{P}[X_1 = 0 | S_1 \subseteq G] = \mathbb{P}[X_1 = 0] \geq \mathbb{P}[N(C_k, G) = 0] \geq (1 - \varepsilon) e^{-\lambda^k/(2k)}.$$

The first equality follows as X_1 considers the subset of k-cycles which are disjoint from S_1, the second uses that X_1 is counting only considering a subset of all k-cycles in K_n, and the third follows from Lemma 2.1 and $n \gtrsim_{k,\delta} 1$. □

Corollary 2.7. *Fix an integer $k \geq 3$, reals $\varepsilon, \delta > 0$ and suppose $n \gtrsim_{k,\varepsilon,\delta} 1$. Then for $np = \lambda \in [\delta, \delta^{-1}]$ and any nonnegative integer $t \leq n^{1/2k}$ we have*

$$\mathbb{P}_{G \sim \mathbb{G}(n,p)} \left[D_{k,t} \subset G \right] \geq (1 - \varepsilon) \mathbb{P}_{X \sim \mathrm{Pois}((\lambda^k)(2k))} [X = t].$$

3. Upper bound of Theorem 1.4

The upper bound is more involved than the lower bound; the goal is to show that the probability of containing t k-cycles is roughly equal to the probability either t disjoint k-cycles or a clique with t k-cycles appears.

The proof strategy will consist of four main steps:

(1) bounding the probability that t k-cycles appears with the probability a subgraph containing only its k-cycles, called the *skeleton*, appears.
(2) for each positive integer i, bounding the probability that a connected component of the skeleton contains at least i k-cycles.
(3) for each sequence of nonnegative integers $s_1, s_2, s_3, \ldots$ satisfying $s_1 + 2s_2 + 3s_3 + \ldots \geq t$, bounding the probability that the skeleton contains s_i components with at least i k-cycles for every positive integer i.
(4) summing this probability over all such sequence $s_1, s_2, s_3, \ldots$ and using convexity ideas to show that the resulting sum is approximately equal to the term corresponding to the sequence $t, 0, 0, \ldots$ (where all terms are zero except for $s_1 = t$) or the term corresponding to the sequence $0, 0, \ldots, 0, 1, 0, \ldots$ (where all terms are zero except for $s_t = 1$).

The most difficult of these steps is step (2). The bound depends on the size of i and consists of three main substeps:

(2a) bounding the probability for $i = 1$.
(2b) bounding the probability for $2 \leq i \leq (\log n)^{1/2}$ by the probability for $i = 2$.
(2c) bounding the probability for $i \geq (\log n)^{1/2}$ using a seed/core argument.

The most technical of these substeps is substep (2c); it consists of four main parts:

(i) proving, using a moments argument, that most of the edges in the connected component are contained in a *seed* with high probability.
(ii) proving that every seed must contain a *core*.

(iii) proving that, when the number of edges in the core is small, the probability a core appears is approximately equal to the probability that a clique containing c k-cycles appears.
(iv) proving that, when the number of edges in the core is large, the probability a core appears is negligible compared to the probability that a clique containing c k-cycles appears.

At a high level, a seed is a subgraph that contains most of the k-cycles and few edges, and a core is a subgraph of a seed with its "superfluous" edges removed. It is important to note that the proof of (iv) will depend on the size of c. Additionally, care needs to be taken in step (3) to obtain the desired Poisson bound; it is not sufficient to simply raise the result in (2) for $i = 1$ to the power of s_1.

Definition 3.1. Let $k \geq 3$ be an integer. A k-cycle induced graph (k-CIG) is a graph whose every edge belongs to at least one k-cycle.

Definition 3.2. Given a graph G, define its *k-skeleton* $\mathcal{S}_k(G)$ to be the union of its k-cycles. Define also the *essential skeleton* $\mathcal{S}^*(G)$ of G to be $\mathcal{S}_k(G)$ where every connected component isomorphic to a k-cycle was removed.

That is, $\mathcal{S}_k(G)$ is the maximal k-CIG contained in G.

Before we begin the proof, let us note that thanks to Claim 2.4, (2.1) and decreasing ε, δ if needed it suffices to prove the following result.

Lemma 3.3. *Fix an integer $k \geq 3$, positive reals $\varepsilon, \delta < 1$ and suppose $n \gtrsim_{k,\varepsilon,\delta} 1$. Then for $np = \lambda \in [\delta, \delta^{-1}]$ and for any nonnegative integer $\lceil \sqrt{\log n} \rceil \leq t \leq \frac{k!}{2k}\binom{n}{k}$ we have*

$$\mathbb{P}_{G \sim \mathbb{G}(n,p)}[N(C_k, G) \geq t] \leq \max\left\{(1+\varepsilon)\frac{e^{\frac{-\lambda^k}{2k}}}{t!}\left(\frac{\lambda^k}{2k}\right)^t,\ p^{(1-\varepsilon)(2kt)^{2/k}/2}\right\}.$$

Definition 3.4. Let $l, m \in \mathbb{Z}^+$. $X(l,m)$ will denote that event that for a random graph G, the skeleton $\mathcal{S}_k(G)$ contains at least l k-cycles in connected components which have exactly m k-cycles each. Let also $X_+(l,m) = \bigcup_{m' \geq m} X(l,m)$, $X_-(l,m) = \bigcup_{m' \leq m} X(l,m)$ and $X(l,[m_1,m_2] = \bigcup_{m \in [m_1,m_2]} X(l,m)$. Finally, let $Y(l,m)$ denote the event that the skeleton of G contains exactly l cycles in the connected components with exactly m k-cycles.

We will also require the following well known estimate on binomial coefficients (see e.g. [23, (5.14)]).

Lemma 3.5. *For every pair of nonnegative integers (n,k),*

$$\binom{n}{k} \leq \left(\frac{en}{k}\right)^k.$$

We will also use the following simple graph-theoretic lemma several times.

Lemma 3.6. *Fix an integer $k \geq 3$, and let G be a connected graph that is a union of k-cycles. Then*

$$v(G) \leq \left(1 - \frac{1}{k}\right)e(G) + 1.$$

Moreover if G is not isomorphic to a k-cycle, then

$$v(G) \leq \left(1 - \frac{1}{k^2+k}\right)e(G).$$

Proof. Let the k-cycles be $\mathcal{C}_1, \ldots, \mathcal{C}_l$, ordered such that for every integer j satisfying $2 \leq j \leq c$, there exists an integer i satisfying $1 \leq i < j$ such that $\mathcal{C}_i$ and $\mathcal{C}_j$ share a vertex; this is possible because G is connected. Since $G = \mathcal{C}_1 \cup \ldots \cup \mathcal{C}_c$, it suffices to prove by induction that

$$|V(\mathcal{C}_1 \cup \ldots \cup \mathcal{C}_i)| \leq \left(1 - \frac{1}{k}\right) |E(\mathcal{C}_1 \cup \ldots \cup \mathcal{C}_i)| + 1$$

for every integer $1 \leq i \leq c$. The base case $i = 1$ follows trivially. For the inductive step, assume that the inductive hypothesis holds for some integer $1 \leq i < l$. Note that the subgraph induced by edges in $E(\mathcal{C}_1 \cup \ldots \cup \mathcal{C}_{i+1}) \setminus E(\mathcal{C}_1 \cup \ldots \cup \mathcal{C}_i)$ is either a proper nonempty subgraph of C_k and by Claim 2.2

$$\left|V(\mathcal{C}_1 \cup \ldots \cup \mathcal{C}_{i+1}) \setminus V(\mathcal{C}_1 \cup \ldots \cup \mathcal{C}_i)\right| < \left|E(\mathcal{C}_1 \cup \ldots \cup \mathcal{C}_{i+1}) \setminus E(\mathcal{C}_1 \cup \ldots \cup \mathcal{C}_i)\right| - 1 \tag{3.1}$$

or it is isomorphic to C_k, in which case addition of $\mathcal{C}_{i+1}$ to $\mathcal{C}_1 \cup \ldots \cup C_i$ adds k new edges, but by the assumption on connectivity, it can add at most $k-1$ new vertices, so that (3.1) holds in both cases. Now (3.1) together with $E(\mathcal{C}_1 \cup \ldots \cup \mathcal{C}_{i+1}) \setminus E(\mathcal{C}_1 \cup \ldots \cup \mathcal{C}_i) \leq k$ and the inductive assumption proves that

$$|V(\mathcal{C}_1 \cup \ldots \mathcal{C}_{i+1})| = k = \left(1 - \frac{1}{k}\right) |E(\mathcal{C}_1 \cup \ldots \cup \mathcal{C}_{i+1})| + 1\,.$$

Finally, when G is a k-CIG not isomorphic to a k-cycle, then adding $1 \leq \frac{e(G)}{k+1}$ to the first result allows us to obtain the second one. □

Remark 3.7. We will need later stated implicitly in the proof of Lemma 3.6: edges of any k-CIG H can be partitioned into sets $A_1, \ldots, A_l$, such that for any $i \leq l$, $A_1 \cup \ldots \cup A_i$ is a k-CIG whose every connected component is nonisomorphic to a k-cycle and $|A_i| \leq 2k$.

Proof. When H is connected $A_1 = E(\mathcal{C}_1) \cup E(\mathcal{C}_2)$ and $A_i = E(\mathcal{C}_{i+1}) \setminus \bigcup_{j \leq i} E(\mathcal{C}_j)$ for $i \geq 2$ suffices. When H is not connected, use the same construction component by component. □

We will need the following lemma, which is [12, Lemma 2.5].

Lemma 3.8. *Fix an integer $k \geq 3$, and let G be a graph. Then*

$$e(G) \geq \frac{1}{2}(2kN(C_k, G))^{2/k}.$$

We will also require a bound of similar spirit, but involving $N_{uv}(C_k, G)$ instead of $N(C_k, G)$.

Lemma 3.9. *Fix an integer $k \geq 3$, let G be a graph, and let uv be an edge of G. Then*

$$N_{uv}(C_k, G) \leq (2e(G))^{(k-3)/2} \min(\deg u, \deg v)\,, \qquad N_v(C_k, G) \leq \deg v (2e(G))^{(k-1)/2}$$

if k is odd, and

$$N_{uv}(C_k, G) \leq \min\left((2e(G))^{(k-2)/2}, (2e(G))^{(k-4)/2} \deg u \deg v\right)$$

if k is even.

Proof. If k is odd, then every k-cycle containing uv is determined by the other vertex adjacent to u, along with the locations and orientations of the $\frac{k-3}{2}$ edges that alternately span the remaining $k-3$ vertices in the k-cycle. Therefore, the number of k-cycles containing uv is at most $(2e(G))^{(k-3)/2} \deg u$. Repeating the same reasoning for v gives the desired bound. For $N_v(C_k, G)$ it suffices to note that for a fixed v, k-cycle containing v can be specified by providing $\frac{k-1}{2}$ edges together with their orientations.

If k is even, then every k-cycle containing uv is determined by the locations and orientations of the other $\frac{k-2}{2}$ edges that form every other edge in the k-cycle. Therefore, the number of k-cycles containing uv is at most $(2e(G))^{(k-2)/2}$. Alternatively, every k-cycle containing uv is determined by the two edges adjacent to uv, along with the locations and orientations of the $\frac{k-4}{2}$ edges that alternately span the remaining $k-4$ vertices in the k-cycle. Therefore, the number of k-cycles containing uv is at most $(2e(G))^{(k-4)/2} \deg u \deg v$. □

Claim 3.10. *Let $H \neq C_k$ be a connected graph where every edge belongs to at least one k-cycle. Then $v(H)+1 \leq e(H)$.*

Proof. Such H has at least $k+1$ vertices, so the claim follows from Lemma 3.6. □

We can now start executing step (2).

Lemma 3.11. *Fix an integer $k \geq 3$, reals $\varepsilon, \delta > 0$ and suppose $n \gtrsim_{k,\varepsilon,\delta} 1$ and $np = \lambda \in [\delta, \delta^{-1}]$. Then for any nonnegative integer $t \leq \sqrt{n}$,*

$$\mathbb{P}_{G\sim\mathbb{G}(n,p)}\big[X(t,1)\big] \leq (1+\varepsilon)\frac{e^{-\lambda^k/2k}}{t!}\left(\frac{\lambda^k}{2k}\right)^t$$

and for any nonnegative integer $t \leq \frac{k!}{2k}\binom{n}{k}$,

$$\mathbb{P}_{G\sim\mathbb{G}(n,p)}\big[X(t,1)\big] \leq (1+\varepsilon)\frac{1}{t!}\left(\frac{\lambda^k}{2k}\right)^t .$$

The proof will be similar to the corresponding lower bound, i.e. Lemma 2.5.

Proof. Let us first prove that

$$\mathbb{P}_{G\sim\mathbb{G}(n,p)}\big[Y(t,1)\big] \leq (1+\varepsilon)\frac{e^{-\lambda^k/2k}}{t!}\left(\frac{\lambda^k}{2k}\right)^t \tag{3.2}$$

when $t \leq \sqrt{n}$.

Similarly as in the proof of Lemma 2.5 let $G \sim \mathbb{G}(n,p)$, $D_{k,t}$ be the disjoint union of l k-cycles and $M = N(D_{k,t}, K_n)$ be the number of copies of $D_{k,t}$ in K_n. Let us label these copies $S_1, \dots, S_M$ and note that

$$M = \frac{1}{t!}\prod_{j=0}^{t-1}\left(\frac{k!}{2k}\binom{n-kj}{k}\right) \leq \frac{n^{kt}}{t!(2k)^t} . \tag{3.3}$$

Also clearly $\mathbb{P}[S_i \subseteq G] = p^{kt}$ for all $1 \leq i \leq M$. Thus by the union bound and symmetry it suffices to prove

$$\mathbb{P}[S_1 \subseteq G \cap N(C_k, G) = t] \leq (1+\varepsilon)e^{-\lambda^k/(2k)}(\lambda/n)^{kt}.$$

Let Z denote the number of k-cycles in G that are disjoint from S_1. Note that

$$\mathbb{P}[N(C_k, G) = t | S_1 \subseteq G] \leq \mathbb{P}[Z = 0 | S_1 \subseteq G].$$

First note that by considering subsets J of S_1 we have

Thus it suffices to prove that $\mathbb{P}[Z = 0|S_1 \subseteq G] \leq (1+\varepsilon)e^{-\lambda^k/(2k)}$. Note that

$$\mathbb{P}[Z=0|S_1 \subseteq G] = \mathbb{P}[Z=0] \leq \mathbb{P}_{G'\sim\mathbb{G}(n-kt,p)}[N(C_k, G) = 0] \leq (1+\varepsilon)e^{-\lambda^k/(2k)} . \tag{3.4}$$

The first equality follows as Z considers the subset of k-cycles which are disjoint from S_1, the second uses that X is counting only considering a subset of all k-cycles in K_n, and the third follows from Lemma 2.1 and $n \gtrsim_{k,\delta} t$. This establishes (3.2). Moreover, note that we needed $t \leq \sqrt{n}$ only to ensure that $n - kt \gtrsim_{k,\varepsilon,\delta} 1$ in the last step to use Lemma 2.1 in (3.4). We can now lift this assumption and in (3.4) use instead a trivial bound $\mathbb{P}[Z = 0] \leq 1$ to obtain

$$\mathbb{P}_{G\sim\mathbb{G}(n,p)}\big[Y(t,1)\big] \leq (1+\varepsilon)\frac{1}{t!}\left(\frac{\lambda^k}{2k}\right)^t \tag{3.5}$$

assuming only $t \leq \frac{k!}{2k}\binom{n}{k}$.

Now note that with $t \geq \sqrt{\log n} \gtrsim_{k,\varepsilon,\delta} 1$, we can assume that $\frac{\lambda^k}{2kt} < \varepsilon/2$. Then we can conclude the statement of this lemma by simple summation based on (3.2):

$$\mathbb{P}_{G\sim\mathbb{G}(n,p)}\big[X(t,1)\big] \leq \sum_{t'=t}^{\infty}(1+\varepsilon)\frac{e^{-\lambda^k/2k}}{t!}\left(\frac{\lambda^k}{2k}\right)^t$$
$$\leq (1+\varepsilon)\frac{e^{-\lambda^k/2k}}{t!}\left(\frac{\lambda^k}{2k}\right)^t\sum_{l=0}^{\infty}(\varepsilon/2)^l < (1+3\varepsilon)\frac{e^{-\lambda^k/2k}}{t!}\left(\frac{\lambda^k}{2k}\right)^t$$

assuming that $\varepsilon < 1$. Replacing ε with $\varepsilon/3$ finishes the proof for $t \leq \sqrt{n}$. Running the same summation based on (3.5) instead of (3.2) yields the result for $t \leq \frac{k!}{2k}\binom{n}{k}$ and finishes the proof. □

Now continue to the step (2b).

Lemma 3.12. *Fix an integer $k \geq 3$, reals $\varepsilon, \delta > 0$ and suppose $n \gtrsim_{k,\varepsilon,\delta} 1$. Then for $np = \lambda \in [\delta, \delta^{-1}]$ and any nonnegative integer $\sqrt{\log n} \leq t \leq e^{\sqrt{\log n}/3}$,*

$$\mathbb{P}_{G\sim\mathbb{G}(n,p)}\big[X_-(t,\sqrt{\log n})\big] \leq (1+\varepsilon)\frac{e^{-\lambda^k/2k}}{t!}\left(\frac{\lambda^k}{2k}\right)^t.$$

Proof. Any component of a skeleton of G containing more than one k-cycle has more edges than vertices by Claim 3.10. For a given number $E < k\sqrt{\log n}$ there is at most $\binom{n}{E-1}\binom{(E-1)^2}{E}$ graphs in K_n which are k-CIGs with E edges. Now the probability that $G \sim \mathbb{G}(n,p)$ contains such a graph is

$$p^E\binom{n}{E-1}\binom{(E-1)^2}{E} \leq \delta^{-\sqrt{\log n}}(ek\sqrt{\log n})^{k\sqrt{\log n}}n^{-1} < n^{-1/2}$$

when $n \gtrsim_{k,\delta,\varepsilon} 1$. Now conditional on $X(t_2,[2,\sqrt{\log n}])$, G must contain at least $t_2/\sqrt{\log n}$ disjoint connected k-CIGs with at most $k\sqrt{\log n}$ edges. Note that

$$\mathbb{P}\big[X_-(t,\sqrt{\log n})\big] \leq \sum_{t_1+t_2=t}\mathbb{P}\big[X(t_1,1)\big]\mathbb{P}\big[X(t_2,[2,\sqrt{\log n}])\big] \tag{3.6}$$
$$\leq (1+\varepsilon)\frac{e^{-\lambda^k/2k}}{t!}\left(\frac{\lambda^k}{2k}\right)^t + \sum_{t_2=1}^{t}(1+\varepsilon)\frac{e^{-\lambda^k/2k}}{(t-t_2)!}\left(\frac{\lambda^k}{2k}\right)^{t-t_2}n^{-t_2/2\sqrt{\log n}} \tag{3.7}$$

so it suffices to show that

$$\sum_{t_2=1}^{t}\frac{e^{-\lambda^k/2k}}{(t-t_2)!}\left(\frac{\lambda^k}{2k}\right)^{t-t_2}n^{-t_2/2\sqrt{\log n}} < \varepsilon\frac{e^{-\lambda^k/2k}}{t!}\left(\frac{\lambda^k}{2k}\right)^t \tag{3.8}$$

(which will yield the final result for 4ε instead of ε). Using $(t-t_2)! \leq t!/t^{t_2}$ shows that it suffices to prove

$$\left(\frac{2kt}{\lambda^k}\right)^{t_2} < \frac{\varepsilon}{t}n^{t_2/2\sqrt{\log n}}.$$

This inequality is, however, true for $n \gtrsim_{k,\varepsilon,\delta} 1$ as $t < n^{1/3\sqrt{\log n}}$. □

Let us now introduce a *core* – a central object for our study of upper tails.

Definition 3.13. Fix an integer $k \geq 3$ and positive real $\varepsilon < 1$. Additionally, let t be a positive integer. An *(ε,t)-core* is a graph G without isolated vertices such that

(C1) $N(C_k,G) \geq (1-6\varepsilon)t$,
(C2) $e(G) \leq \varepsilon^{-1}t^{2/k}$,
(C3) $N_{uv}(C_k,G) \geq \varepsilon^2 t^{1-2/k}$ for every edge $uv \in E(G)$,

(C4) G has at most $2\varepsilon^{-(k+2)/(k-2)}$ connected components, and every connected component of G has at least $\frac{1}{2}k^{2/k}\varepsilon^{4/(k-2)}t^{2/k}$ edges, and

(C5) $\deg u \deg v \leq 2^{k+1}\varepsilon^{-(k+4)}t^{2/k}$ for every edge $uv \in E(G)$.

Let also $\mathrm{CORE}(\varepsilon, t, G)$ denote the event that a random graph G contains an (ε, t)-core.

Now the main purpose of these definitions is to show that conditional on $N(G, C_k)$ being large, G contains a core with high probability and then bound the probability of a core appearing in G to obtain Lemma 3.3. The proof of the existence of a core will be different for small and large values of t.

Lemma 3.14. *Fix an integer $k \geq 3$ and reals $\varepsilon, \delta > 0$. Assume $\varepsilon \lesssim_{k,\delta} 1$. Then for $n \gtrsim_{k,\varepsilon,\delta} 1$ the following holds. For $np = \lambda \in [\delta, \delta^{-1}]$ and for any nonnegative integer $t \leq n^{1/4k}$,*

$$\mathbb{P}_{G\sim\mathbb{G}(n,p)}\left[e(\mathcal{S}_k^*(G)) \geq \varepsilon^{-1}t^{2/k}\right] < \varepsilon p^{(1+\varepsilon)(2kt)^{2/k}/2}.$$

Proof. Assume that $e(\mathcal{S}_k^*(G)) \geq \varepsilon^{-1}t^{2/k}$. By Remark 3.7 we can select a subgraph $H \subseteq \mathcal{S}_k^*(G)$ which is a k-CIG, none of its connected components is isomorphic to a k-cycle and $\varepsilon t^{2/k} \leq e(H) < \varepsilon t^{2/k} + 2k$. Note that Lemma 3.6 applied to every connected component of H yields

$$v(H) \leq \left(1 - \frac{1}{k^2+k}\right) e(H).$$

Hence for a given E, there is at most

$$\binom{n}{\left(1-\frac{1}{k^2+k}\right)E}\binom{E^2/2}{E} \leq n^{E(1-1/(k^2+k))}(2E)^E$$

possible candidates for H with $e(H) = E$ in a K_n (where we have used Lemma 3.5). Let $\mathcal{H}$ be the family of these candidates. Now

$$\begin{aligned}
\mathbb{P}_{G\sim\mathbb{G}(n,p)}\left[e(\mathcal{S}_k^*(G)) \geq \varepsilon^{-1}t^{2/k}\right] &\leq \sum_{H\in\mathcal{H}} \mathbb{P}_{G\sim\mathbb{G}(n,p)}(H \subseteq G) \\
&\leq \sum_{E=\lceil\varepsilon^{-1}t^{2/k}\rceil}^{\lceil\varepsilon^{-1}t^{2/k}\rceil+2k-1} p^E n^{E(1-1/(k^2+k))}(2E)^E \\
&\leq \sum_{E=\lceil\varepsilon^{-1}t^{2/k}\rceil}^{\lceil\varepsilon^{-1}t^{2/k}\rceil+2k-1} \delta^{-E} n^{-E/(k^2+k)}(2E)^E \\
&\leq \varepsilon p^{(1+\varepsilon)(2kt)^{2/k}/2}
\end{aligned}$$

where we used $1 \lesssim_{k,\varepsilon,\delta} \sqrt{\log n}$, $t \leq n^{1/4k}$ and $\varepsilon \lesssim_k 1$. □

Lemma 3.15. *Fix an integer $k \geq 3$ and a positive real number $\varepsilon \lesssim_k 1$. Additionally, let t be a positive integer. Then any graph with t k-cycles and at most $\varepsilon^{-1}t^{2/k}$ edges contains a (ε, t)-core with at least $(1 - 3\varepsilon)t$ edges as a subgraph.*

Proof. If $\varepsilon \geq 1/6$, the empty graph is an (ε, t)-core. Assume thus that $\varepsilon < 1/6$.

Let G be a graph with t k-cycles and at most $\varepsilon^{-1}t^{2/k}$ edges. The idea is to successively remove edges from G according to the following algorithm and show that the resulting graph is a core.

(1) If k is even: remove all edges $uv \in E(G)$ for which $\deg u \deg v > 2^{k+1}\varepsilon^{-(k+1)}t^{2/k}$ to get a graph G_0.
If k is odd: remove all vertices $v \in V(G)$ such that $\deg v > 2^{(k+1)/2}\varepsilon^{-(k+1)/2}t^{1/k}$ to get a graph G_0.

(2) Define a sequence of graphs $G_0, G_1, \dots$ by repeatedly setting $G_{i+1} = G_i \setminus \{u_i v_i\}$ for some edge $u_i v_i \in G_i$ satisfying $N_{u_i v_i}(C_k, G_i) < \varepsilon^2 t^{1-2/k}$, if such an edge exists. Let G_s be the last graph in the sequence.
(3) Remove every connected component of G_s with less than $\frac{1}{2}k^{2/k}\varepsilon^{4/(k-2)}t^{2/k}$ edges to get a graph G'.

To show that G' satisfies (C1), it suffices to check that at most εc k-cycles were removed during each step.

(1) Assume first that k is even. By double counting the number of directed paths of length three by the middle edge,

$$\sum_{uv \in E(G)} 2 \deg u \deg v \le (2e(G))^2$$

since each ordered pair of oriented edges in $E(G)$ determines the starting edge and the ending edge of at most one directed path of length three. Therefore, the number of edges $uv \in E(G)$ for which $\deg u \deg v \ge 2^k \varepsilon^{-(k+4)} t^{2/k}$ is at most

$$\frac{\frac{1}{2} \cdot (2e(G))^2}{2^{k+1}\varepsilon^{-(k+1)}t^{2/k}}.$$

Now, every edge $uv \in E(G)$ is part of at most $(2e(G))^{(k-2)/2}$ k-cycles by Lemma 3.9. Thus, removing all edges $uv \in E(G)$ for which $\deg u \deg v \ge 2^{k/2}\varepsilon^{-(k+1)/2}t^{2/k}$ removes at most

$$\begin{aligned}\frac{\frac{1}{2} \cdot (2e(G))^2}{2^{k+1}\varepsilon^{-(k+1)}t^{2/k}} \cdot (2e(G))^{(k-2)/2} &< \frac{\varepsilon^{(k+4)/2} e(G)^{(k+2)/2}}{t^{2/k}} \\ &\le \frac{\varepsilon^{(k+4)/2}\left(\varepsilon^{-1}t^{2/k}\right)^{(k+2)/2}}{t^{2/k}} = \varepsilon t\end{aligned}$$

k-cycles from G, since $e(G) \le \varepsilon^{-1}t^{2/k}$ by hypothesis.
Now assume k is odd. There are at most

$$\frac{\sum_{v \in V(G)} \deg v}{2^{(k+1)/2}\varepsilon^{-(k+1)/2}t^{1/k}} = \frac{2e(G)}{2^{(k+1)/2}\varepsilon^{-(k+1)/2}t^{1/k}}$$

removed vertices, each of which is contained in at most $(2e(G))^{(k-1)/2}$ cycles by lemma Lemma 3.9. Thus removing the specified vertices removes at most

$$\begin{aligned}\frac{2e(G)}{2^{(k+1)/2}\varepsilon^{-(k+1)/2}t^{1/k}} \cdot (2e(G))^{(k-1)/2} &= \frac{\varepsilon^{-(k+1)/2}e(G)^{(k+1)/2}}{t^{1/k}} \\ &\le \frac{\varepsilon^{-(k+1)/2}(\varepsilon^{-1}t^{2/k})^{(k+1)/2}}{t^{1/k}} \\ &= \varepsilon t\end{aligned}$$

k-cycles as desired.

(2) By construction, $N_{u_i v_i}(C_k, G_i) \le \varepsilon^2 t^{1-2/k}$ for each integer i satisfying $0 \le i < s$. Hence,

$$\begin{aligned}N(C_k, G_s) - N(C_k, G_0) &= \sum_{i=0}^{s-1} N(C_k, G_{i+1}) - N(C_k, G_i) \\ &= \sum_{i=0}^{s-1} N_{u_i v_i}(C_k, G_i) \\ &< \sum_{i=0}^{s-1} \varepsilon^2 t^{1-2/k}\end{aligned}$$

$$
\begin{aligned}
&= s\varepsilon^2 t^{1-2/k} \\
&\leq \left(\varepsilon^{-1} t^{2/k}\right)\left(\varepsilon^2 t^{1-2/k}\right) \\
&= \varepsilon t,
\end{aligned}
$$

since $s \leq |E(G_0)| \leq \varepsilon^{-1} t^{2/k}$ by hypothesis. Thus, at most εt k-cycles can be removed during this step.

(3) For any connected component H in G_s and any edge $uv \in E(H)$,

$$\varepsilon^2 t^{1-2/k} \leq N_{uv}(C_k, H) \leq (2|E(H)|)^{(k-2)/2}$$

where the first inequality follows by construction from step (2) and the second inequality follows from Lemma 3.9. Therefore,

$$|E(H)| \geq \frac{1}{2}\left(\varepsilon^2 t^{1-2/k}\right)^{2/(k-2)} = \frac{1}{2}\varepsilon^{4/(k-2)} t^{2/k}$$

for every connected component H in G_s containing at least one edge. Thus, the number of connected components in G_s containing at least one edge is at most

$$\frac{|E(G_s)|}{\frac{1}{2}\varepsilon^{4/(k-2)} t^{2/k}} \leq \frac{\varepsilon^{-1} t^{2/k}}{\frac{1}{2}\varepsilon^{4/(k-2)} t^{2/k}} = 2\varepsilon^{-(k+2)/(k-2)}$$

since $|E(G_s)| \leq \varepsilon^{-1} t^{2/k}$ by hypothesis. Hence, removing all the connected components in G_s with at fewer than $\frac{1}{2}k^{2/k}\varepsilon^{4/(k-2)} t^{2/k}$ edges removes at most

$$2\varepsilon^{-(k+2)/(k-2)} \cdot \frac{1}{2k}\left(2 \cdot \frac{1}{2}k^{2/k}\varepsilon^{4/(k-2)} t^{2/k}\right)^{k/2} = \varepsilon t$$

k-cycles from G, since Lemma 3.8 states that each such connected component has at most $\frac{1}{2k}\left(2 \cdot \frac{1}{2}k^{2/k}\varepsilon^{4/(k-2)} t^{2/k}\right)^{k/2}$ unlabeled k-cycles.

Therefore G' satisfies (C1) since $N(C_k, G) \geq c$ by hypothesis and each of the three steps above removes at most εt k-cycles from the graph. Finally, G' satisfies

- (C2) because G has at most $\varepsilon^{-1} t^{2/k}$ edges by hypothesis and G' is a subgraph of G,
- (C3) because G_s satisfies (C3) by construction in step (2), and removing connected components in step (3) cannot change this property,
- (C4) because all components with less than $\frac{1}{2}k^{2/k}\varepsilon^{4/(k-2)} t^{2/k}$ edges were removed in step (3) and there were at most $2\varepsilon^{-(k+2)/(k-2)}$ connected components, and
- (C5) if k is even, because step (1) removed all edges $uv \in E(G)$ for which $\deg u \deg v > 2^{k/2}\varepsilon^{-(k+4)/2} t^{2/k}$, and removing further edges cannot increase $\deg u \deg v$ for any edge $uv \in G$.

Hence G' is a (ε, t)-core, as desired. ☐

Finally, note that for $\varepsilon \lesssim_k 1$, $n \gtrsim_{k,\varepsilon,\delta} 1$ and any $t \leq \frac{k!}{2k}\binom{n}{k}$, the clique $K_{\min(n, \lceil (2kt)^{1/k} \rceil + k)}$ is an (ε, t)-core and occurs in $S^*(G)$ for $G \sim \mathbb{G}(n,p)$ with $np = \lambda \in [\delta, \delta^{-1}]$ with probability at least $p^{(1+\varepsilon)(2kt)^{2/k}/2}$ (see Lemma 2.6) if $t \geq \sqrt{\log n}$. Hence

$$\mathbb{P}_{G \sim \mathbb{G}(n,p)}[\mathrm{CORE}(\varepsilon, t, S^*(G))] \geq p^{(1+\varepsilon)(2kt)^{2/k}/2} \tag{3.9}$$

which, together with Lemma 3.14 and Lemma 3.15, allows us to conclude:

Corollary 3.16. *Fix an integer $k \geq 3$, reals $\varepsilon, \delta > 0$ and suppose $\varepsilon \lesssim_k 1$ and $n \gtrsim_{k,\varepsilon,\delta} 1$. Then for $np = \lambda \in [\delta, \delta^{-1}]$ and for any nonnegative integer $\sqrt{\log n} \leq t \leq n^{1/4k}$,*

$$\mathbb{P}_{G \sim \mathbb{G}(n,p)}\left[N(C_k, S^*(G)) \geq t\right] \leq (1+\varepsilon)\mathbb{P}_{G \sim \mathbb{G}(n,p)}[\mathrm{CORE}(\varepsilon, t, S^*(G))].$$

Now we will need to extend Corollary 3.16 also to the regime $t > n^{1/4k}$ of large t. This is more difficult and requires the moment method introduced in [14]. First, we need to obtain a more loose structure of a seed and then refine it to show the existence of a core.

Definition 3.17. Fix an integer $k \geq 3$, positive real ε and suppose $n \gtrsim_{k,\varepsilon,\delta} 1$. Then for any nonnegative integer $t \leq \frac{k!}{2k}\binom{n}{k}$ define an (ε,t)-*seed* as a fixed embedding into K_n of a graph H without isolated vertices which satisfies

(S1) $\mathbb{E}_{G\sim\mathbb{G}(n,p)}[N(C_k,G) \mid H \subseteq G] \geq (1-\varepsilon)t$

(S2) $e(G) \leq \varepsilon^{-1}k(2k)^{2/k}t^{2/k}\log\frac{1}{p}$.

Let also $\mathrm{SEED}(\varepsilon,t,G)$ denote the event that a random graph G contains an (ε,t)-seed.

It is worth noting that our Definition 3.17 of seed corresponds roughly to the seed from [12] and preseed from [2]. However, [2] uses a bound corresponding to the result of Claim 3.19 instead of (S2) in their definition of seed and later also core.

Claim 3.18. *Fix an integer $k \geq 3$, reals $\varepsilon,\delta > 0$ and suppose $n \gtrsim_{k,\varepsilon,\delta} 1$ and $\varepsilon \lesssim_k 1$. Then for $np = \lambda \in [\delta,\delta^{-1}]$ and for any nonnegative integer $t \leq \frac{k!}{2k}\binom{n}{k}$ we have*

$$\mathbb{P}_{G\sim\mathbb{G}(n,p)}\left[N(C_k,G) \geq t\right] \leq (1+\varepsilon)\mathbb{P}[\mathrm{SEED}(\varepsilon,t,G)]\,.$$

Proof. In this proof we will use a simplified notation of

$$\mathbb{E}_H[X] := \mathbb{E}_{G\sim\mathbb{G}(n,p)}[X \mid H \subseteq G]$$

where H is some fixed embedding of a graph in K_n and X is a random variable being a function of a random graph G. Also, $\mathbb{E}[X]$ and $\mathbb{P}[X]$ should mean $\mathbb{E}_{G\sim\mathbb{G}(n,p)}[X]$ and $\mathbb{P}_{G\sim\mathbb{G}(n,p)}[X]$ respectively if not specified otherwise.

Let Z be the indicator random variable for G not containing an (ε,t)-seed. Let $Z_H = 1$ if H is not a seed and $Z_H = 0$ otherwise. Additionally, let $\mathcal{C}$ be the set of k-cycles in K_n, and for every $T \in \mathcal{C}$ let Y_T be the indicator random variable for $T \subseteq G$. Define $l = \varepsilon^{-1}(2k)^{2/k}t^{2/k}\log\frac{1}{p}$. Then for every positive integer $\ell' \leq \ell + 1$,

$$\begin{aligned}
\mathbb{E}_{G\sim\mathbb{G}(n,p)}\left[N(C_k,G)^{\ell'}Z\right] &= \mathbb{E}\left[\left(\sum_{T\in\mathcal{T}} Y_T\right)^{\ell'} Z\right] \\
&= \mathbb{E}\left[\sum_{T_1,\ldots,T_{\ell'}\in\mathcal{T}} Y_{T_1}\cdots Y_{T_{\ell'}}\cdot Z\right] \\
&\leq \mathbb{E}\left[\sum_{T_1,\ldots,T_{\ell'}\in\mathcal{T}} Y_{T_1}\cdots Y_{T_{\ell'}}\cdot Z_{T_1\cup\ldots\cup T_{\ell'-1}}\right] \\
&= \mathbb{E}\left[\sum_{T_1,\ldots,T_{\ell'-1}\in\mathcal{T}} Y_{T_1}\cdots Y_{T_{\ell'-1}}\cdot \mathbb{E}_{T_1\cup\ldots\cup T_{\ell'-1}}[N(C_k,G)]\cdot Z_{T_1\cup\ldots\cup T_{\ell'-1}}\right] \\
&\leq (1-\varepsilon)t\cdot\mathbb{E}\left[\sum_{T_1,\ldots,T_{\ell'-1}\in\mathcal{T}} Y_{T_1}\cdots Y_{T_{\ell'-1}}\cdot Z_{T_1\cup\ldots\cup T_{\ell'-1}}\right] \\
&\leq \cdots \\
&\leq ((1-\varepsilon)t)^{\ell'},
\end{aligned}$$

since if $\mathbb{E}_{T_1\cup\ldots\cup T_{\ell'-1}}[N(C_k,G)] \ge (1-\varepsilon)t$, then $Z=0$. Note that this only holds when $\ell'-1 \le \ell = \frac{\sqrt[k]{(2k)^2}\log\varepsilon}{\log(1-\varepsilon)}\sqrt[k]{t^2}\log\frac{1}{p}$, since the maximum number of edges in a seed is kl, and $T_1\cup\ldots\cup T_{\ell'-1}$ has at most $k(\ell'-1)$ edges. Thus, Markov's inequality gives

$$(1-\varepsilon)^\ell t^\ell \ge \mathbb{E}\big[N(C_k,G)^\ell Z\big] \ge \mathbb{P}\big[N(C_k,G)^\ell Z \ge t^\ell\big]\cdot t^\ell = \mathbb{P}\big[N(C_k,G)Z\ge t\big]\cdot t^\ell\,.$$

Therefore,

$$\mathbb{P}\big[N(C_k,G)Z\ge t\big] \le (1-\varepsilon)^\ell \le \varepsilon p_n^{\frac{1}{2}(1+\varepsilon)\sqrt[k]{(2kt_n)^2}}$$

as $\varepsilon \lesssim_k 1$. Since for $n \gtrsim_{k,\varepsilon,\delta} 1$ a complete graph on $n_0 = \min(\lfloor (2kt)^{1/k}\rfloor + k, n)$ vertices has at most $(1+\varepsilon)\frac{1}{2}\sqrt[k]{(2kt)^2}$ edges and at least t k-cycles, it is a seed. Therefore

$$\begin{aligned}\mathbb{P}(X_n\ge t \text{ and } G \text{ has no seed}) &= \mathbb{P}\big[N(C_k,G)Z\ge t\big]\\ &\le \varepsilon\mathbb{P}(K_{n_0}\subseteq G)\\ &\le \varepsilon\mathbb{P}[\mathrm{SEED}(\varepsilon,t,G)]\,.\end{aligned}$$

Adding

$$\mathbb{P}\big[\{N(C_k,G)\ge t\}\cap \mathrm{SEED}(\varepsilon,t,G)\big] \le \mathbb{P}\big[\,\mathrm{SEED}(\varepsilon,t,G)\big]$$

to both sides gives $\mathbb{P}\big[N(C_k,G)\ge t\big] \le (1+\varepsilon)\mathbb{P}[\mathrm{SEED}(\varepsilon,t,G)]$, as desired. ☐

Before we determine the existence of a core, we will need some preprocessing.

Claim 3.19. *Fix an integer $k\ge 3$ and positive real $\varepsilon<1$ and let $(\log n)^{k^2/2} < t \le \frac{k!}{2k}\binom{n}{k}$ be an integer. Then for $n\gtrsim_{k,\varepsilon} 1$ every (ε,t)-seed G satisfies $N(C_k,G)\ge(1-2\varepsilon)t$.*

Proof. Note that (see a similar idea in [2, Equation 3.1])

$$\mathbb{E}_{G\sim\mathbb{G}(n,p)}\big[N(C_k,G)\big] \le \sum_{J\subseteq C_k} N(J,G)n^{k-v(J)}p^{k-e(J)} \tag{3.10}$$

where the sum runs over all graphs J which have no isolated vertices and are subgraphs of C_k. Consider any $\varnothing\ne J\subsetneq C_k$ without isolated vertices. Such J is a disjoint union of paths $P_{l_1},\ldots,P_{l_j}$ where $l_1+\cdots+l_j\le k$. We may overestimate

$$N(J,G) \le \prod_{i=1}^{j} N(P_{l_i},G)\,. \tag{3.11}$$

Selection of the directions of the first, third, ..., $(l-1)$st edge gives at most $2e(G)$ choices, so $N(P_l,G)\le \big(2e(G)\big)^{l/2}$. When $2\nmid l$, we can select the first vertex of P_l in at most $v_G\le n$ ways and the remaining $l-1$ vertices in at most $\big(2e(G)\big)^{\frac{l-1}{2}}$ ways, as P_l without the first vertex becomes P_{l-1} and we can recall the result for even l. Thus, for odd l, $N(P_l,G)\le\big(2e(G)\big)^{\frac{l-1}{2}}n$. As $e(G)<n^2$, we have a general bound

$$N(P_l,G)\le Ce(G)^{\frac{l-1}{2}} \tag{3.12}$$

valid for all $2\le l\le k$. Note that a more general bound can be found in [12, Theorem 5.4], where some of the properties of a fractional independence number are used. However, in our case the proof is much simpler and hence is provided in this paper.

Now we can plug (3.11) and (3.12) to (3.10) (note that we need to treat $J=\varnothing$ and $J=C_k$ separately):

$$\mathbb{E}_{G\sim\mathbb{G}(n,p)}\big[N(C_k,G)\big] \le n^kp^k + N(C_k,G) + \sum_{l_1,\ldots,l_j} n^{k-l_1-\cdots-l_j}p^{k-(l_1-1)-\cdots-(l_j-1)}\prod_{i=1}^{j} e(G)^{\frac{l_i-1}{2}}n$$

$$= n^k p^k + N(C_k, G) + \sum_{l_1,\dots,l_j} (np)^{k+j-l_1-\dots-l_j} e(G)^{\frac{l_1+\dots+l_j-j}{2}}$$

$$\le \delta^{-k} + N(C_k, G) + C'(k,\varepsilon) e(G)^{\frac{k-1}{2}} . \tag{3.13}$$

where $C'(k,\varepsilon,\delta)$ is some constant depending only on k, ε and δ. The sum in the first line is over all tuples $(l_1,\dots,l_j)$ such that every l_i is a positive integer and $l_1 + \dots + l_j \le k$. In the third line we use $l_1 + \dots + l_j - j \le k-1$. Finally G is a seed so $\mathbb{E}_{G\sim\mathbb{G}(n,p)}\big[N(C_k,G)\big] \ge (1-\varepsilon)t_n$ and $e(G) \le \varepsilon^{-1} k (2k)^{2/k} t^{2/k} \log \frac{1}{p}$. Hence (3.13) results in

$$(1-\varepsilon)t_n \le N(C_k, G) + \delta^{-k} + C'' t_n^{k-1} k (\log n)^{\frac{k-1}{2}} \tag{3.14}$$

where $C'' = C''(k,\lambda,\varepsilon)$ is a constant depending only on k, λ and ε. Finally, note that

$$\varepsilon t^{\frac{k-1}{k}} (\log n)^{\frac{k-1}{2}} + \delta^{-k}$$

as we assumed that $t > (\log n)^{\frac{k^2}{2}}$. Thus for $n \gtrsim_{k,\varepsilon,\delta} 1$ we must have

$$(1-2\varepsilon)t \le N(C_k, G)$$

which finishes the proof. □

Definition 3.20. Fix an integer $k \ge 3$ and positive real $\varepsilon < 1$. Additionally, let t be a positive integer. An (ε,t)-*precore* is a graph G without isolated vertices such that

(PC1) $N(C_k, G) \ge (1-3\varepsilon)t$,
(PC2) $e(G) \le \varepsilon^{-1} k (2k)^{2/k} t^{2/k} \log \frac{1}{p}$,
(PC3) $N_{uv}(C_k, G) \ge \frac{\varepsilon^2}{k(2k)^{2/k}\log\frac{1}{p}} t^{1-2/k}$ for every edge $uv \in E(G)$.

Let also $\mathrm{PC}(\varepsilon,t,G)$ denote the event that a random graph G contains an (ε,t)-precore.

Claim 3.21. *Every (ε,t)-seed contains an (ε,t)-precore.*

Proof. The proof is very similar to step (2) in the proof of Lemma 3.15.

Let G be an n-seed, and repeatedly remove edges from G that violate (PC3) (if a vertex becomes isolated after edge deletion we remove that vertex too) to get a sequence of graphs

$$G = G_1 \longrightarrow G_2 \longrightarrow \dots \longrightarrow G_i$$

that terminates at G_i. By construction, G_i must satisfy (C2) and (C3). To verify G_i satisfies (C1), note that

$$\begin{aligned} N(C_k, G_1) - N(C_k, G_i) &= \sum_{i'=1}^{i-1} \big(N(C_k, G_{i'}) - N(C_k, G_{i'+1})\big) \\ &< (i-1)\cdot \frac{\varepsilon^2 t^{1-2/k}}{k(2k)^{2/k}\log\frac{1}{p_n}} \\ &< \varepsilon^{-1} k (2k)^{2/k} t^{2/k} \log\frac{1}{p_n} \cdot \frac{\varepsilon^2 t^{1-2/k}}{k(2k)^{2/k}\log\frac{1}{p_n}} \\ &= \varepsilon t_n. \end{aligned}$$

Therefore by Claim 3.19

$$N(C_k, G_i) \ge N(C_k, G) - \varepsilon t_n \ge (1-3\varepsilon)t_n\,.$$ □

Claim 3.22. *Fix an integer $k \geq 3$, reals $\varepsilon, \delta > 0$ and suppose $n \gtrsim_{k,\varepsilon,\delta} 1$ and $\varepsilon \lesssim_k 1$. Then for $np = \lambda \in [\delta, \delta^{-1}]$ and any nonnegative integer $\sqrt{\log n} \leq t \leq \frac{k!}{2k}\binom{n}{k}$ we have*

$$\mathbb{P}_{G\sim\mathbb{G}(n,p)}[\mathrm{PC}(\varepsilon,t,G)] \leq (1+\varepsilon)\mathbb{P}_{G\sim\mathbb{G}(n,p)}[\mathrm{CORE}(\varepsilon,t,G)]\,.$$

Proof. It suffices to show that

$$\mathbb{P}_{G\sim\mathbb{G}(n,p)}[\mathrm{PC}(\varepsilon,t,G)\setminus\mathrm{CORE}(\varepsilon,t,G)] \leq \varepsilon\mathbb{P}_{G\sim\mathbb{G}(n,p)}[\mathrm{CORE}(\varepsilon,t,G)]\,.$$

Let $\mathrm{PC}'(\varepsilon,t,G)$ denote that event that G contains an (ε,t)-precore with more than $\varepsilon^{-1}t^{2/k}$ edges. Thanks to Lemma 3.15 it suffices to show

$$\mathbb{P}_{G\sim\mathbb{G}(n,p)}[\mathrm{PC}'(\varepsilon,t,G)] \leq \varepsilon\mathbb{P}_{G\sim\mathbb{G}(n,p)}[\mathrm{CORE}(\varepsilon,t,G)]\,. \tag{3.15}$$

Let H be an (ε,t)-precore with $e(H) > \varepsilon^{-1}t^{2/k}$. Note that due to Lemma 3.9 for any edge $uv \in E(H)$ we have

$$\begin{aligned}\frac{\varepsilon^2}{k(2k)^{2/k}\log\frac{1}{p}}t^{1-2/k} &\leq N_{uv}(C_k,H)\\ &\leq (\deg u\deg v)(2e(H))^{\frac{k-4}{2}}\\ &\leq (\deg u\deg v)2^{(k-4)/2}\varepsilon^{-(k-4)/2}k^{(k-4)/2}(2k)^{(k-4)/k}t^{(k-4)/k}(\log 1/p)^{(k-4)/2}\end{aligned} \tag{3.16}$$

when k is even and

$$\begin{aligned}\frac{\varepsilon^2}{k(2k)^{2/k}\log\frac{1}{p}}t^{1-2/k} &\leq N_{uv}(C_k,H)\\ &\leq \min(\deg u,\deg v)(2e(H))^{\frac{k-3}{2}}\\ &\leq \min(\deg u,\deg v)2^{(k-3)/2}\varepsilon^{-(k-3)/2}k^{(k-3)/2}(2k)^{(k-3)/k}t^{(k-3)/k}(\log 1/p)^{(k-3)/2}\end{aligned} \tag{3.17}$$

when k is odd. Then (3.16) rearranges to

$$\deg u\deg v \geq \varepsilon^{k/2}2^{k-4}k^{-(k-2)/2}(2k)^{(k-2)/k}t^{2/k}(\log n)^{-(k-2)/2} \tag{3.18}$$

when $n \geq \delta^{-1}$ and (3.17) gives

$$\deg u\deg v \geq \varepsilon^{k+1}2^{-2(k+3)}k^{-(k-1)}(2k)^{-2(k-1)/k}t^{2/k}(\log n)^{-(k-1)} \tag{3.19}$$

when $n \geq \delta^{-1}$. When $n \gtrsim_{k,\varepsilon,\delta} 1$ we can combine (3.18) and (3.19) to

$$\deg u\deg v \geq t^{2/k}(\log n)^{-2k} \tag{3.20}$$

true (regardless of the parity of k) for any $uv \in E(H)$ for a (ε,t)-precore H.

For a positive integer j define

$$V_j = \left\{v \in V(H) \mid 2^{j-1} \leq \deg v < \min(2^j, e(H))\right\}$$

for $j \leq \lceil\log_2 e(H)\rceil$. Since $\sum_{v\in V(H)}\deg v = 2e(H)$, so for all $j \leq \lceil\log_2 e(H)\rceil$

$$|V_j| < \frac{2e(H)}{2^{j-1}}\,. \tag{3.21}$$

Also, (3.15) ensures that for any edge $uv \in E(H)$ if $u \in V_j$, then $\deg v \geq \frac{t^{2/k}}{2^{j-1}(\log n)^{2k}}$.

Now assume we are given $v(H)$ vertices together with the list of their degrees (so that the vertices can be divided into sets V_j as above), but we are not yet given any edges. For a given $u \in V_j$ only those v with $\deg v \geq 2^{-j}(\log n)^{-2k}t^{2/k}$ are eligible for the other endpoint of an edge incident to u. The number of such possible v is at most

$$M_j = \frac{2e(H)}{2^{-j}(\log n)^{-2k}t^{2/k}} = 2^{j+1}e(H)t^{-2/k}(\log n)^{2k}\,.$$

Thus the number of pairs $\{u,v\}$ such that uv can be an edge is at most

$$M = \sum_{j=1}^{\lceil \log_2 e(H)\rceil} M_j|V_j| \leq 8e(H)^2t^{-2/k}(\log n)^{2k} \tag{3.22}$$

of them.

Finally, note that Lemma 3.6 it follows that $v(H) \leq \lceil \big(1-1/(k^2+k)\big)e(H)\rceil$. Hence for given m all (ε,t)-precores in K_n with m edges are generated in the following process:

(1) select $\lceil \big(1-1/(k^2+k)\big)m\rceil$ vertices of K_n,
(2) fix a degree sequence which possibly includes vertices of degree 0,
(3) divide vertices into sets V_j and select m edges among M possible pairs, where M is defined in (3.22).

Recall that $t \geq \sqrt{\log n} \gtrsim_{k,\varepsilon,\delta} 1$, so $\lceil \big(1-1/(k^2+k)\big)m\rceil < (2k^2-1)m/2k^2$, so there is at most $n^{(2k^2-1)m/2k^2}$ ways to select vertices in step (1). There are also at most $\binom{2m}{m} \leq 4^m$ possible degree sequences (as they are sequences of m nonnegative integers summing to m) at and most $\binom{M}{m} < (eM/m)^m$ ways of selecting edges, where according to (3.22) we have $M \leq 8m^2t^{-2/k}(\log n)^{2k}$.

Hence we can provide the following upper bound for the number N_m of possible (ε,t)-precores with m edges

$$N_m \leq 4^m n^{(2k^2-1)m/2k^2}(8mt^{-2/k}(\log n)^{2k})^m\,. \tag{3.23}$$

Recall that for $n \gtrsim_{k,\varepsilon,\delta} 1$ and any (ε,t)-precore H we have

$$e(H) \leq \varepsilon^{-1}k(2k)^{2/k}t^{2/k}(\log 1/p) < t^{2/k}(\log n)^2\,.$$

Hence (3.23) is transformed to

$$N_m < n^{(2k^2-1)m/2k^2}(32\log n)^{m(2k+2)}\,. \tag{3.24}$$

Now we can bound the probability that $G \sim \mathbb{G}(n,p)$ contains an (ε,t)-precore with more than $\varepsilon^{-1}t^{2/k}$ edges. Thus by the union bound

$$\begin{aligned}
&\mathbb{P}_{G\sim\mathbb{G}(n,p)}[n \text{ contains a } (\varepsilon,t)\text{-precore with more than } \varepsilon^{-1}t^{2/k} \text{ edges}]\\
&\qquad\leq \sum_{m=\lceil \varepsilon^{-1}t^{2/k}\rceil}^{\lfloor t^{2/k}(\log n)^2\rfloor} p^m N_m\\
&\qquad\leq \sum_{m=\lceil \varepsilon^{-1}t^{2/k}\rceil}^{\lfloor t^{2/k}(\log n)^2\rfloor} n^{-m/2k^2}\Big(32\delta^{-1}(\log n)^{2k+2}\Big)^m\\
&\qquad\leq 2n^{-\varepsilon^{-1}t^{2/k}/2k^2}(32\delta^{-1}(\log n)^{2k+2})^{\varepsilon^{-1}t^{2/k}}
\end{aligned}$$

where in the last line we used $n \gtrsim_{k,\delta,\varepsilon} 1$, so $32\delta^{-1}n^{-1/2k^2}(\log n)^{2k+2} < 1/2$.

Finally, using (3.9) we conclude that it suffices to check that

$$2n^{-\varepsilon^{-1}t^{2/k}/2k^2}(32\delta^{-1}(\log n)^{2k+2})^{\varepsilon^{-1}t^{2/k}} < \varepsilon p^{(1+\varepsilon)(2kt)^{2/k}/2}$$

which is trivially true when $\varepsilon \lesssim_k 1$. ☐

Finally, Claim 3.18, Claim 3.19, Claim 3.21, and Claim 3.22 allows us to infer the following corollary.

Corollary 3.23. *Fix an integer $k \geq 3$, reals $\varepsilon, \delta > 0$ and suppose $\varepsilon \lesssim_k 1$ and $n \gtrsim_{k,\varepsilon,\delta} 1$. Then for $np = \lambda \in [\delta, \delta^{-1}]$ and for any nonnegative integer $(\log n)^{k^2/2} \leq t \leq \frac{k!}{2k}\binom{n}{k}$*

$$\mathbb{P}_{G\sim\mathbb{G}(n,p)}\left[N(C_k, G) \geq t\right] \leq (1+\varepsilon)\mathbb{P}_{G\sim\mathbb{G}(n,p)}\left[\mathrm{CORE}(\varepsilon, t, G)\right].$$

Having established Corollaries 3.16 and 3.23 we can proceed to providing an upper bound for existence of an (ε, t)-core in $G \sim \mathbb{G}(n,p)$ to ultimately conclude Lemma 3.3.

Lemma 3.24. *Fix an integer $k \geq 3$ and a positive real number ε. Additionally, let t be a positive integer. Then for every connected component of a (ε, t)-core, either*

(i) *every vertex in this connected component has degree at least $C(k,\varepsilon)^{-1}t^{1/k}$, or*
(ii) *this connected component is bipartite, every vertex has degree at least two, and the product of the sizes of the two parts is at most $C(k,\varepsilon)t^{2/k}$*

for some positive real number $C(k,\varepsilon)$ only depending on k and ε.

Proof. Let H be a (ε, c)-core. If k is odd, then for every edge $uv \in E(H)$,

$$\min(\deg u, \deg v) \geq \frac{N_{uv}(C_k, H)}{(2e(H))^{(k-3)/2}} \geq \frac{\varepsilon^2 t^{1-2/k}}{(2\varepsilon^{-1}t^{2/k})^{(k-3)/2}} = 2^{-(k-3)/2}\varepsilon^{(k+1)/2}t^{1/k}$$

by Lemma 3.9, (C2), and (C3).

If k is even, then for every edge $uv \in E(H)$,

$$\begin{aligned}2^{k/2}\varepsilon^{-(k+4)/2}t^{2/k} \geq \deg u \deg v &\geq \frac{N_{uv}(C_k, H)}{(2e(H))^{(k-4)/2}} \\ &\geq \frac{\varepsilon^2 t^{1-2/k}}{(2\varepsilon^{-1}t^{2/k})^{(k-4)/2}} = 2^{-(k-4)/2}\varepsilon^{k/2}t^{2/k}\end{aligned}$$

by (EC5), Lemma 3.9, (C2), and (C3); for brevity, define

$$D(k,\varepsilon) = \max\left(2^{k/2}\varepsilon^{-(k+4)/2}, 2^{(k-4)/2}\varepsilon^{-k/2}, 1+\varepsilon\right)$$

so that

$$\frac{1}{D(k,\varepsilon)}t^{2/k} \leq \deg u \deg v \leq D(k,\varepsilon)t^{2/k}$$

for all edges $uv \in E(H)$. Now, define sets

$$\begin{aligned}
S_0 &= \{v \in V(H) \mid \deg v = 2\} \\
S_1 &= \left\{v \in V(H) \mid \frac{1}{2D(k,\varepsilon)}t^{2/k} \leq \deg v \leq \frac{D(k,\varepsilon)}{2}t^{2/k}\right\} \\
S_2 &= \left\{v \in V(H) \mid 2 < \deg v \leq 2D(k,\varepsilon)^2\right\} \\
S_3 &= \left\{v \in V(H) \mid \frac{1}{2D(k,\varepsilon)^3}t^{2/k} \leq \deg v < \frac{1}{2D(k,\varepsilon)}t^{2/k}\right\} \\
&\vdots \\
S_i &= \left\{v \in H \mid 2D(k,\varepsilon)^{i-2} < \deg V \leq 2D(k,\varepsilon)^i\right\} && \text{when } i \text{ is even and } 2 \leq i \leq s \\
S_i &= \left\{v \in H \mid \frac{1}{2D(k,\varepsilon)^i}t^{2/k} \leq \deg V < \frac{1}{2D(k,\varepsilon)^{i-2}}t^{2/k}\right\} && \text{when } i \text{ is odd and } 2 \leq i \leq s
\end{aligned}$$

where s is the largest integer for which

$$2D(k,\varepsilon)^s < D(k,\varepsilon)^{-1/2}t^{1/k} \iff \frac{1}{2D(k,\varepsilon)^s}t^{2/k} > D(k,\varepsilon)^{1/2}t^{1/k}$$

holds. Next, define $S_{s+1} = V(H) \setminus (S_1 \cup \ldots \cup S_s)$ so that the sets $S_1, \ldots, S_{s+1}$ partition $V(H)$ by degree; note that by maximality of s,

$$\min_{v \in S_{s+1}} \deg v \geq D(k,\varepsilon)^{-2} D(k,\varepsilon)^{-1/2} t^{1/k}.$$

Lastly, for convenience set $S_i = \varnothing$ for every integer i satisfying $i < 0$ or $i > s+1$.

Since $\frac{1}{D(k,\varepsilon)}t^{2/k} \leq \deg u \deg v \leq D(k,\varepsilon)t^{2/k}$ for all edges $uv \in E(H)$,

- all edges with an endpoint in S_0 have an endpoint in S_1,
- for each integer i satisfying $1 \leq i \leq s$, all edges with an endpoint in S_i have an endpoint in $S_{i-1} \cup S_{i+1}$, and
- all edges with an endpoint in S_{s+1} have an endpoint in $S_s \cup S_{s-1}$ or have both endpoints in S_{s+1}.

In particular, if $uv \in E(H)$ is an edge with $u \in S_i$ and $v \in S_j$, then $|i-j| \leq 1$ ($i = j$ only if $i = j = s+1$) or $|i-j| = 2$ and at least one of i, j is equal to $s+1$.

Now, for every connected component J of H, let

$$a(J) = \min\left\{i \in \mathbb{Z} \mid 0 \leq i \leq s+1, |V(J) \cap S_i| \geq 1\right\}$$

and

$$b(J) = \max\left\{i \in \mathbb{Z} \mid 0 \leq i \leq s+1, |V(J) \cap S_i| \geq 1\right\};$$

since J is connected, there exists an edge $u_i v_i \in E(J)$ with $u_i \in S_i$ and $v_i \in S_{i+1}$ for every integer i satisfying $a(J) \leq i < b(J) - 1$. For every such edge,

$$\begin{aligned}
\varepsilon^2 t^{1-2/k} &\leq N_{u_i v_i}(C_k, H) \\
&= N_{u_i v_i}(C_k, J) \\
&= N_{u_i v_i}(C_k, J[S_{i-k/2+1} \cup \ldots \cup S_{i+k/2}]) \\
&\leq \left(2|E(J[S_{i-k/2+1} \cup \ldots \cup S_{i+k/2}])|\right)^{(k-2)/2}
\end{aligned}$$

because the first inequality follows from (C3), the first equality follows from the fact that k-cycles is connected, the second equality follows from the fact that any k-cycle with vertices in both S_i and S_{i+1} must be contained in the sets $S_{i-k/2+1} \cup \ldots \cup S_{i+k/2}$, and the last inequality follows from Lemma 3.9. Therefore for every connected component J of H,

$$|E(J[S_{i-k/2+1} \cup \ldots \cup S_{i+k/2}])| \geq \frac{1}{2}\left(\varepsilon^2 t^{1-2/k}\right)^{2/(k-2)} = \frac{1}{2}\varepsilon^{4/(k-2)} t^{2/k}$$

for every integer i satisfying $a(J) \leq i < b(J)$. Summing this inequality over all such i gives

$$\begin{aligned}
(b(J) - a(J) - 1) \cdot \frac{1}{2}\varepsilon^{4/(k-2)} t^{2/k} &\leq \sum_{i=a}^{b-1} |E(J[S_{i-k/2+1} \cup \ldots \cup S_{i+k/2}])| \\
&\leq k|E(J)| \\
&\leq k\varepsilon^{-1} t^{2/k};
\end{aligned}$$

the second inequality follows because every edge in J is counted at most k times in the sum, and the third inequality follows from (C2). Thus,

$$b(J) - a(J) \leq \frac{k\varepsilon^{-1} t^{2/k}}{\frac{1}{2}\varepsilon^{4/(k-2)} t^{2/k}} + 1 = 2k\varepsilon^{-(k+2)/(k-2)} + 1$$

for every connected component J of H.

To finish the proof of the lemma, there are two cases.

- If $b(J) = s+1$, then $a(J) \geq s+1-2k\varepsilon^{-(k+2)/(k-2)}$. Therefore, the minimum degree of any vertex in J is at least
$$\begin{aligned} 2D(k,\varepsilon)^{a(J)-2} &\geq D(k,\varepsilon)^{s-2-2k\varepsilon^{-(k+2)/(k-2)}} \\ &\geq 2\left(D(k,\varepsilon)^{-3/2}t^{1/k}\right)\left(D(k,\varepsilon)^{-2-2k\varepsilon^{-(k+2)/(k-2)}}\right) \\ &= 2D(k,\varepsilon)^{-7/2-2k\varepsilon^{-(k+2)/(k-2)}}t^{1/k}, \end{aligned}$$
where the second inequality follows from the maximality of s.
- If $b(J) < s+1$, recall that every edge $uv \in E(H)$ with $u \in S_i$, $v \in S_j$, and $u, v \leq s$ satisfies $|i-j| = 1$. Hence, the vertices of J can be partitioned into the parts
$$A = V(J) \cap \bigcup_{\substack{a \leq i \leq b \\ i \text{ even}}} S_i$$
and
$$B = V(J) \cap \bigcup_{\substack{a \leq i \leq b \\ i \text{ odd}}} S_i.$$
Since the vertices of J are contained in $S_{a(J)} \cup \ldots \cup S_{b(J)}$ and the bounds defining S_{i+2} differ from the bounds defining S_i by a factor of D^2 for each integer i satisfying $0 \leq i \leq s-2$,
$$\begin{aligned} \max_{u \in A} \deg u &\leq D(k,\varepsilon)^{b(J)-a(J)+1} \min_{u \in A} \deg u \\ &\leq D(k,\varepsilon)^{2+2k\varepsilon^{-(k+2)/(k-2)}} \min_{u \in A} \deg u \end{aligned}$$
and similarly
$$\max_{v \in B} \deg v \leq D(k,\varepsilon)^{2+2k\varepsilon^{-(k+2)/(k-2)}} \min_{v \in B} \deg v.$$
Therefore,
$$\begin{aligned} |A||B| &\leq \frac{|E(J)|}{\min\limits_{u \in A} \deg u} \cdot \frac{|E(J)|}{\min\limits_{v \in B} \deg v} \\ &\leq \frac{D(k,\varepsilon)^{2+2k\varepsilon^{-(k+2)/(k-2)}}|E(J)|}{\max\limits_{u \in A} \deg u} \cdot \frac{D(k,\varepsilon)^{2+2k\varepsilon^{-(k+2)/(k-2)}}|E(J)|}{\max\limits_{v \in B} \deg v} \\ &\leq \frac{D(k,\varepsilon)^{4+4k\varepsilon^{-(k+2)/(k-2)}}|E(J)|^2}{\max\limits_{uv \in E(J)} \deg u \deg v} \\ &\leq \frac{D(k,\varepsilon)^{4+4k\varepsilon^{-(k+2)/(k-2)}}\left(\varepsilon^{-1}t^{2/k}\right)^2}{2^{k/2}\varepsilon^{-(k+4)/2}t^{2/k}} = 2^{-k/2}\varepsilon^{k/2}D(k,\varepsilon)^{4+4k\varepsilon^{-(k+2)/(k-2)}}t^{2/k} \end{aligned}$$
where the last inequality follows from (C2) and (C5). Finally, note that every vertex in $A \cup B$ must have degree at least two, since cores do not have isolated vertices and any edge incident to a degree-one vertex cannot be contained in any k-cycles.

Finally, choosing
$$C(k,\varepsilon) = \max\left(2^{-(k-3)/2}\varepsilon^{(k+1)/2}, 2D(k,\varepsilon)^{-7/2-2k\varepsilon^{-(k+2)/(k-2)}}, 2^{-k/2}\varepsilon^{k/2}D(k,\varepsilon)^{4+4k\varepsilon^{-(k+2)/(k-2)}}\right)$$
covers all cases for both k odd and k even, as desired. □

Before we continue, let us provide a useful graph theoretic lemma first. It is (especially after transformation to Corollary 3.26) is very similar to [2, Lemma 4.4] for H being a k-cycle, but the slightly stronger statement will be needed for stability considerations.

Lemma 3.25. *Fix an integer $k \geq 3$, and let G be a bipartite graph with parts A and B and no isolated vertices. Then*

$$kN(C_k, G) \leq (e(G) - |A|)^{k/2-1} e(G).$$

Proof. Assume k is even, since $e(G) \geq |A|$ and $N(C_k, G) = 0$ when k is odd. For every positive integer i, let v_i be the number of vertices in A with degree i. Then for every choice of $\frac{k}{2}$ positive integers $d_1, d_3, \ldots, d_{k-1} \in \mathbb{Z}^+$, the number of labeled k-cycles in G whose i^{th} vertex is in A and has degree d_i for each odd integer $1 \leq i \leq k$ is at most

$$\left(v_{d_1} v_{d_3} \cdots v_{d_{k-1}}\right) \cdot d_1 \cdot (d_3 - 1)(d_5 - 1) \cdots (d_{k-1} - 1) = v_1 v_{d_1} \prod_{\substack{3 \leq i \leq k \\ i \text{ odd}}} (d_i - 1) v_{d_i}$$

by choosing the odd-indexed vertices of the k-cycle first, choosing the second vertex next, and choosing the location of the remaining even-indexed vertices in order. Thus, the number of labeled k-cycles in G whose odd-indexed vertices are in A is at most

$$\sum_{\substack{d_1, d_3, \ldots, d_{k-1} \in \mathbb{Z} \\ 2 \leq d_1, d_3, \ldots, d_{k-1} \leq |B|}} d_1 v_{d_1} \prod_{\substack{3 \leq i \leq k \\ i \text{ odd}}} (d_i - 1) v_{d_i} = \left(\sum_{d=1}^{|B|} d v_d\right) \left(\sum_{d=1}^{|B|} (d-1) v_d\right)^{k/2-1}$$

where the first inequality follows because $d_1 \geq 2$ implies $d_1 \leq 2(d_1 - 1)$. Now, the number of ways an unlabeled k-cycle in G can be labeled such that odd-indexed vertices lie in A is $\frac{k}{2} \cdot 2 = k$ by choosing a starting vertex and a direction, so

$$kN(C_k, G) \leq \left(\sum_{d=1}^{|B|} d v_d\right) \left(\sum_{d=2}^{|B|} (d-1) v_d\right)^{k/2}.$$

Finally,

$$e(G) = \sum_{d=1}^{|B|} d v_d$$

since every edge in G is incident to exactly one vertex in A, so combining this result with the previous inequality gives the desired result. ☐

Corollary 3.26. *Fix an integer $k \geq 3$, and let G be a bipartite graph with parts A and B such that $\deg v \geq 2$ for every vertex v in A. Then*

$$e(G) - |A| \geq \left(\frac{k}{2} N(C_k, G)\right)^{2/k}.$$

Proof. Use the bound $e(G) \leq 2(e(G) - |A|)$, then rearrange. ☐

Remark 3.27. Lemma 3.25 and Corollary 3.26 are tight up to lower order terms, as witnessed by $G = K_{2,m}$ (with A being the larger part).

Lemma 3.28. *Fix an integer $k \geq 3$, reals $\varepsilon, \delta > 0$ and suppose $\varepsilon \lesssim_k 1$ and that $n \gtrsim_{k,\varepsilon,\delta} 1$ is a positive integer. Let also $\sqrt{\log n} \leq t \leq \frac{k!}{2k}\binom{n}{k}$ be a positive integer. Then for $pn = \lambda \in [\delta, \delta^{-1}]$ we have*

$$\mathbb{P}_{G \sim \mathbb{G}(n,p)}\left[\text{CORE}(\varepsilon, t, G)\right] \leq p^{(1-\varepsilon)(2kt)^{2/k}/2}.$$

Proof. Let H be any (ε, t)-core. By (C4) and Lemma 3.24, there is a real number $C(k,\varepsilon)$ depending only on k and ε and a nonnegative integer $s \le 2\varepsilon^{-(k+2)/(k-2)}$ for which the vertices of H can be split into sets A_1, B_1, A_2, B_2, …, A_s, B_s, and S such that

- for each integer i satisfying $1 \le i \le s$, $H[A_i \cup B_i]$ is a bipartite connected component of H with parts A_i and B_i such that every vertex in $A_i \cup B_i$ has degree at least two,
- for each integer i satisfying $1 \le i \le s$, $|A_i| \ge |B_i|$ and $|A_i||B_i| \le C(k,\varepsilon)t^{2/k}$, and
- $H[S]$ is a union of connected components, and $\deg v \ge C(k,\varepsilon)^{-1}t^{1/k}$ for every vertex $v \in S$.

Note that $s = 0$ if k is odd, since the number of odd cycles in a bipartite graph is zero. Using this characterization, the probability that H appears in $G \sim \mathbb{G}(n,p)$ is at most

$$\begin{aligned} n^{v(H)}p^{e(H)} &= n^{|S|}p^{e(H[S])}\prod_{i=1}^{s} n^{|A_i|+|B_i|}p^{e(H[A_i\cup B_i])} \\ &\le \delta^{-v(H)}p^{e(H[S])-|S|}\prod_{i=1}^{s}p^{e(H[A_i\cup B_i])-|A_i|-|B_i|} \\ &\le \delta^{-v(H)}p^{(2kN(C_k,H[S]))^{2/k}/2-|S|}\prod_{i=1}^{s}p^{(kN(C_k,H[A_i\cup B_i])/2)^{2/k}-|B_i|} \\ &= \delta^{-v(H)}p^{-(|S|+|B_1|+\cdots+|B_s|)}p^{(2k)^{2/k}/2\cdot N(C_k,H[S])^{2/k}+(k/2)^{2/k}\sum_{i=1}^{s}N(C_k,H[A_i\cup B_i])^{2/k}} \end{aligned} \tag{3.25}$$

where the first inequality follows from Lemma 3.8 and Lemma 3.25. Now, every vertex in S has degree at least $C(k,\varepsilon)^{-1}t^{1/k}$, so

$$|S| \le \frac{2e(H[S])}{C(k,\varepsilon)^{-1}t^{1/k}} \le \frac{2\varepsilon^{-1}t^{2/k}}{C(k,\varepsilon)^{-1}t^{1/k}} = 2\varepsilon^{-1}C(k,\varepsilon)t^{1/k}\,.$$

Additionally, for each integer i satisfying $1 \le i \le s$,

$$|B_i| \le (|A_i||B_i|)^{1/2} \le C(k,\varepsilon)^{1/2}t^{1/k}.$$

Since $s \le 2\varepsilon^{-(k+2)/(k-2)}$,

$$|S| + |B_1| + \ldots + |B_s| \le 2\varepsilon^{-1}C(k,\varepsilon)t^{1/k} + 2\varepsilon^{-(k+2)/(k-2)}C(k,\varepsilon)^{1/2}t^{1/k} \le \varepsilon(2kt)^{1/k}/2\,, \tag{3.26}$$

since $t \ge \sqrt{\log n} \gtrsim_{k,\varepsilon,\delta} 1$.

Now, when k is even (so $k \ge 4$), convexity of the function $f(x) = x^{2/k}$ gives

$$\begin{aligned} &\frac{(2k)^{2/k}}{2}N(C_k,H[S])^{2/k} + \left(\frac{k}{2}\right)^{2/k}\sum_{i=1}^{s}N(C_k,H[A_i\cup B_i])^{2/k} \\ &\ge \min\left(\frac{(2k)^{2/k}}{2},\left(\frac{k}{2}\right)^{2/k}\right)\left(N(C_k,H[S]) + \sum_{i=1}^{s}N(C_k,H[A_i\cup B_i])\right)^{2/k} \\ &= \frac{(2k)^{2/k}}{2}N(C_k,H)^{2/k} \\ &= \frac{((1-3\varepsilon)2kt)^{2/k}}{2}, \end{aligned} \tag{3.27}$$

where the last line follows from (C1). Also, when k is odd, $N(C_k,H[A_i\cup B_i]) = 0$, so

$$\begin{aligned} &\frac{(2k)^{2/k}}{2}N(C_k,H[S])^{2/k} + \left(\frac{k}{2}\right)^{2/k}\sum_{i=1}^{s}N(C_k,H[A_i\cup B_i])^{2/k} \\ &= \frac{(2k)^{2/k}}{2}N(C_k,H)^{2/k} \end{aligned}$$

$$= \frac{((1-3\varepsilon)2kt)^{2/k}}{2}, \tag{3.28}$$

Finally, plugging (3.26) to (3.28) into (3.25) shows that the probability that H appears in $G \sim \mathbb{G}(n,p)$ is at most

$$\delta^{-v(G)} p^{((1-3\varepsilon)2kt)^{2/k}/2-\varepsilon(2kt)^{2/k}} \leq \delta^{-\varepsilon^{-1}t^{2/k}} p^{(1-4\varepsilon)(2kt)^{2/k}/2} \tag{3.29}$$

as $v(G) < e(G) \leq \varepsilon^{-1}t^{2/k}$. To find the total probability that any (ε, t)-core appears, it suffices to sum (3.29) over all (ε, t)-cores.

There are at most n^{2s+1} ways to pick the sizes of the sets A_1, B_1, …, A_s, B_s, and S. There are at most $\binom{n}{2}^{s+1}$ ways to choose the sizes of $E(H[A_1 \cup B_1])$, …, $E(G[A_s \cup B_s])$, and $E(H[S])$. For each integer i satisfying $1 \leq i \leq s$, there are at most

$$\begin{aligned} \binom{|A_i||B_i|}{e(H[A_i \cup B_i])} &\leq \left(\frac{e|A_i||B_i|}{e(H[A_i \cup B_i])}\right)^{e(H[A_i \cup B_i])} \\ &\leq \max\left(1, \frac{eC(k,\varepsilon)t^{2/k}}{\frac{1}{2}k^{2/k}\varepsilon^{4/(k-2)}t^{2/k}}\right)^{\varepsilon^{-1}t^{2/k}} \\ &= \max\left(1, 2ek^{-2/k}\varepsilon^{-4/(k-2)}C(k,\varepsilon)\right)^{\varepsilon^{-1}t^{2/k}} \end{aligned}$$

ways to choose the edges in $E(H[A_i \cup B_i])$ given the number of vertices and edges in $H[A_i \cup B_i]$, since the first inequality follows from Lemma 3.5 and the second inequality follows from Lemma 3.24, (C4), and (C2). Additionally, there are at most

$$\binom{\binom{|S|}{2}}{e(H[S])} \leq \left(\frac{e\binom{|S|}{2}}{e(H[S])}\right)^{e(H[S])} \leq \left(\frac{e|S|^2}{2e(H[S])}\right)^{e(H[S])}$$

ways to choose the edges in $E(H[S])$ given the number of vertices and edges in $H[S]$ (the first inequality follows from Lemma 3.5). Since S is a union of connected components and every vertex in S has degree at least $C(k,\varepsilon)^{-1}t^{1/k}$,

$$|S| \leq \frac{2|E(H[S])|}{C'(k,\varepsilon)^{-1}t^{1/k}} \leq 2C(k,\varepsilon)t^{-1/k}e(H[S])\,;$$

substituting this into the previous inequality gives the number of ways to choose the edges in $E(H[S])$ as at most

$$\begin{aligned} \left(\frac{e\left(2C(k,\varepsilon)t^{-1/k}e(H[S])\right)^2}{2e(H[S])}\right)^{e(H[S])} &\leq \left(\frac{2eC(k,\varepsilon)^2 e(H[S])}{t^{2/k}}\right)^{e(H[S])} \\ &\leq \left(\frac{2eC(k,\varepsilon)^2\varepsilon^{-1}t^{2/k}}{t^{2/k}}\right)^{e(H[S])} \\ &= \left(2eC(k,\varepsilon)^2\varepsilon^{-1}\right)^{e(H[S])} \\ &\leq \max\left(1, 2eC(k,\varepsilon)^2\varepsilon^{-1}\right)^{\varepsilon^{-1}t^{2/k}}, \end{aligned}$$

where both inequalities follow from (C2). Therefore the number of (ε, t)-cores is at most

$$n^{2s+1} \cdot \binom{n}{2}^{s+1} \cdot \left(\max\left(1, 2ek^{-2/k}\varepsilon^{-4/(k-2)}C(k,\varepsilon)\right)^{\varepsilon^{-1}t^{2/k}}\right)^s \cdot \max\left(1, 2eC(k,\varepsilon)^2\varepsilon^{-1}\right)^{\varepsilon^{-1}t^{2/k}}$$

$$
\begin{aligned}
&\le n^{4s+3}\left(\max\left(1, 2ek^{-2/k}\varepsilon^{-4/(k-2)}C(k,\varepsilon)\right)^{\varepsilon^{-1}s}\cdot \max\left(1, 2eC'(k,\varepsilon)^2\varepsilon^{-1}\right)^{\varepsilon^{-1}}\right)^{t^{2/k}}\\
&\le n^{8\varepsilon^{-(k+2)/(k-2)}+3}\left(\max\left(1, 2ek^{-2/k}\varepsilon^{-4/(k-2)}C'(k,\varepsilon)\right)^{2\varepsilon^{-4/(k-2)}}\cdot \max\left(1, 2eC(k,\varepsilon)^2\varepsilon^{-1}\right)^{\varepsilon^{-1}}\right)^{t^{2/k}}\\
&= n^{C_2(k,\varepsilon)}C_3(k,\varepsilon)^{t^{2/k}} \qquad (3.30)
\end{aligned}
$$

where

$$
C_2(k,\varepsilon) = 8\varepsilon^{-(k+2)/(k-2)+3} + 3
$$

and

$$
C_3(k,\varepsilon) = \max\left(1, 2ek^{-2/k}\varepsilon^{-4/(k-2)}C(k,\varepsilon)\right)^{2\varepsilon^{-4/(k-2)}}\cdot \max\left(1, 2eC(k,\varepsilon)^2\varepsilon^{-1}\right)^{\varepsilon^{-1}};
$$

note that the second inequality follows from the upper bound on s.

Finally, the probability any (ε, t)-core appears is at most

$$
\begin{aligned}
&n^{C_2(k,\varepsilon)}C_3(k,\varepsilon)^{t^{2/k}}\delta^{-\varepsilon^{-1}t^{2/k}}p^{(1-4\varepsilon)(2kt)^{2/k}/2}\\
&\le p^{(1-5\varepsilon)(2kt)^{2/k}/2}
\end{aligned}
$$

when $n \gtrsim_{k,\varepsilon,\delta} 1$ and $t \ge \sqrt{\log n} \gtrsim_{k,\varepsilon,\delta} 1$. Replacing ε with $\varepsilon/5$ finshes the proof. □

The last component needed in the final bound will be an upper bound for the probability that $G \sim \mathbb{G}(n,p)$ has small but nonempty essential skeleton. Fortunately, only a simple rough bound will suffice.

Lemma 3.29. *Fix an integer $k \ge 3$ and a positive real δ. Let n be a positive integer and let a positive real $p < 1$ be such that $pn = \lambda \in [\delta, \delta^{-1}]$. Then*

$$
\mathbb{P}_{G\sim\mathbb{G}(n,p)}\left[e(S^*(G)) \ge 1\right] \le \frac{(k!)^2\delta^{-4k}}{4k}n^{-1}.
$$

Proof. Note first that $e(S^*(G)) \ge 1$ if and only if G contains two k-cycles that share (at least one) vertex. For every nonnegative integer v, the number of pairs of k-cycles sharing v vertices is at most

$$
\binom{n}{v}\binom{n}{k-v}^2\left(\frac{k!}{2k}\right)^2 \le \left(\frac{k!}{2k}\right)^2 n^{2k-v}
$$

by choosing the shared vertices first, the remaining vertices next, and the edges within each k-cycle last. Additionally, for any two distinct unlabeled cycles $\mathcal{C}_1$ and $\mathcal{C}_2$ that share a vertex,

$$
e(\mathcal{C}_1 \cup \mathcal{C}_2) - v(\mathcal{C}_1 \cup \mathcal{C}_2) \ge \frac{e(\mathcal{C}_1 \cup \mathcal{C}_2)}{k} - 1 \ge \frac{k+1}{k} - 1 > 0
$$

by Lemma 3.6, so if k-cycles $\mathcal{C}_1$ and $\mathcal{C}_2$ share v vertices, $\mathcal{C}_1 \cup \mathcal{C}_2$ has at least $2k - v + 1$ edges.

Therefore

$$
\begin{aligned}
\mathbb{P}_{G\sim\mathbb{G}(n,p)}\left[e(S^*(G)) \ge 1\right] &\le \sum_{v=1}^{k}\sum_{\substack{\mathcal{C}_1,\mathcal{C}_2\ k\text{-cycles}\\ v(\mathcal{C}_1\cap\mathcal{C}_2)=v}} p^{2k-e(\mathcal{C}_1\cap\mathcal{C}_2)}\\
&\le \sum_{v=1}^{k}\sum_{\substack{\mathcal{C}_1,\mathcal{C}_2\ k\text{-cycles}\\ |V(\mathcal{C}_1\cap\mathcal{C}_2)|=v}}^{2k} p^{2k-v+1}\\
&\le \sum_{v=1}^{k}\left(\frac{k!}{2k}\right)^2 n^{2k-v}p^{2k-v+1}
\end{aligned}
$$

$$\le \frac{(k!)^2\delta^{-4k}}{4k} n^{-1}. \qquad \square$$

Now we are finally ready to prove Lemma 3.3.

Proof of Lemma 3.3. In this proof $\mathbb{P}[X]$ and $\mathbb{E}[Y]$ should mean $\mathbb{P}_{G\sim\mathbb{G}(n,p)}[X]$ and $\mathbb{E}_{G\sim\mathbb{G}(n,p)}[Y]$ unless specified otherwise.

First let us consider $t \ge (\log n)^{k^2/2}$. Then by Remark 1.5 it suffices to show that

$$\mathbb{P}\big[N(C_k, G) \ge t\big] \le p^{(1-\varepsilon)(2kt)^{2/k}/2}.$$

However, we have

$$\begin{aligned}\mathbb{P}\big[N(C_k, G) \ge t\big] &\le (1+\varepsilon)\mathbb{P}\big[\,\mathrm{CORE}(\varepsilon, t, G)\big]\\ &\le p^{(1-\varepsilon)(2kt)^{2/k}/2}\end{aligned}$$

for $n \gtrsim_{k,\varepsilon,\delta}$, where we used Corollary 3.23 in the first inequality, and Lemma 3.28 in the second one. This settles the result for $t \ge (\log n)^{k^2/2}$.

Now assume $\sqrt{\log n} \le t < (\log n)^{k^2/2}$. Since $N(C_k, G) = N(C_k, \mathcal{S}(G))$ we have

$$\mathbb{P}_{G\sim\mathbb{G}(n,p)}\big[N(C_k, G) \ge t\big] \le \sum_{t_1=0}^{t} \mathbb{P}\big[Y(t_1,1)\big]\mathbb{P}\big[X_+(t-t_1,2)\big] + \mathbb{P}\big[X(t+1,1)\big] \tag{3.31}$$

since $\mathbb{P}[X_+(t_2,2) \mid Y(t_1,1)] \le \mathbb{P}[X_+(t_2,2)]$. Denote $t_2 = t - t_1$. Since $t \ge \sqrt{\log n}$, we can use Lemma 3.11 and Lemma 2.5 to conclude that

$$\mathbb{P}\big[X(t+1,1)\big] \le \frac{1+\varepsilon}{(t+1)!}\left(\frac{\lambda^k}{2k}\right)^{t+1} < \frac{\lambda^k e^{\lambda^k/2k}}{kt}\mathbb{P}\big[Y(1,t)\big] < \varepsilon\mathbb{P}\big[Y(1,t)\big] \tag{3.32}$$

so that (3.31) gives

$$\begin{aligned}&\mathbb{P}_{G\sim\mathbb{G}(n,p)}\big[N(C_k, G) \ge t\big] \le\\ &\quad\le (1+\varepsilon)\mathbb{P}\big[Y(t,1)\big] + \mathbb{P}\big[X_+(t,2)\big] + \sum_{t_1=1}^{t-1}\mathbb{P}\big[Y(t_1,1)\big]\mathbb{P}\big[X_+(t-t_1,2)\big]\\ &\quad\le (1+3\varepsilon)\frac{e^{-\lambda^k/2k}}{t!}\left(\frac{\lambda^k}{2k}\right)^t + \mathbb{P}\big[X_+(t,2)\big] + 2\sum_{t_1=1}^{t-1}\frac{e^{-\lambda^k/2k}}{t_1!}\left(\frac{\lambda^k}{2k}\right)^{t_1}\mathbb{P}\big[X_+(t-t_1,2)\big].\end{aligned} \tag{3.33}$$

Now we can use Lemma 3.29 to bound $\mathbb{P}[X_+(t_2,2)]$ (note that $S^*(G)$ is nonempty if and only if $X_+(t_2,2)$ for $t_2 \ge 1$) in the regime $t_2 \le \sqrt{\log n}$. This yields

$$\begin{aligned}&\mathbb{P}_{G\sim\mathbb{G}(n,p)}\big[N(C_k, G) \ge t\big]\\ &\quad\le (1+3\varepsilon)\frac{e^{-\lambda^k/2k}}{t!}\left(\frac{\lambda^k}{2k}\right)^t + \mathbb{P}\big[X_+(t,2)\big] + 2(k!)^2\delta^{-4k}n^{-1}\sum_{t_2=1}^{\lfloor\sqrt{\log n}\rfloor}\frac{1}{(t-t_2)!}\left(\frac{\lambda^k}{2k}\right)^{t-t_2} +\\ &\qquad 2\sum_{t_2=\lceil\sqrt{\log n}\rceil}^{t-1}\frac{e^{-\lambda^k/2k}}{(t-t_2)!}\left(\frac{\lambda^k}{2k}\right)^{t-t_2}\mathbb{P}\big[X_+(t_2,2)\big].\end{aligned} \tag{3.34}$$

At the same time note that

$$\frac{1}{(t-t_2)!}\left(\frac{\lambda^k}{2k}\right)^{t-t_2} < \frac{e^{-\lambda^k/2k}}{t!}\left(\frac{\lambda^k}{2k}\right)^t \cdot e^{\lambda^k/2k}\left(\frac{2kt}{\lambda^k}\right)^{t_2} < \frac{e^{-\lambda^k/2k}}{t!}\left(\frac{\lambda^k}{2k}\right)^t t^{2t_2}$$

for $t \geq \sqrt{\log n} \gtrsim_{k,\delta} 1$. This yields

$$2(k!)^2\delta^{-4k}n^{-1}\sum_{t_2=1}^{\sqrt{\log n}}\frac{1}{(t-t_2)!}\left(\frac{\lambda^k}{2k}\right)^{t-t_2} \leq \frac{e^{-\lambda^k/2k}}{t!}\left(\frac{\lambda^k}{2k}\right)^t 2(k!)^2\delta^{-4k}n^{-1}t^{2\sqrt{\log n}}\sqrt{\log n} \leq$$
$$\leq \varepsilon\frac{e^{-\lambda^k/2k}}{t!}\left(\frac{\lambda^k}{2k}\right)^t$$

for $n \gtrsim_{k,\delta} 1$ and $t < (\log n)^{k^2/2}$. Using these results on (3.34) gives

$$\mathbb{P}_{G\sim\mathbb{G}(n,p)}\left[N(C_k,G) \geq t\right]$$
$$\leq (1+4\varepsilon)\frac{e^{-\lambda^k/2k}}{t!}\left(\frac{\lambda^k}{2k}\right)^t + \mathbb{P}\left[X_+(t,2)\right] + 2\sum_{t_2=\lceil\sqrt{\log n}\rceil}^{t-1}\frac{e^{-\lambda^k/2k}}{(t-t_2)!}\left(\frac{\lambda^k}{2k}\right)^{t-t_2}\mathbb{P}\left[X_+(t_2,2)\right]. \quad (3.35)$$

Now we can estimate the sum for $t_2 \geq \sqrt{\log n}$. Note that

$$\begin{aligned}\mathbb{P}\left[X_+(t_2,2)\right] &= \mathbb{P}\left[N(C_k,S^*(G)) \geq t_2\right]\\ &\leq (1+\varepsilon)\mathbb{P}\left[\,\mathrm{CORE}(\varepsilon,t_2,S^*(G))\right]\\ &\leq (1+\varepsilon)p^{(1-\varepsilon)(2kt_2)^{2/k}/2}\\ &\leq p^{(1-2\varepsilon)(2kt_2)^{2/k}/2}\end{aligned}$$

where in the second line we used Corollary 3.16 and Corollary 3.23, in the third line we used Lemma 3.28 and in the fourth line we used $n \gtrsim_{k,\varepsilon,\delta} 1$. Now (3.35) gives

$$\mathbb{P}_{G\sim\mathbb{G}(n,p)}\left[N(C_k,G) \geq t\right]$$
$$\leq (1+4\varepsilon)\frac{e^{-\lambda^k/2k}}{t!}\left(\frac{\lambda^k}{2k}\right)^t + p^{(1-2\varepsilon)(2kt)^{2/k}/2} + 2\sum_{t_2=\lceil\sqrt{\log n}\rceil}^{t-1}\frac{e^{-\lambda^k/2k}}{(t-t_2)!}\left(\frac{\lambda^k}{2k}\right)^{t-t_2}p^{(1-2\varepsilon)(2kt_2)^{2/k}/2}. \quad (3.36)$$

We will need the following fact.

Lemma 3.30. *Fix an integer $k \geq 3$, reals $\varepsilon,\delta > 0$ and suppose $\varepsilon \lesssim_k 1$ and that $n \gtrsim_{k,\varepsilon,\delta} 1$ is a positive integer. Let also $\sqrt{\log n} \leq t \leq (\log n)^{k^2/2}$, $\sqrt{\log n} \leq t_2 \leq t-1$ and $t_1 = t-t_2$ be positive integers. Then for any $pn = \lambda \in [\delta,\delta^{-1}]$ we have*

$$\frac{e^{-\lambda^k/2k}}{t_1!}\left(\frac{\lambda^k}{2k}\right)^{t_1}p^{(1-2\varepsilon)(2kt_2)^{2/k}/2} \leq \frac{\varepsilon}{t}\cdot\frac{e^{-\lambda^k/2k}}{t!}\left(\frac{\lambda^k}{2k}\right)^t \quad (3.37)$$

or

$$\frac{e^{-\lambda^k/2k}}{t_1!}\left(\frac{\lambda^k}{2k}\right)^{t_1}p^{(1-2\varepsilon)(2kt_2)^{2/k}/2} \leq \frac{\varepsilon}{t}p^{(1-3\varepsilon)(2kt)^{2/k}/2}. \quad (3.38)$$

Suppose that Lemma 3.30 holds. Then we can use it to bound the sum in (3.35) to obtain

$$\mathbb{P}_{G\sim\mathbb{G}(n,p)}\left[N(C_k,G) \geq t\right]$$
$$\leq (1+4\varepsilon)\frac{e^{-\lambda^k/2k}}{t!}\left(\frac{\lambda^k}{2k}\right)^t + p^{(1-2\varepsilon)(2kt)^{2/k}/2} + 2\sum_{t_2=\lceil\sqrt{\log n}\rceil}^{t-1}\frac{e^{-\lambda^k/2k}}{(t-t_2)!}\left(\frac{\lambda^k}{2k}\right)^{t-t_2}p^{(1-2\varepsilon)(2kt_2)^{2/k}/2} \quad (3.39)$$

$$\leq (1+6\varepsilon)\frac{e^{-\lambda^k/2k}}{t!}\left(\frac{\lambda^k}{2k}\right)^t + (1+2\varepsilon)p^{(1-3\varepsilon)(2kt)^{2/k}/2} \quad (3.40)$$

$$\leq (1+6\varepsilon)\frac{e^{-\lambda^k/2k}}{t!}\left(\frac{\lambda^k}{2k}\right)^t + p^{(1-4\varepsilon)(2kt)^{2/k}/2}\,. \tag{3.41}$$

Now if $p^{(1-4\varepsilon)(2kt)^{2/k}/2} < \varepsilon\frac{e^{-\lambda^k/2k}}{t!}\left(\frac{\lambda^k}{2k}\right)^t$, we conclude

$$\mathbb{P}_{G\sim\mathbb{G}(n,p)}\left[N(C_k,G)\geq t\right] \leq (1+7\varepsilon)\frac{e^{-\lambda^k/2k}}{t!}\left(\frac{\lambda^k}{2k}\right)^t$$

and in the opposite case

$$\mathbb{P}_{G\sim\mathbb{G}(n,p)}\left[N(C_k,G)\geq t\right] \leq (\varepsilon^{-1}+1)p^{(1-4\varepsilon)(2kt)^{2/k}/2} \leq p^{(1-5\varepsilon)(2kt)^{2/k}/2}\,.$$

In both cases replacing ε with $\varepsilon/7$ finishes the proof. Therefore it remains to show Lemma 3.30. □

Proof of Lemma 3.30. Define $c=(2k)^{2/k}/2$. Let us start from some basic bounds. We have

$$\begin{aligned} A_1 &:= \frac{e^{-\lambda^k/2k}}{t_1!}\left(\frac{\lambda^k}{2k}\right)^{t_1} p^{(1-2\varepsilon)(2kt_2)^{2/k}/2}\cdot\left(\frac{\varepsilon}{t}p^{(1-3\varepsilon)(2kt)^{2/k}/2}\right)^{-1} \\ &\leq \frac{\varepsilon^{-1}e^{-\lambda^k/2k}t}{e\sqrt{t_1}(t_1/e)^{t_1}}\left(\frac{\lambda^k}{2k}\right)^{t_1} p^{(1-2\varepsilon)(2kt_2)^{2/k}/2-(1-3\varepsilon)(2kt)^{2/k}/2} && (3.42) \\ &\leq \varepsilon^{-1}e^{-\lambda^k/2k-1}t\left(\frac{e\lambda^k}{2kt_1}\right)^{t_1} p^{\varepsilon ct_2^{2/k}-2(1-3\varepsilon)ct_1^{2/k}/kt_2^{\frac{k-2}{k}}} && (3.43) \\ & && (3.44) \end{aligned}$$

where in the first inequality we used $l! < e\sqrt{l}(l/e)^l$ (true for any integer l) and in the second inequality we used convexity $t^{2/k}\leq t_2^{2/k}+\frac{2}{k}t_1^{2/k}/t_2^{\frac{k-2}{k}}$. If $A_1\leq 1$, the lemma (specifically, (3.38)) is satisfied. Hence assume $A_1\geq 1$ and take natural logarithm of (3.43)

$$\begin{aligned} 0\leq \log A_1 &\leq -\frac{\lambda^k}{2k}-1+\log\varepsilon^{-1}+\log t+t_1\left(\log\frac{e\lambda^k}{2k}-\log t_1\right) \\ &\quad+\left(\varepsilon(2kt_2)^{2/k}/2-(1-3\varepsilon)\frac{2ct_1}{kt_2^{\frac{k-2}{k}}}\right)(\log\lambda-\log n) \\ &\leq -t_1\log t_1+(1-2\varepsilon)\frac{2ct_1\log n}{kt_2^{(k-2)/k}}-\varepsilon(1-\varepsilon)ct_2^{2/k}\log n && (3.45) \\ &\quad+\log\varepsilon^{-1}+t_1\log\frac{e\lambda^k}{2k}+\log t && (3.46) \\ & && (3.47) \end{aligned}$$

where in the second line we used $\varepsilon\lesssim_{k,\delta}1$, and $n\gtrsim_\delta 1$. Now if $t_1<(\log n)^{1/4k}$, then since $t_2\geq\sqrt{\log n}$ we would have for $n\gtrsim_{k,\varepsilon,\delta}$

$$\varepsilon(1-\varepsilon)ct_2^{2/k}\log n > (\log n)^{1+1/1/k} > \frac{(2kt_1)^{2/k}\log n}{2}+\log\varepsilon^{-1}+t_1\log\frac{e\lambda^k}{2k}+\log t$$

(note that $\log t < k^2\log\log n$) so the right side of (3.46) would be negative. Therefore $t_1\geq(\log n)^{1/2k}\gtrsim_{k,\varepsilon,\delta}1$ and hence

$$-t_1\log t_1+\log\varepsilon^{-1}+t_1\log\frac{e\lambda^k}{2k}+\log t < -(1-\varepsilon)t_1\log t_1\,.$$

Therefore from (3.46) we infer

$$0 \le \log A_1 \le -(1-\varepsilon)t_1 \log t_1 + (1-2\varepsilon)\frac{2ct_1 \log n}{kt_2^{(k-2)/k}} - \varepsilon(1-\varepsilon)ct_2^{2/k}\log n \tag{3.48}$$

and hence

$$0 \le -t_2^{\frac{k-2}{k}} \log t_1 + (1-\varepsilon)\frac{2c}{k}\log n - \varepsilon c\left(\frac{t_2}{t_1}\right)\log n\,. \tag{3.49}$$

Therefore we must have

$$t_2 \le \varepsilon^{-1} t_1 \tag{3.50}$$

as well as

$$t_2^{\frac{k-2}{k}} \log t_1 \le (1-\varepsilon)\frac{2c}{k}\log n\,. \tag{3.51}$$

Now let us analyze (3.37). We have

$$\begin{aligned} A_2 :&= \frac{e^{-\lambda^k/2k}}{t_1!}\left(\frac{\lambda^k}{2k}\right)^{t_1} p^{(1-2\varepsilon)(2kt_2)^{2/k}/2} \cdot \left(\frac{\varepsilon}{t}\frac{e^{-\lambda^k/2k}}{t!}\left(\frac{\lambda^k}{2k}\right)^{t}\right)^{-1} \\ &\le \frac{t!\varepsilon^{-1}t}{t_1!}\left(\frac{\lambda^k}{2k}\right)^{-t_2} p^{(1-2\varepsilon)ct_2^{2/k}} \\ &\le \varepsilon^{-1}t^{t_2+1}\left(\frac{\lambda^k}{2k}\right)^{-t_2} p^{(1-2\varepsilon)ct_2^{2/k}}\,. \end{aligned}$$

Assuming that (3.37) doesn't hold, we have

$$\begin{aligned} 0 \le \log A_2 &\le (t_2+1)\log t + \log \varepsilon^{-1} - t_2 \log\frac{\lambda^k}{2k} + (1-2\varepsilon)ct_2^{2/k}(\log\lambda - \log n) \\ &\le (1+\varepsilon)t_2 \log t - (1-3\varepsilon)ct_2^{2/k}\log n \end{aligned} \tag{3.52}$$

where in the second inequality we used $t \ge t_2 \ge \sqrt{\log n} \gtrsim_{k,\varepsilon,\delta} 1$. Now

$$(1-4\varepsilon)c\log n \le t_2^{\frac{k-2}{k}}\log t \le t_2^{\frac{k-2}{k}}\log(2\varepsilon^{-1}t_1) \le (1+\varepsilon)t_2^{\frac{k-2}{k}}\log t_1\,. \tag{3.53}$$

due to (3.50) and $\varepsilon < 1$. Combining this with (3.51) yields

$$\frac{k}{2}(1-4\varepsilon)t_2^{\frac{k-2}{k}}\log t_1 \le (1-4\varepsilon)c\log n \le (1+\varepsilon)t_2^{\frac{k-2}{k}}\log t_1$$

which is a contradiction when $\varepsilon < \frac{k-2}{5k}$. Since we assume $\varepsilon \lesssim_k 1$, the proof is finished. □

References

[1] A. D. Barbour, *Poisson convergence and random graphs*, Math. Proc. Cambridge Philos. Soc. **92** (1982), 349–359. 3

[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 (2019). 2, 13, 14, 21

[3] Béla Bollobás, *Threshold functions for small subgraphs*, Math. Proc. Cambridge Philos. Soc. **90** (1981), 197–206. 3

[4] Béla Bollobás, *Threshold functions for small subgraphs*, Mathematical Proceedings of the Cambridge Philosophical Society **90** (1981), 197–206. 1

[5] Suman Chakraborty, Remco van der Hofstad, and Frank den Hollander, *Sparse random graphs with many triangles*, 2021. 1

[6] Sourav Chatterjee, *The missing log in large deviations for triangle counts*, Random Structures Algorithms **40** (2012), 437–451. 1

[7] Sourav Chatterjee and S. R. S. Varadhan, *The large deviation principle for the Erdős-Rényi random graph*, European J. Combin. **32** (2011), 1000–1017. 1

[8] B. Demarco and J. Kahn, *Tight upper tail bounds for cliques*, Random Structures Algorithms **41** (2012), 469–487. 1

[9] B. DeMarco and J. Kahn, *Upper tails for triangles*, Random Structures Algorithms **40** (2012), 452–459. 1

[10] Paul L. Erdos and Alfréd Rényi, *On the evolution of random graphs*, Transactions of the American Mathematical Society **286** (1984), 257–257. 1

[11] Alan Frieze and Michał Karoński, *Introduction to random graphs*, Cambridge University Press, Cambridge, 2016. 1

[12] Matan Harel, Frank Mousset, and Wojciech Samotij, *Upper tails via high moments and entropic stability*, Duke Mathematical Journal **171** (2022), 2089 – 2192. 1, 2, 7, 13, 14

[13] Svante Janson, Krzysztof Oleszkiewicz, and Andrzej Ruciński, *Upper tails for subgraph counts in random graphs*, Israel J. Math. **142** (2004), 61–92. 1

[14] Svante Janson, Krzysztof Oleszkiewicz, and Andrzej Ruciński, *Upper tails for subgraph counts in random graphs*, Israel J. Math. **142** (2004), 61–92. 13

[15] Svante Janson and Andrzej Ruciński, *The infamous upper tail*, vol. 20, 2002, Probabilistic methods in combinatorial optimization, pp. 317–342. 1

[16] Svante Janson and Andrzej Ruciński, *The deletion method for upper tail estimates*, Combinatorica **24** (2004), 615–640. 1

[17] Svante Janson and Lutz Warnke, *The lower tail: Poisson approximation revisited*, Random Structures Algorithms **48** (2016), 219–246. 1

[18] J. H. Kim and V. H. Vu, *Divide and conquer martingales and the number of triangles in a random graph*, Random Structures Algorithms **24** (2004), 166–174. 1

[19] Eyal Lubetzky and Yufei Zhao, *On replica symmetry of large deviations in random graphs*, Random Structures Algorithms **47** (2015), 109–146. 1

[20] Eyal Lubetzky and Yufei Zhao, *On the variational problem for upper tails in sparse random graphs*, Random Structures Algorithms **50** (2017), 420–436. 1

[21] Andrzej Ruciński, *When are small subgraphs of a random graph normally distributed?*, Probab. Theory Related Fields **78** (1988), 1–10. 3

[22] K. Schürger, *Limit theorems for complete subgraphs of random graphs*, Period. Math. Hungar. **10** (1979), 47–53. 1

[23] J. Spencer and L. Florescu, *Asymptopia*, Student Mathematical Library, vol. 71, American Mathematical Society, Providence, RI, 2014. 6

[24] Van H. Vu, *A large deviation result on the number of small subgraphs of a random graph*, Combin. Probab. Comput. **10** (2001), 79–94. 1

[25] Yufei Zhao, *On the lower tail variational problem for random graphs*, Combin. Probab. Comput. **26** (2017), 301–320. 1

Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
Email address: `{hsmui,asah,msawhney,tomaszsl}@mit.edu`