The upper tail problem for induced 4-cycles in sparse random graphs

Asaf Cohen Antonir

School of Mathematical Sciences, Tel Aviv University, Tel Aviv, Israel

Correspondence
Asaf Cohen Antonir, School of Mathematical Sciences, Tel Aviv University, Tel Aviv, 6997801, Israel.
Email: asafc1@tauex.tau.ac.il

Abstract
Building on the techniques from the breakthrough paper of Harel, Mousset and Samotij, which solved the upper tail problem for cliques, we compute the asymptotics of the upper tail for the number of induced copies of the 4-cycle in the binomial random graph $G_{n,p}$. We observe a new phenomenon in the theory of large deviations of subgraph counts. This phenomenon is that, in a certain (large) range of $p$, the upper tail of the induced 4-cycle does not admit a naive mean-field approximation.

KEYWORDS
non linear large deviations, random graphs, upper tail

1 | INTRODUCTION

The study of the random variable counting the number of copies of a given graph in the binomial random graph $G_{n,p}$ has a very long history and many things are known about it. In particular, for every graph $H$, when the expected number of copies of $H$ in $G_{n,p}$ tends to infinity a ‘law of large numbers’, see Bollobás [6], and a central limit theorem are known, see Ruciński [28].

After establishing such results, it is natural to ask what is the probability of the event that this random variable differs from its expectation by a significant amount. In this paper, we will consider only the ‘upper tail’ problem: what is the probability of a random variable exceeding its expectation by a multiplicative factor $1 + \delta$, where $\delta$ is some positive real.

We continue by recalling some definitions to make our discussion more rigorous. For every non-negative integer $n$ and a sequence $p = p(n) \in (0, 1)$, we let $G_{n,p}$ be the binomial random graph with $n$ vertices and density $p$. Furthermore, for any graph $H$, let $X_{n,p}^H$ be the random variable counting...
the number of (labeled) copies of $H$ in $G_{n,p}$. Further, for any $\delta$ a positive real and a random variable $Y$ we write UT$(Y, \delta)$ for the event \( \{ Y \geq (1 + \delta)\mathbb{E}[Y] \} \).

The work on the problem of estimating (the logarithm of) the upper tail probability of $X_{n,p}^H$ was initiated by the famous works of Kim and Vu [24], Vu [29], Janson and Ruciński [23]. This problem turns out to be so difficult\(^1\) because when $H$ is a connected graph with at least two edges $X_{n,p}^H$ is not a linear function of independent Bernoulli random variables. In the case of linear functions, life is easier and much is known about its large deviation properties, see [15].

In a famous paper [22], Janson, Oleszkiewicz, and Ruciński estimated the logarithm of the upper tail probability for every graph $H$ up to a multiplicative factor of $O(\log(1/p))$. Seven years later, Chatterjee [8] and DeMarco–Kahn [14] closed this gap up to a multiplicative factor of $O(1)$ when $H$ is a triangle and DeMarco–Kahn [13] generalized this for a clique of arbitrary size.

Next, one would like to obtain a first order approximation of the logarithm of the upper tail probability. The first to do that were Chatterjee and Varadhan [10]. They obtained a first order approximation of the upper tail probability for any graph, under the assumption that $p$ is a constant. Their proof relied on the regularity method and therefore extends only to $p$ tending to zero very slowly, for more discussion about this see [26].

The general strategy for estimating the logarithm of the upper tail probability, used by Chatterjee and Varadhan as well as all later works on this subject, is to establish a ‘large deviation principle’. That is to prove that the logarithm of the upper tail probability is asymptotically equal to the solution to a minimization problem over a non-trivial set of product probability measures. These minimization problems are called ‘variational problems’. After achieving such large deviation principle one is then left with solving the variational problem.

In a breakthrough paper of Chatterjee–Dembo [9], the authors established a ‘large deviation principle’ not only for $X_{n,p}^H$ when $p = \Omega(n^{-\epsilon})$, but also for a large class of functions of $N$ independent Bernoulli random variables with mean $p = \Omega(N^{-\epsilon})$. They proved that for any ‘smooth enough’ function $f : \{0, 1\}^N \to \mathbb{R}$ and a sequence of independent Bernoulli random variables $Y = (Y_1, Y_2, \ldots , Y_N) \in \{0, 1\}^N$, all with mean $p$, we have the following:

\[
-\log \mathbb{P}(\text{UT}(f(Y), \delta)) = (1 + o(1)) \inf \left\{ \sum_{i=1}^{N} I_p(q_i) : \mathbb{E}[f(\tilde{Y})] \geq (1 + \delta)\mathbb{E}[f(Y)] \right\},
\]

where $I_p(q) = q \log \left( \frac{q}{p} \right) + (1-q) \log \left( \frac{1-q}{1-p} \right)$ is the relative entropy (the Kullback–Leibler divergence) between Ber$(q)$ and Ber$(p)$, and $\tilde{Y} = (\tilde{Y}_1, \tilde{Y}_2, \ldots , \tilde{Y}_N) \in \{0, 1\}^N$ is a sequence of independent Bernoulli random variables with $\mathbb{E}[\tilde{Y}_i] = q_i$ for each $i$.

This was further developed by Eldan [16] and Augeri [2, 3]. These methods are completely different from the ones used in the dense regime.

Revisiting the ideas from Chatterjee–Varhadan [10], Cook–Dembo–Pham [12] developed a decomposition theorem similar to Szemerédi’s regularity lemma and a corresponding counting lemma which are suitable for sparse graphs. Using these they extended the range of sparsity of $p$ where the variational approximation (1) holds (for every graph $H$). Generalizing this method, Cook–Dembo–Pham [12] pushed the bounds even further in the case of subgraph count. They also obtained an approximation of the logarithm of the upper tail probability for the count of uniform sub-hypergraphs in the binomial random uniform hypergraph model, and the induced count of graphs in $G_{n,p}$. Their result combined with the solution to the corresponding variational problem for uniform hypergraph cliques, and one

\(^1\)This problem was called ‘the infamous upper tail’ by Janson and Ruciński [23].
more non-trivial 3-uniform hypergraph which were given by Liu–Zhao in [25], yields estimations for the sub-hypergraph count in the binomial random hypergraph model in the sparse regime. We also note that to the best of our knowledge the solution to the ‘induced variational problem’ is not known apart from the case where $H$ is a clique. The case of the induced subgraph count will be of main interest in this paper, and will be discussed later.

To estimate the asymptotics of the logarithm of the upper tail probability, one also needs to solve the variational problem in the right-hand side of (1). In the case of the random variable $X^H_{n,p}$, Lubetzky and Zhao [27], and Bhattacharya, Ganguly, Lubetzky, and Zhao [5] solved this variational problem for all $H, n$ and $p$ satisfying $n^{-1/\Delta_H} \ll p \ll 1$. This then leads to a first order approximation of the logarithm of the upper tail probability for $X^H_{n,p}$ for every graph $H$ and $p \geq n^{-c_H}$. We wish to emphasize that, in all known cases, when $f$ counts the number of copies of a given $H$ in $G_{n,p}$, the solution to the corresponding variational problem (the right hand side of (1)) always satisfies $q_i \in \{p, 1\}$ when $p = o(1)$.

A recent surprising result due to Gunby [20] studied the upper tail probability for subgraph counts in a random regular graph $G_{n,d}$, where he also proved a large deviation principle. However, for a particular graph, and certain range of $d$ (the regularity) the answer for this variational problem is achieved with three possible values of $q_i$ and not two as in the binomial random graph model.

In a breakthrough paper of Harel, Mousset and Samotij [21], using a combinatorial approach, the authors managed to extend the range where the variational problem (the same one as before) bounds from above the logarithm of the upper tail probability for the count of cliques to the optimal range of $p$, which is $p^{\frac{r-1}{2}} \gg n^{-1}(\log(n))^{\frac{r-2}{2}}$ for a clique on $r$ vertices$^2$. We wish to emphasize that their approach for solving the upper tail problem is completely different from any previous techniques in the papers mentioned above. This, together with the solution to the variational problem, settled the problem of the first order approximation of the logarithm of the upper tail probability for cliques. Moreover, they established the correct range of $p$, up to a polylogarithmic factor, where the variational problem bounds from above the logarithm of the upper tail probability for non-bipartite regular graphs$^3$.

Building on the combinatorial ideas of Harel, Mousset and Samotij, Basak and Basu [4] extended the range of $p$ to the optimal range for all regular graphs, including bipartite graphs. Again, this with the solution to the variational problem settled the problem of estimating the logarithm of the upper tail probability for regular graphs. We summarize this discussion with the following first order approximation of the upper tail problem for regular graphs in the ‘localized regime’. The optimality of the following theorem follows below the lower threshold, the upper tail probability is of different type (resembles that of a Poisson random variable, see [21]). Further, the upper threshold is also needed, and this was considered in [10]. Let us also note that below the lower threshold (up to logarithmic factors) in the following theorem we have $\mathbb{E}[X^H_{n,p}] = o(1)$.

**Theorem 1.1** (Due to [4, 21]). For any $\Delta \geq 2$ and a $\Delta$-regular connected graph $H$ the following holds:

$$
-\log \mathbb{P}(X^H_{n,p} \geq (1 + \delta)\mathbb{E}[X^H_{n,p}]) = (1 + o(1)) \begin{cases} 
\min\{\theta_H, \frac{1}{2} \delta^{2/\nu_H}\} & \text{if } \frac{1}{\sqrt{n}} \ll p^{\Delta/2} \ll 1, \\
\frac{1}{2} \delta^{2/\nu_H} & \text{if } \frac{\log^{\nu_H - 2}(n)}{n} \ll p^{\Delta/2} \ll \frac{1}{\sqrt{n}}.
\end{cases}
$$

$^2$That is because below this threshold they also showed a Poisson approximation.

$^3$For bipartite regular graphs their result was not optimal, but also a lot of progress was made.
where $\theta_H$ is the unique solution to the equation $P_H(\theta) = 1 + \delta$ where $P_H$ is the independence polynomial$^4$ of $H$.

Before proceeding to the main result of this paper, let us give some intuition for Theorem 1.1. A strategy to bound from below the upper tail probability is to observe that, conditioned on the event $C \subseteq G_{n,p}$ for some $C$ satisfying $\mathbb{E}[X_{n,p}^H | C \subseteq G_{n,p}] \geq (1 + \delta)\mathbb{E}[X_{n,p}^H]$, the upper tail event holds with `decent’ probability. This allows us to bound the upper tail probability from below by the probability of $C \subseteq G_{n,p}$ (which is $p^{|C|}$) times the `decent’ probability above (which is $p^{|\theta_H(C)|}$). We refer to this as ‘planting a copy of $C$’. Of course one would like to consider the smallest such $C$, to increase the lower bound as much as possible. For $\Delta$-regular graphs there are two natural candidates for $C$: First, a clique on $\delta^1/n$ vertices, and second, a spanning complete bipartite graph with the smaller side of size $\Theta(np^\Delta)$. The second construction is often called a Hub. Indeed, conditioning on the existence of the clique in $G_{n,p}$ the above condition holds. That is because we get a factor of $\delta \mathbb{E}[X_{n,p}^H]$ copies of $H$ in the clique. For the Hub the situation is a bit different. For example, if $H$ is not bipartite, then the Hub does not contain even a single copy of $H$. Nevertheless, in expectation the Hub will boost the number of copies of $H$ by a constant multiplicative factor. To see this, let $J = (A, B)$ be a spanning bipartite subgraph of $H$ with sides $A$ and $B$, where $A$ consists of vertices of degree $\Delta$. Note that the number of edges in $H$ and not in $J$ is exactly $\Delta(|B| - |A|)/2$. Therefore, each copy of $J$ contained in the Hub will boost the expected number of copies of $H$ by $\Theta(p^\Delta(|B| - |A|)/2)$. Note that there are $\Theta(n^{|A|+|B|}p^{|A|\Delta})$ ways to embed $J$ in the Hub, where $|A|$ is placed in the small side and $B$ is placed in the large side. All together yields a boost of the order of $\Theta(n^2p^\Delta)$ copies of $H$ in expectation, as we wished for. What was essentially proven in Theorem 1.1 is that the logarithm of the upper tail probability is asymptotically equal to the logarithm of the probability of the existence of either the clique or the Hub. This explains both the order of magnitude of the logarithm of the upper tail probability, the minimum in the normalized limit, and the different behaviour depending on the sparsity of $p$. Indeed, the number of edges of both the clique and the Hub is $\Theta(n^2p^\Delta)$, and hence the logarithm of the probability of containing either one is $\Theta(n^2p^\Delta \log(p))$ — this is the order of magnitude of the logarithm of the upper tail probability. The minimum in the normalized limit, stands for the two competing examples. Lastly, we note that when $p$ is sufficiently sparse so that $np^\Delta = o(1)$, the Hub construction is no longer valid, as then the small side of the Hub cannot be an integer.

With this intuition in mind, let us turn to the main concern of this paper. Our main result in this paper is an asymptotic estimate of the logarithm of the upper tail probability for the count of induced copies of $C_4$ in $G_{n,p}$ in the localized regime up to logarithmic factors. To the best of our knowledge, this is the first$^5$ exact result for the upper tail probability of a random variable counting the number of induced copies of a given non-complete graph in $G_{n,p}$. Our proofs rely heavily on the combinatorial approach of Harel, Mousset and Samotij.

From now on we let $X$ to be the random variable counting the number of induced copies of $C_4$ in $G_{n,p}$. Before presenting the main result of this paper, let us note that the order of magnitude of the the logarithm of the upper tail probability of $X$ is clearly $n^2p^2 \log(p)$ in the localized regime. The upper bound follows from the Theorem 1.1, the lower bound can be easily seen as planting a complete and balanced bipartite graph with sides of size $\Theta(np)$ boosts $\mathbb{E}[X]$ by a lot. Let us now state the main result of this paper.

$^4$The independent polynomial of $H$ is $P_H(x) = \sum_{k=0}^{\Delta} s_k X^k$ where $s_k$ is the number of independent sets in $H$ of size $k$.

$^5$An independent result regarding induced counts was obtained by Cook–Dembo–Pham [12].
Theorem 1.2. There is an explicit sequence \( 0 = c_1 < c_2 < \ldots \leq 1/3 \) such that the following holds for \( p \gg \frac{\log^q(n)}{n} \):

\[
-\log \mathbb{P}(X \geq (1+\delta)\mathbb{E}[X]) \frac{n^2p^2 \log(1/p)}{n^2p^2 \log(1/p)} = (1+o(1)) \begin{cases} 
\rho_k(n,p) \sqrt{\frac{2}{k}} & \text{if } n^{-1+c_{k-1}} \leq p \leq n^{-1+c_k}, \\
\sqrt{\frac{2}{2} + \frac{1}{16}} - \frac{1}{\sqrt{16}} & \text{if } n^{-1/2} \ll p \ll \frac{n^{1/2}}{\log(n)}, \\
\sqrt{\frac{2}{2} + \frac{1}{16}} - \frac{1}{\sqrt{16}} & \text{if } n^{-1/2} \ll p \ll 1,
\end{cases}
\]

where \( \rho_k(n,p) = \sqrt{\frac{k}{k-1}} \left( 1 - \frac{2}{k} + \frac{\log(n)}{k \log(p)} \right) < 1 \).

The sequence above is explicitly defined in Section 6.

There are several interesting phenomena which distinguish Theorem 1.2 from previous results concerning the upper tail problem for subgraph counts.

The first, and most noticeable difference is the infinite number of phase transitions. To the best of our knowledge, there is no earlier example of an upper tail problem exhibiting infinitely many phase transitions in its first order approximation. To understand the reason for this, let us compare the situation with the Theorem of 1.1, and in particular the graphs that are being planted to obtain a lower bound on the logarithm of the upper tail probability. Even though \( C_4 \) is a regular graph, the presence of a clique in \( G_{n,p} \) does not boost the expected number of induced 4-cycles by a significant amount. Therefore, the construction of a clique is no longer valid. The analog of this construction in our case is a balanced and complete bipartite graph with \( \Theta(np) \) vertices. Surprisingly, although amongst all graphs with a given number of edges the complete and balanced bipartite graph maximizes the number of induced copies of \( C_4 \) (see [7, 19] and Lemma 6.2) planting it does not always give the strongest lower bound on the upper tail probability.

Informally, this happens because in a wide range of densities of \( p \), there is a large set of graphs boosting the expected number of copies of \( G \) by a significant amount. This set is so large so that the logarithm of the probability of containing some graph from this family is asymptotically larger than the logarithm of the probability of containing any particular subgraph in the family. More formally, one can define \( \mathcal{K}_p \) to be the set of subgraphs of \( K_n \), where each \( C \in \mathcal{K}_p \) satisfies \( \mathbb{E}[X|C \subseteq G_{n,p}] \geq (1+\delta)\mathbb{E}[X] \) and has only slightly suboptimal size (amongst all \( C \) with this property). I turns out that \( \mathcal{K}_p \) is so ‘large’ that

\[
\log \mathbb{P}(C \subseteq G_{n,p} \text{ for some } C \in \mathcal{K}_p) \geq -(1 - \Omega(1))\min_C \mathbb{E}(C) \log(1/p),
\]

where the minimum ranges over all graphs \( C \) with \( \mathbb{E}[X|C \subseteq G_{n,p}] \geq (1+\delta)\mathbb{E}[X] \); moreover, conditioned on the event that \( C \subseteq G_{n,p} \) for some \( C \in \mathcal{K}_p \), the upper tail event still holds with ‘decent’ probability. We wish to emphasize that this is very different from the case of non-induced regular graphs, as here we condition on the appearance of some graph from \( \mathcal{K}_p \) rather than a single graph (which corresponds to tilting the measure of \( G_{n,p} \) to \( G_{n,\bar{q}} \) for some \( \bar{q} \in (p, 1) \)). In fact, we will show (in Section 7) that the lower bound obtained with the use of \( \mathcal{K}_p \) is stronger than any bound corresponding to the more general tilting \( G_{n,p} \) to \( G_{n,\bar{q}} \) for some \( \bar{q} \in [0, 1) \).

Let us elaborate on the structure of the graphs in \( \mathcal{K}_p \). The family \( \mathcal{K}_p \) depends on \( p \) in the following way: provided \( n^{-1+c_{k-1}} \leq p \leq n^{-1+c_k} \) for some integer \( k \geq 2 \), the set \( \mathcal{K}_p \) comprises all complete bipartite graphs with sides \( k \) and \( \ell = 2 \sqrt{\frac{\delta \mathbb{E}[X]}{k(k-1)}} \). Note that, for any fixed \( k \) and large \( n \), the graph \( K_{k,\ell} \) has more
edges than the complete and balanced bipartite graph with the same number of induced copies of $C_4$. However, the key point here is that we are no longer planting a single copy of this graph. Since $\ell$ is large, there are many embeddings of $K_{k,\ell}$ in $K_n$ and the probability that $G_{n,p}$ contains some such copy is affected by the combinatorial factor of the number of possible embeddings. It turns out that the lower bound obtained by considering these families is actually tight in the range where $\frac{\log^3(n)}{n} \ll p \ll n^{-2/3-o(1)}$. Further, between $\frac{\log^3(n)}{n}$ and $n^{-2/3-o(1)}$ the family $K_p$ changes infinitely many times depending on $p$, which explains the infinite number of phase transitions. This will be explained in more details in Section 7. When $p$ is denser, the ‘entropic’ contribution mentioned above becomes negligible. Then one can consider just planting a complete and balanced bipartite graph with sides of sizes $\Theta(np)$ when $n^{-2/3} \leq p \ll n^{-2/3}$ the Hub becomes available and this is the construction we plant, that is a complete bipartite graph between a set of size $\Theta(np^2)$ and the rest of the graph.

Organisation of the paper

In Section 2, we prove a general large deviation principle for nonnegative functions of independent Bernoulli variables which is a version of Theorem 9.1 in [21], which was stated in [21] but was not proven. In Section 3, we continue by connecting the general large deviation principle to our large deviation problem. To this end we define special classes of graphs called core graphs. Then we modify the general result proved in Section 2 using this notion of core graphs.

Section 4 is independent of the others and gives lower bounds for the upper tail probability by planting graphs in the denser regimes and families of graphs in the sparser regimes.

In Section 5, we solve the variational problem presented in Section 3. Furthermore, we make a connection between the number of core graphs with $m$ edges and the maximum number of vertices a core graph with $m$ edges might have.

Section 6 uses the results from Section 5 to provide upper bounds for the logarithm of the upper tail probability. This section is divided into three parts. First is the dense regime $n^{-1/2} \ll p \ll 1$, second is the sparse regime $\frac{\log^3(n)}{n} \ll p \ll \frac{1}{\sqrt{n \log(n)}}$ which we also split into two cases: the dense case in the sparse regime $n^{-2/3} \ll p \ll \frac{1}{\sqrt{n \log(n)}}$ and the sparser cases in the sparse regime $n^{-1+c_3} \ll p \ll n^{-1+c_4}$. Before the second split, we develop some tools used in both cases.

In Section 7 we solve the naive mean field variational problem, showing that it is different the solution to the variational problem we used.

Appendix A was added for completeness, where we reprove results of [5, 27] in a language of graphs rather than graphons.

2 | MAIN TOOLS–POLYNOMIALS IN THE HYPERCUBE

We start with some notations and definitions which can also be found in [21]. We work within the probability space $(\{0, 1\}^N, \text{Ber}(p)^N)$. Suppose $I \subseteq [N]$ and $x \in \{0, 1\}^N$. We say that

$$F(I, x) = \{ y \in \{0, 1\}^N : y_n = x_n \text{ for all } n \in I \}$$

is a subcube centered at $(I, x)$ with codimension $|I|$ denoted by codim($F$) = $|I|$. Note that every subcube is centered at some pair $(I, x)$. Moreover, if a subcube $F$ is centered at $(I, x)$ and it is also centered at $(J, y)$ then $I = J$. Hence, the codimension is well defined. For every subcube $F$ centered at $(I, x)$ we define the one-supcube and the zero-supcube of $F$ to be $F^{(1)} = F(I_1, x)$ and $F^{(0)} = F(I_0, x)$ where
Let $I_i = \{n \in I : x_n = i\}$ for $i = 0, 1$. Moreover, for every subcube $F$ we say that the one-codimension of it is the codimension of $F^{(1)}$ and the zero-codimension of it is the codimension of $F^{(0)}$ denoted by $\text{codim}(F)$ for $i = 1$ and $i = 0$ respectively. A simple observation is that any subcube $F$ always satisfies $F = F^{(0)} \cap F^{(1)}$ and $\text{codim}(F) = \text{codim}(F^{(0)}) + \text{codim}(F^{(1)})$. For every $X : \{0, 1\}^N \to \mathbb{R}_{\geq 0}$ we define the complexity of $X$ denoted by $\text{comp}(X)$ to be the smallest integer $d$ for which it is possible to represent $X$ as a linear combination with nonnegative coefficients of indicator functions of subcubes with codimension at most $d$. Note that the complexity is well defined as for every such function $X$ we can write

$$X = \sum_{x \in \{0, 1\}^N} X(x) \mathbb{1}_{F(x)}.$$  

Moreover, this shows that for any $X$ we have $\text{comp}(X) \leq N$. Assume now that $Y$ is a random variable taking values in $\{0, 1\}^N$ and that $X = X(Y)$. Given a subcube $F \subseteq \{0, 1\}^N$, we write $E_F[X] = E[X|Y \in F]$ for the expectation of $X$ conditioned on $Y \in F$. We further define $\Phi_X : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ by

$$\Phi_X(\delta) = \min\{-\log P(Y \in F) : F \subseteq \{0, 1\}^N \text{ is a subcube with } E_F[X] \geq (1 + \delta)E[X]\}.$$ 

Our main tool is [Theorem 9.1][21], which was stated but not proved. We prove a slightly different version of the theorem, but the essence remains the same.

**Theorem 2.1.** For every positive integer $d$ and all positive real numbers $\epsilon$ and $\delta$ with $\epsilon < \frac{1}{2}$, there is a positive constant $K = K(\epsilon, \delta, d)$ such that the following holds. Let $Y$ be a sequence of $N$ independent $\text{Ber}(p)$ random variables for some $p \in (0, 1/2)$ and assume that $X = X(Y)$ is nonnegative with complexity at most $d$ and satisfies $\Phi_X(\delta - \epsilon) \geq K \log(\frac{1}{p})$. Denote by $F$ the collection of all subcubes $F \subseteq \{0, 1\}^N$ satisfying

(F1)  $E_F[X] \geq (1 + \delta - \epsilon)E[X]$,

(F2)  $\text{codim}(F) \leq K \cdot \Phi_X(\delta + \epsilon)$.

Then,

$$P(X \geq (1 + \delta)E[X]) \leq (1 + \epsilon)P(Y \in F \text{ for some } F \in F).$$

To prove the theorem we state and prove the following lemmas.

**Lemma 2.2.** Let $Y$ be a random variable taking values in $\{0, 1\}^N$ and let $X = X(Y)$ be a real-valued function of $Y$. Suppose that $E[X] > 0$ and that $X \leq M$ always. Then for all positive $\epsilon$ and $\delta$,

$$-\log P(X \geq (1 + \delta)E[X]) \leq \Phi_X(\delta + \epsilon) + \log \left( \frac{M}{\epsilon E[X]} \right).$$

**Proof.** Let $t = (1 + \delta)E[X]$. If $\Phi_X(\delta + \epsilon) = \infty$ then the claim is vacuously true. Otherwise, there exists a subcube $F$ such that $-\log P(Y \in F) = \Phi_X(\delta + \epsilon)$ and $E_F[X] \geq (1 + \delta + \epsilon)E[X] = t + \epsilon E[X]$. By $X \leq M$ we have $E_F[X] \leq M \cdot P(X \geq t|Y \in F) + t$. Thus, $P(X \geq t|Y \in F) \geq \frac{\epsilon E[X]}{M}$ and hence,

$$P(X \geq t) \geq P(X \geq t|Y \in F)P(Y \in F) \geq \frac{\epsilon E[X]}{M} P(Y \in F).$$

Now taking the negative logarithm gives the assertion of the lemma. \qed
Fact 2.3. Suppose $F_1, \ldots, F_k$ are subcubes of $\{0, 1\}^N$. If $F_1 \cap \cdots \cap F_k$ is nonempty, then it is also a subcube of $\{0, 1\}^N$ and, moreover, $\text{codim}(F_1 \cap \cdots \cap F_k) \leq \sum_{i=1}^k \text{codim}(F_i)$.

Proof. We prove the statement for $k = 2$; the case $k > 2$ follows by a simple inductive argument. Let $I_1, I_2$ and $x_1, x_2$ be such that $F_1$ is a subcube centered at $(I_i, x_i)$ for $i = 1, 2$. Define $x \in \{0, 1\}^N$ in the following way: for all $i \in I_1$ and $j \in I_2$ put $(x)_i = (x_1)_i$ and $(x)_j = (x_2)_j$ and for all $i \notin I_1 \cup I_2$ put $(x)_i = 0$. This is indeed a well defined element in $\{0, 1\}^N$ as for all $i \in I_1 \cap I_2$ we have, $(x_1)_i = (x_2)_i$, otherwise it will contradict our assumption that $F_1 \cap F_2 \neq \emptyset$. It follows from the definition of $x$ that $F_1 \cap F_2 = F(I_1 \cup I_2, x)$ and hence $F_1 \cap F_2$ is a subcube. Furthermore, we have

$$\text{codim}(F_1 \cap F_2) = |I_1 \cup I_2| \leq |I_1| + |I_2| = \text{codim}(F_1) + \text{codim}(F_2).$$

Lemma 2.4. Let $Y$ be a random variable taking values in $\{0, 1\}^N$ and let $X = X(Y)$ be a nonnegative real-valued function with complexity bounded by $d$. Then for every positive integer $\ell$ and all positive real numbers $\epsilon$ and $\delta$ with $\epsilon < 1 + \delta$,

$$\mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X] \text{ and } Y \notin F \text{ for all } F \in \mathcal{F}) \leq \left(\frac{1 + \delta - \epsilon}{1 + \delta}\right)^\ell,$$

where

$$\mathcal{F} = \{F \in \{0, 1\}^N : F \text{ is a subcube and } \text{codim}(F) \leq d\ell \text{ and } \mathbb{E}_F[X] \geq (1 + \delta - \epsilon)\mathbb{E}[X]\}.$$

Proof. Given $S \subseteq \{0, 1\}^N$ a subcube, let $Z_S$ be the indicator random variable of the event that $Y \notin F$ for all $F \in \mathcal{F}$ with $F \supseteq S$. Note that $S \subseteq S'$ implies $Z_S \leq Z_{S'}$ and let $Z = Z_\emptyset$. Since $XZ \geq 0$ and $Z^{\ell'} = Z$, Markov’s inequality gives

$$\mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X] \text{ and } Z = 1) = \mathbb{P}(XZ \geq (1 + \delta)\mathbb{E}[X]) \leq \frac{\mathbb{E}[X^{\ell'} Z]}{(1 + \delta)\mathbb{E}[X]^{\ell'}}. \quad (2)$$

To simplify the notation, for every $S \subseteq \{0, 1\}^N$ we write $1_S$ for $1_{\{y \in S\}}$. Write $X = \sum_F \alpha_F 1_F$, where the sum ranges over all subcubes of $\{0, 1\}^N$, each coefficient $\alpha_F$ is nonnegative, and $\alpha_F = 0$ for all $F$ with $\text{codim}(F) > d$. Then for every $k \in [\ell]$,

$$\mathbb{E}[X^k Z] = \sum_{F_1, \ldots, F_k} \alpha_{F_1} \cdots \alpha_{F_k} \mathbb{E}[1_{F_1} \cdots 1_{F_k} \cdot Z]$$

$$\leq \sum_{F_1, \ldots, F_k} \alpha_{F_1} \cdots \alpha_{F_k} \mathbb{E}[1_{F_1} \cdots 1_{F_k} \cdot Z_{F_1 \cap \cdots \cap F_k}]$$

$$\leq \sum_{F_1, \ldots, F_{k-1}} \alpha_{F_1} \cdots \alpha_{F_{k-1}} \mathbb{E}[1_{F_1} \cdots 1_{F_{k-1}} \cdot Z_{F_1 \cap \cdots \cap F_{k-1}}] \cdot \mathbb{E}[X|1_{F_1} \cdots 1_{F_{k-1}} = 1],$$

where we may let the third sum range only over sequences $F_1, \ldots, F_{k-1}$ for which the event $\{1_{F_1} \cdots 1_{F_{k-1}} \cdot Z_{F_1 \cap \cdots \cap F_{k-1}} = 1\}$ has a positive probability of occurring.
Claim 2.5. For any such sequence,
\[ F_1 \cap \ldots \cap F_{k-1} \notin \mathcal{F} \quad \text{and} \quad 1_{F_1} \cdots 1_{F_{k-1}} \cdot Z_{F_1 \cap \ldots \cap F_{k-1}} = 1_{F_1} \cdots 1_{F_{k-1}}. \]

Proof. To see this, note that if \( \{ 1_{F_1} \cdots 1_{F_{k-1}} \cdot Z_{F_1 \cap \ldots \cap F_{k-1}} = 1 \} \) has positive probability of occurring, then there exists a \( y \) such that \( y \in F_1 \cap \ldots \cap F_{k-1} \) and \( Z_{F_1 \cap \ldots \cap F_{k-1}}(y) = 1 \). That means \( y \in F_1 \cap \ldots \cap F_{k-1} \) and \( y \notin F \) for any subcube \( F \in \mathcal{F} \) such that \( F \supseteq F_1 \cap \ldots \cap F_{k-1} \). Hence, \( F_1 \cap \ldots \cap F_{k-1} \notin \mathcal{F} \). For the second part assume towards a contradiction that \( 1_{F_1} \cdots 1_{F_{k-1}} \cdot Z_{F_1 \cap \ldots \cap F_{k-1}} < 1_{F_1} \cdots 1_{F_{k-1}} \), meaning that there is \( y' \in F_1 \cap \ldots \cap F_{k-1} \) such that \( Z_{F_1 \cap \ldots \cap F_{k-1}}(y') = 0 \). Therefore, there exists a subcube \( F^* \in \mathcal{F} \) such that, \( F^* \supseteq F_1 \cap \ldots \cap F_{k-1} \). This is a contradiction as now, \( y \in F_1 \cap \ldots \cap F_{k-1} \subseteq F^* \) but by our assumption, \( y \notin F^* \) as \( F^* \in \mathcal{F} \).

Since \( \{ 1_{F_1} \cdots 1_{F_{k-1}} = 1 \} \) has a positive probability of occurring, Fact 2.3 asserts that \( F_1 \cap \ldots \cap F_{k-1} \) is a subcube, and \( \text{codim}(F_1 \cap \ldots \cap F_{k-1}) \leq d(k-1) \leq d\ell \). Therefore
\[ \mathbb{E}[X | 1_{F_1} \cdots 1_{F_{k-1}} \cdot Z_{F_1 \cap \ldots \cap F_{k-1}} = 1] = \mathbb{E}_{F_1 \cap \ldots \cap F_{k-1}}[X] < (1 + \delta - \varepsilon)\mathbb{E}[X], \]
as otherwise \( F_1 \cap \ldots \cap F_{k-1} \) would belong to \( \mathcal{F} \). It follows that
\[
\sum_{F_1, \ldots, F_k} \alpha_{F_1} \cdots \alpha_{F_k} \mathbb{E}[1_{F_1} \cdots 1_{F_k} \cdot Z_{F_1 \cap \ldots \cap F_k}] < (1 + \delta - \varepsilon)\mathbb{E}[X] \sum_{F_1, \ldots, F_{k-1}} \alpha_{F_1} \cdots \alpha_{F_{k-1}} \mathbb{E}[1_{F_1} \cdots 1_{F_{k-1}} \cdot Z_{F_1 \cap \ldots \cap F_{k-1}}].
\]
By induction, we see that \( \mathbb{E}[X^\varepsilon Z] < ((1 + \delta - \varepsilon)\mathbb{E}[X])^\varepsilon \). Substituting this inequality into \( (2) \) completes the proof.

Now we use the previous lemmas to prove the theorem.

Proof of Theorem 2.1. Let \( K \) be a large constant that may depend on \( \varepsilon, \delta \) and \( d \). Furthermore, let \( t = (1 + \delta)\mathbb{E}[X] \) and \( \ell = \lfloor K/d \cdot \Phi_X(\delta + \varepsilon) \rfloor \). Define
\[ \mathcal{F} = \{ F \subseteq \{0,1\}^N : F \text{ is a subcube with codim}(F) \leq d\ell \quad \text{and} \quad \mathbb{E}_F[X] \geq (1 + \delta - \varepsilon)\mathbb{E}[X] \}. \]
It follows from Lemma 2.4 that
\[ \mathbb{P}(X \geq t \quad \text{and} \quad Y \notin F \quad \text{for all} \quad F \in \mathcal{F}) \leq \left( 1 - \frac{\varepsilon}{1 + \delta} \right)^\ell. \]
As \( \text{comp}(X) \leq d \), we can write
\[ X = \sum_{j \in J} \alpha_j 1_{\{ Y \in F_j \}} \]
for some \( J \) so that, for all \( j \in J \), the coefficient \( \alpha_j \) is nonnegative and \( F_j \) is a subcube with \( \text{codim}(F_j) \leq d \). Put \( M = \sum \alpha_j \) and note that \( X \leq M \) always. Applying Lemma 2.2 gives
\[ -\log \mathbb{P}(X \geq t) \leq \Phi_X(\delta + \varepsilon) + \log \left( \frac{M}{\varepsilon \mathbb{E}[X]} \right). \]
Note also that $\mathbb{E}[X] \geq Mp^d$ as $\text{comp}(X) \leq d$ and therefore $\mathbb{P}(Y \in F_j) \geq \min\{p, 1-p\}^d = p^d$ for every $j \in J$. Therefore,

$$- \log \mathbb{P}(X \geq t) \leq \Phi_X(\delta + \varepsilon) + \log \left( \frac{1}{\varepsilon p^d} \right).$$

Thus, provided $K$ is sufficiently large we also have

$$- \log \mathbb{P}(X \geq t) \leq (1 + \varepsilon)\Phi_X(\delta + \varepsilon),$$

as we assumed that $\Phi_X(\delta + \varepsilon) \geq \Phi_X(\delta - \varepsilon) \geq K \log(1/p)$. Putting all of this together gives that for sufficiently large $K$, we have

$$\mathbb{P}(X \geq t \text{ and } Y \notin F \text{ for all } F \in \mathcal{F}) \leq \left(1 - \frac{\varepsilon}{1 + \delta}\right)^{\ell} \leq e^{-\frac{\varepsilon}{\ell} \cdot 1_{\mathcal{F}}[K/\delta]} \Phi_X(\delta + \varepsilon)[1]
\leq (\varepsilon/2) \cdot \mathbb{P}(X \geq t).$$

The assertion of the theorem now follows:

$$\mathbb{P}(X \geq t) \leq (1 - \varepsilon/2)^{-1} \cdot \mathbb{P}(Y \in F \text{ for some } F \in \mathcal{F}) \leq (1 + \varepsilon)\mathbb{P}(Y \in F \text{ for some } F \in \mathcal{F})$$

where the last inequality holds for $\varepsilon < 1/2$.

When $p \ll 1$ and $\text{comp}(X) = O(1)$, the majority of the contribution to both $\mathbb{E}_F[X]$ and $\mathbb{P}(Y \in F)$ comes from the one-supcube $F^{(1)} \supseteq F$. We prove a straightforward corollary of Theorem 2.1 which will be more convenient to work with. We start by proving the following lemma.

**Lemma 2.6.** The following holds for every positive integer $d$ and all positive real numbers $\varepsilon$ and $\delta$ with $\varepsilon < 1$. Let $Y$ be a sequence of $N$ independent $\text{Ber}(p)$ random variables for some $p < 1 - \left(\frac{1 + \delta - \varepsilon}{1 + \delta - \varepsilon/2}\right)^{1/d}$ and assume that $X = X(Y)$ has complexity at most $d$. Let $F$ be a subcube satisfying $\mathbb{E}_F[X] \geq (1 + \delta - \varepsilon/2)\mathbb{E}[X]$. Then, $\mathbb{E}_{F^{(1)}}[X] \geq (1 + \delta - \varepsilon)\mathbb{E}[X]$ where $F^{(1)}$ is the one-supcube of $F$.

**Proof.** As $X$ has complexity $d$ we can write

$$X = \sum_{j \in J} \alpha_j 1_{\{Y \in F_j\}}$$

for some $J$ so that, for all $j \in J$, the coefficient $\alpha_j$ is nonnegative and $F_j$ is a subcube with $\text{codim}(F_j) \leq d$. Moreover, by our assumptions on $F_j$, we have $\text{codim}(F_j) < \text{codim}(F_j) \leq d$. Thus, we have the following inequality,

$$\mathbb{P}\left(Y \in F_j \mid Y \in F^{(1)}\right) \geq (1 - p)^d \cdot \mathbb{P}(Y \in F_j \mid Y \in F).$$

It follows that

$$\mathbb{E}_{F^{(1)}}[X] = \sum_{j \in J} \alpha_j \mathbb{E}_{F^{(1)}}[1_{\{Y \in F_j\}}] \geq \sum_{j \in J} \alpha_j (1 - p)^d \cdot \mathbb{E}_F[1_{\{Y \in F_j\}}]$$

$$\geq \sum_{j \in J} \alpha_j \left(\frac{1 + \delta - \varepsilon}{1 + \delta - \varepsilon/2}\right) \mathbb{E}_F[1_{F_j}] = \left(\frac{1 + \delta - \varepsilon}{1 + \delta - \varepsilon/2}\right) \mathbb{E}_F[X] \geq (1 + \delta - \varepsilon)\mathbb{E}[X], \quad (3)$$
where the first inequality in (3) follows from the assumption on \( p \) and the second inequality in (3) follows from the assumption on \( \mathbb{E}_F[X] \).

Using this lemma we now state and prove a straightforward corollary of Theorem 2.1.

**Corollary 2.7.** For every positive integer \( d \) and all positive real numbers \( \epsilon \) and \( \delta \) with \( \epsilon < 1 \), there is a positive constant \( K = K(\epsilon, \delta, d) \) such that the following holds. Let \( Y \) be a sequence of \( N \) independent \( \text{Ber}(p) \) random variables for some \( p < \min\{1 - \left(\frac{1 + \delta - \epsilon}{1 + \delta - \epsilon/2}\right)^{1/d}, 1/2\} \) and assume that \( X = X(Y) \) has complexity at most \( d \) and satisfies \( \Phi_X(\delta - \epsilon) \geq K \log(\frac{1}{p}) \).

Denote by \( \mathcal{F}_1 \) the collection of all subcubes \( F \subseteq \{0, 1\}^N \) satisfying

\[
(H1) \quad \mathbb{E}_F[X] \geq (1 + \delta - \epsilon) \mathbb{E}[X],
\]

\[
(H2) \quad \text{codim} F \leq K \cdot \Phi_X(\delta + \epsilon),
\]

\[
(H3) \quad \text{codim}_1(F) = \text{codim}(F).
\]

Then

\[
\mathbb{P}(X \geq 1 + \delta) \mathbb{E}[X]) \leq (1 + \epsilon) \mathbb{P}(Y \in F \text{ for some } F \in \mathcal{F}_1).
\]

**Proof.** Applying Theorem 2.1 with \( \epsilon \) replaced by \( \epsilon/2 \) we obtain

\[
\mathbb{P}(X \geq 1 + \delta) \mathbb{E}[X]) \leq (1 + \epsilon) \mathbb{P}(Y \in F \text{ for some } F \in \mathcal{F}),
\]

where \( \mathcal{F} \) is the collection of all subcubes \( F \subseteq \{0, 1\}^N \) satisfying \((F1)\) and \((F2)\) where \( \epsilon \) is replaced with \( \epsilon/2 \). Letting \( \mathcal{F}_1 = \{ F^{(1)} : F \in \mathcal{F} \} \), we obtain \((H2)\) and \((H3)\) for every subcube \( F \in \mathcal{F}_1 \) due to \((F2)\) and the definition of one-supcubes. Noting that for every subcube \( F \) we have \( F \subseteq F^{(1)} \) we obtain also that

\[
\mathbb{P}(Y \in F \text{ for some } F \in \mathcal{F}) \leq \mathbb{P}(Y \in F \text{ for some } F \in \mathcal{F}_1).
\]

Combining (4) and (5) we obtain that

\[
\mathbb{P}(X \geq 1 + \delta) \mathbb{E}[X]) \leq (1 + \epsilon) \mathbb{P}(Y \in F \text{ for some } F \in \mathcal{F}_1).
\]

Furthermore, Lemma 2.6 implies that \( \mathcal{F}_1 \) also satisfies \((H1)\). \[\square\]

We call a subcube \( F \) a **seed** if \( F \) is an element of \( \mathcal{F}_1 \) defined in Corollary 2.7. A formal definition is given in the following section. A general way one would use Corollary 2.7 to bound upper tails of counts of induced subgraphs is to define a special type of seeds called **structured seeds**. These structured seeds are subcubes in \( \mathcal{F}_1 \) from Corollary 2.7, with a stronger condition than \((H1)\). This condition is a combinatorial condition involving various supcubes of the subcubes in \( \mathcal{F}_1 \) (this will be explained further in Section 3). Then one would define a **core subcube** to be a subcube containing a structured seed such that ‘every coordinate counts’. In the following section we define these special subcubes. We will also show that Corollary 2.7 can be modified to bound the upper tail probability via the probability of \( Y \) being an element in a core subcube.

### 3 FROM SEEDS TO MODIFIED CORES

Note that for any subcube \( F \subseteq \{0, 1\}^{E(K_n)} \), \( F^{(1)} \) the one-supcube of \( F \) corresponds to all subgraphs of \( K_n \) containing a specific subgraph of \( K_n \) with \( \text{codim}_1(F) \) edges. Therefore, from this point onwards we
think of one-supcubes as a family of subgraphs of $K_n$ containing a specific subgraph. In particular, instead of writing $E_f[X]$ for some one-supcube (of some subcube) we will write $E[G][X]$ for the graph corresponding to $F$. The subgraphs corresponding to the members of $P_1$ from Corollary 2.7 will be called seeds. We start with a definition which will be useful in this section as well as the next ones.

**Definition.** Suppose $H$ and $G$ are graphs and let $e$ be an edge of $G$. We define:

- $N_{ind}(H, G)$ is the number of induced copies of $H$ in $G$.
- $N_{ind}(e, H, G)$ is the number of induced copies of $H$ in $G$ that contain $e$.

Our main concern in this paper is the random variable counting the number of induced copies of $C_4$ in $G_n$. Therefore, we let $X = N_{ind}(C_4, G_n)$. We will use this notation from this point onward. We now continue by defining seed graphs. For this, let us remind the reader the definition of $\Phi$.

Let us note that the upper bound on $\Phi(\delta)$ by considering the same optimization problem where we replace ‘cubes’ with ‘one-supcubes’, is close to the actual value of $\Phi(\delta)$.

**Definition.** Let $\epsilon, \delta, K$ be positive reals. In addition let $p \in (0, 1)$. Then we define $S'(\epsilon, \delta, K)$ to be the collection of all spanning subgraphs $G \subseteq K_n$ satisfying:

(S1) $E[G][X] \geq (1 + \epsilon - \delta)E[X]$, and
(S2) $e(G) \leq K \cdot \Phi_X(\delta + \epsilon)$

We call the graphs in $S(\epsilon, \delta, K)$ seeds. Furthermore, for every positive integer $m$ we define $S_m(\epsilon, \delta, K)$ to be the set of all seeds with $m$ edges.

Since $E[G][X]$ is determined by the number of induced copies of various subgraphs of $C_4$ in $G$, it will be convenient to restate (S1) in the above definition with a graph theoretic condition, giving rise to the notion of structured seeds. In the definition we use the graph obtained from $K_{1,2}$ by adding an isolated vertex; we denote this graph by $K_{1,2} \sqcup K_1$. Now we define the structured seeds.

**Definition.** Let $\epsilon, \delta, K$ be positive reals. In addition let $p \in (0, 1)$. Then we define $S'_1(\epsilon, \delta, K)$ to be the collection of all spanning subgraphs $G \subseteq K_n$ satisfying:

(S'1) $N_{ind}(C_4, G) + N_{ind}(K_{1,2} \sqcup K_1, G)p^2 \geq (\delta - \epsilon)E[X]$, and
(S'2) $e(G) \leq K \cdot \Phi_X(\delta + \epsilon)$.

We call the graphs in $S'_1(\epsilon, \delta, K)$ structured seeds. Furthermore, for every positive integer $m$ we define $S'_m(\epsilon, \delta, K)$ to be the set of all structured seeds with $m$ edges.

The following is a lemma relating the seeds and the structured seeds, which we prove later.

**Lemma 3.1.** Let $\epsilon, \delta, K$ be positive reals and suppose also that $\epsilon \leq \delta/2$ and $p \ll 1$. Then, there exists a positive constant $C = C(\epsilon, \delta, K)$ such that for any $m \leq K\Phi_X(\delta - \epsilon)$ and large enough $n$ we have,

$$S_m(\epsilon, \delta, C) \subseteq S'_m(2\epsilon, \delta, C).$$

Given the above lemma and Corollary 2.7, we obtain an upper bound on the upper tail probability of $X$ by the probability of the existence of a structured seed. To continue we would like to use a union bound argument. Unfortunately, there are way too many structured seeds for such an argument to work. The reason for this is that many edges of the structured seed can be ‘useless’. Indeed, for a
(medium sized) structured seed we can add edges and end up with a structured seed. We can choose the new edges in any way we like and this generates many structured seeds. The caveat is that the new edges does not contribute to achieve the first condition of being a structured seed. For this reason we introduce the notion of cores. Cores are essentially ‘minimal’ structured seeds, where we force each edge of the structured seed to be useful. This is given formally in the following definition.

**Definition.** Let $\varepsilon, \delta, K$ be positive reals. In addition let $p \in (0, 1)$. Then we define $C(\varepsilon, \delta, K)$ to be collection of all structured seeds $G \in S'(\varepsilon, \delta, K)$ satisfying the following:

For all $e \in E(G)$

$$N_{\text{ind}}(e, C_4, G) + N_{\text{ind}} \left( e, K_{1,2} \sqcup K_1, G \right) p^2 \geq \varepsilon \mathbb{E}[X] / (2K \cdot \Phi_X(\delta + \varepsilon)).$$

We call the graphs in $C(\varepsilon, \delta, K)$ core graphs. Furthermore, for every positive integer $m$ we define $C_m(\varepsilon, \delta, K)$ to be the set of all cores with $m$ edges.

In cases where the parameters $\varepsilon, \delta$ and $K$ can be understood from the context we omit them. The main aim of this section is to prove the following theorem which we derive using Corollary 2.7 and several lemmas. Essentially, the difference between Theorem 3.2 and Corollary 2.7 is that we bound the upper tail probability by the existence of a core instead of a seed. As explained earlier this is crucial for our next steps to work. The idea of the proof of Theorem 3.2 is to start with Corollary 2.7, and then use Lemma 3.1 to move from seeds to structured seeds and then use Lemma 3.3 to move from structured seeds to cores. Further, let us mention that the informal intuition of Theorem 3.2 is that the upper tail event occurs almost certainly because of the existence of a core.

**Theorem 3.2.** For all positive real numbers $\varepsilon, \delta, p$ with $\varepsilon < \delta$ and $p < \min \left\{ 1 - \left( \frac{1+\delta - \varepsilon}{1+\delta - \varepsilon/2} \right)^{1/6}, 1/2 \right\}$, there is a positive constant $K = K(\varepsilon, \delta)$ such that following holds:

$$\mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X]) \leq (1 + \varepsilon)\mathbb{P}(G \subseteq G_{n,p} \text{ for some } G \in C(\varepsilon, \delta, K)).$$

(6)

In particular,

$$\mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X]) \leq (1 + \varepsilon) \sum_m |C_m(\varepsilon, \delta, K)| p^m.$$

We start by proving Lemma 3.1 relating seeds and structured seeds.

**Proof of Lemma 3.1.** Suppose $G \in S_m(\varepsilon, \delta, K)$. Changed By (S1) we have

$$(\delta - \varepsilon)\mathbb{E}[X] \leq \mathbb{E}_G[X] - \mathbb{E}[X] \leq \sum_{H \not\subseteq C_4} \text{exn}(H, C_4) N_{\text{ind}}(H, G) p^{4-\varepsilon(H)},$$

where $\subseteq$ stands for spanning subgraphs, and exn$(H, C_4)$ is the number of extensions of $H$ to an induced copy of $C_4$ by adding edges. Observe that

$$N_{\text{ind}}(H, G) \leq \begin{cases} 
    mn^2 & \text{if } H = K_2 \cup 2K_1 \\
    m^2 & \text{if } H = M_2 \text{ or } P_4,
\end{cases}$$

where $K_2 \cup 2K_1$ is the graph on four vertices and one edge, $M_2$ is a matching of size two, and $P_4$ is the path with 3 edges; this indeed holds as two edges of $G$ span at most one
induced $M_2$ or $P_4$. Therefore, we obtain Changed

$$(\delta - \varepsilon)\mathbb{E}[X] \leq N_{\text{ind}}(C_4, G) + N_{\text{ind}}(K_{1,2} \cup K_1, G)p^2 + mn^2p^3$$

Next we will use the estimation $\Phi_X(\delta) = O_\delta(\sqrt{\mathbb{E}[X]} \log(1/p))$, this follows from considering $G_{n,p}$ conditioning on the event that $G_{n,p}$ contains $K_{\Theta(np), \Theta(np)}$. For more details the reader is referred to Claim 5.1 in Section 5. Therefore, as $\mathbb{E}[X] = \Theta(n^4p^3)$, we have $m = O_{\delta, \varepsilon}(n^2p^2 \log(1/p))$. Further, as $p \ll 1$ we get from (7) the following inequality for large enough $n$,

$$(\delta - \varepsilon)\mathbb{E}[X] \leq N_{\text{ind}}(C_4, G) + N_{\text{ind}}(K_{1,2} \cup K_1, G)p^2 + O(p\log^2(1/p)\mathbb{E}[X])$$

$$\leq N_{\text{ind}}(C_4, G) + N_{\text{ind}}(K_{1,2} \cup K_1, G)p^2 + \varepsilon\mathbb{E}[X].$$

Therefore, we obtain $N_{\text{ind}}(C_4, G) + N_{\text{ind}}(K_{1,2} \cup K_1, G)p^2 \geq (\delta - 2\varepsilon)\mathbb{E}[X]$ for large enough $n$. This is the assertion of the lemma.

Informally, the next claim is that every structured seed contains a core. Formally this is given in the following claim.

**Lemma 3.3.** Suppose $\varepsilon, \delta, K$ are positive reals. Then for every structured seed $G \in S'$ and every nonnegative real $s$, there exists a subgraph $G^* \subseteq G$ such that:

1. $N_{\text{ind}}(C_4, G^*) + N_{\text{ind}}(K_{1,2} \cup K_1, G^*)p^2 \geq (\delta - \varepsilon)\mathbb{E}[X] - s$,
2. $N_{\text{ind}}(e, C_4, G^*) + N_{\text{ind}}(e, K_{1,2} \cup K_1, G^*)p^2 \geq \frac{s}{K\Phi_X(\delta + \varepsilon)}$ for every edge $e \in E(G^*)$,
3. $e(G) \leq K \cdot \Phi_X(\delta + \varepsilon)$.

**Proof.** In this proof, it would be convenient to let

$$N(G) = N_{\text{ind}}(C_4, G) + N_{\text{ind}}(K_{1,2} \cup K_1, G)p^2.$$

This is because then, (C1) is equivalent to $N(G^*) \geq (\delta - \varepsilon)\mathbb{E}[X] - s$ and (C2) is equivalent to $N(G^*) - N(G^* \setminus e) \geq \frac{s}{K\Phi_X(\delta + \varepsilon)}$ for every $e$ an edge in $G^*$.

Define the sequences $G = G_0 \supseteq G_1 \supseteq \cdots \supseteq G_r = G^*$ and $e_1, e_2, \ldots, e_r \in G$ by repeatedly setting $G_{k+1}$ to be a subgraph of $G_k$ obtained by the deletion of an edge $e_k$ such that

$$N(G_k) - N(G_{k+1}) < \frac{s}{e(G)},$$

as long as such edge exists. The graph $G^*$ satisfies (C3) as it is a subgraph of a structured seed. We also claim that, the subgraph $G^*$ satisfies (C2). That is because, if there is an edge $e$ in $G^*$ with $N(G^*) - N(G^* \setminus e) < \frac{s}{e(G)}$, the process would have continued by deleting this edge — a contradiction. Finally, since $r \leq e(G)$, we have

$$N(G) - N(G^*) = \sum_{k=0}^{r-1} N(G_k) - N(G_{k+1}) \leq \frac{rs}{e(G)} \leq s.$$
Rearranging this inequality and recalling that \( G \) is a structured seed we obtain the assertion of the lemma,

\[
N(G^*) = N_{\text{ind}}(C_4, G^*) + N_{\text{ind}}(K_{1,2} \cup K_1, G^*)p^2 \geq (\delta - \epsilon)\mathbb{E}[X] - s.
\]

Application Lemma 3.3 invoked with \( \epsilon \) replaced by \( \epsilon/2 \) and \( s = \epsilon\mathbb{E}[X]/2 \) yields the following corollary.

**Corollary 3.4.** Suppose \( \epsilon, \delta, K \) are positive reals. Suppose further that \( G \in S'(\epsilon/2, \delta, K) \) is a structured seed. Then, there exists a core \( G^* \in C(\epsilon, \delta, K) \) such that \( G^* \subseteq G \).

Now we derive Theorem 3.2 from the above claims and Corollary 2.7.

**Proof of Theorem 3.2.** Applying Lemma 3.1 with \( \epsilon \) replaced by \( \epsilon/4 \) and Corollary 3.4 with \( \epsilon \) replaced by \( \epsilon/2 \) we find that

\[
\mathbb{P}(G \not\subseteq G_{n,p} \text{ for some } G \in S(\epsilon/4, \delta, K)) \leq \mathbb{P}(G \not\subseteq G_{n,p} \text{ for some } G \in S'(\epsilon/2, \delta, K)) \leq \mathbb{P}(G \not\subseteq G_{n,p} \text{ for some } G \in C(\epsilon, \delta, K)).
\]

Applying Corollary 2.7 with \( \epsilon \) replaced with \( \epsilon/2 \) we obtain

\[
\mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X]) \leq (1 + \epsilon)\mathbb{P}(G \not\subseteq G_{n,p} \text{ for some } G \in S(\epsilon/4, \delta, K)).
\]

Combining the above inequalities we obtain the assertion of the theorem, that is

\[
\mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X]) \leq (1 + \epsilon)\mathbb{P}(G \not\subseteq G_{n,p} \text{ for some } G \in C(\epsilon, \delta, K)).
\]

\[
\square
\]

4 | LOWER BOUNDS

The aim of this section is to give lower bounds for \( \mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X]) \) for every positive \( \delta \). We do so by presenting a family of graphs such that ‘planting’ them in \( G_{n,p} \) increases the expectation of \( X \) by a multiplicative factor of \( 1 + \delta \). More formally, when we say ‘planting’ we mean changing the probability measure on the hypercube \( \{0, 1\}^{(\ast)} \) from the usual product measure of \( G_{n,p} \) to the probability measure of \( G_{n,p} \) conditioned on the existence of a subgraph from a predetermined family of graphs. The graphs in the families that we will consider will satisfy \( \mathbb{E}_G[X] \geq (1 + \delta)\mathbb{E}[X] \). We choose these families as such because, on the event that \( G \subseteq G_{n,p} \) for \( G \subseteq K_n \) such that \( \mathbb{E}_G[X] \geq (1 + \delta)\mathbb{E}[X] \) the probability of \( X \geq (1 + \delta)\mathbb{E}[X] \) is pretty ‘large’.

Since for every labeled graph \( G \subseteq K_n \) the probability of \( G_{n,p} \) containing \( G \) is \( p^{e(G)} \) it makes sense to consider such graphs with the smallest \( e(G) \). In contrast, in our case there is a wide range of \( p \) where the strongest lower bound is obtained by planting some member of a large family of graphs; each individual graph in the family is sub-optimal in terms of the number of edges, but the size of the family compensates for the difference in the number of edges between the optimal graph and the graphs in our family. This family is, roughly speaking all embeddings of an unbalanced complete bipartite graph into \( K_n \).
More precisely, denoting the complete bipartite graph with sides of size \( s \) and \( t \) by \( K_{s,t} \), we will define integers \( m_0, m_2, m_3, \ldots \) and \( m_s \) such that \( k|m_k \) for \( k \neq 0 \) and \( n|m_s \) and \( m_k \approx 2\sqrt{k/(k-1)\delta E[X]} \) for \( k \neq 0 \), \( m_0 \approx 2\sqrt{\delta E[X]} \), and \( m_s \approx (\sqrt{1+2\delta}-1)\sqrt{2E[X]} \).

Note that provided that \( n^{-1} \ll p \ll n^{-1/2} \) the constructions \( K_{k,m_k} \) and \( K_{\sqrt{m_0},\sqrt{m_0}} \) contain \( \delta E[X] \) induced copies of \( C_4 \), up to lower order terms. In addition, we will show later that \( H = K_{2m_s/n,n/2} \) admits \( \mathbb{E}_H[X] \geq (1+\delta)\mathbb{E}[X] \).

Denote by \( \mathcal{E}_k \) the set of all copies of \( K_{k,m_k} \) in \( G_n \) when \( k \neq 0, 1 \). Denote by \( \mathcal{E}_0 \) the set of all \( K_{\sqrt{m_0},\sqrt{m_0}} \) in \( G_n \) and and by \( \mathcal{E}_s \) denote the set of all copies of \( H \) in \( G_n \). Planting one of \( \mathcal{E}_k, \mathcal{E}_0 \) or \( \mathcal{E}_s \) yields a lower bound for the probability of the upper tail event which is valid for all values of \( p \). As was mentioned in the introduction the significant different between this work and previous ones is the need to plant a large family of sub-optimal graphs and not a single optimal graph. This is true only for the families \( \mathcal{E}_k \) where \( k \neq 0 \). In the case of \( \mathcal{E}_0, \mathcal{E}_s \) we could as well plant a single graph from these sets as \( |\mathcal{E}_0| \) and \( |\mathcal{E}_s| \) are negligible.

One can compare these bounds, and see that the best one depends on \( p \) in the following way. There exists an increasing sequence \( \{c_k\}_{k=1}^{\infty} \) with \( c_1 = 0 \) and \( \lim_{k \to \infty} c_k = 1/3 \) so that, provided \( n^{-1+c_k-1} \ll p \ll n^{-1+c_k} \) for some integer \( k \geq 2 \) we obtain the strongest lower bound on the upper tail probability by planting \( \mathcal{E}_k \). The best family to plant when \( n^{-2/3} \ll p \ll n^{-1/2} \) is \( \mathcal{E}_0 \) which should be thought of as the ‘limit’ of \( \mathcal{E}_k \) when \( k \) goes to infinity. Lastly, \( \mathcal{E}_s \) is the best family to plant when \( n^{-1/2} \ll p \ll 1 \).

The main result of this section is a formalisation of the above discussion. In order to make things rigorous from now on we fix \( \delta, \varepsilon \) to be positive reals and let

\[
r_k = \begin{cases} 
2 & \text{for } k = 0, \\
2\sqrt{k/(k-1)} & \text{for } k \geq 2,
\end{cases}
\]

when \( k \) is some non negative integer. Moreover, note the following: For all \( n^{-1} \ll p \ll 1 \) we have \( \mathbb{E}[X] = \Omega(n^4p^4) \). Therefore, for clarity of the presentation we assume that \( n^{-1+c_k-1} \ll p \ll n^{-1+c_k} \) for some integer \( k \geq 2 \) we obtain the strongest lower bound on the upper tail probability by planting \( \mathcal{E}_k \). The best family to plant when \( n^{-2/3} \ll p \ll n^{-1/2} \) is \( \mathcal{E}_0 \) which should be thought of as the ‘limit’ of \( \mathcal{E}_k \) when \( k \) goes to infinity. Lastly, \( \mathcal{E}_s \) is the best family to plant when \( n^{-1/2} \ll p \ll 1 \).

We wish to emphasize that both \( m_k \) and \( m_s \) depend on \( \delta \) and \( \varepsilon \).

**Theorem 4.1.** Let \( \varepsilon, \delta, C \) be positive real numbers and let \( 2 \leq k \leq Cnp \) be a positive integer. Then the following holds

\[
\mathbb{P}(X \geq (1+\delta)\mathbb{E}[X]) \geq \begin{cases} 
\left(p^{m_k}\left(\begin{array}{c} n \\
m_k/k\end{array}\right)\right)^{1+\varepsilon} & \text{for } n^{-1} \ll p \ll n^{-1/2} \\
\left(p^{(1+\varepsilon)m_s}\right)^{1+\varepsilon} & \text{for } n^{-1/2} \ll p \ll 1
\end{cases}
\]

for large enough \( n \).
In the following lemma we claim that the expected number of induced copies of $C_4$ conditioned on one labeled copy of $K_{k,m_k/k}$ or $H$ being a subgraph of $G_{n,p}$ in a suitable range of $p$ is at least $(1 + \delta + \epsilon / 2)\mathbb{E}[X]$ provided $n$ is large enough. To this end let us introduce the following notations. From now on we assume that the vertex set of $G_{n,p}$ is $[n]$. For every positive integer $k \leq n$ and $A \subseteq [n] \setminus [k]$ with $|A| = m_k/k$ define the following events:

(I) $F_{A,k}$ is the event that $n^{k-1}N(i) = A$ and there are no edges between any $i,j \in [k]$,

(II) $F_k = \bigcup_{A \in \binom{[n]\setminus[k]}{m_k/k}} F_{A,k}$.

The reason we define these events in this manner (as opposed to just the existence of a $K_{k,m_k/k}$) is to make sure that for every $A$ the events $F_{A,k}$ are disjoint. Now we are ready to state the lemma.

**Lemma 4.2.** Suppose $\epsilon, \delta, C$ are positive reals, $2 \leq k \leq Cn^p$ is some positive integer and $A \subseteq [n] \setminus [k]$ with $|A| = m_k/k$. Then the following holds:

(1) Suppose $n^{-1} \ll p \ll n^{-1/2}$. Then, for large enough $n$ we have

$$\mathbb{E}[X|F_{A,k}] \geq (1 + \delta + 3\epsilon/4)\mathbb{E}[X].$$

(2) Suppose $n^{-1/2} \ll p \ll 1$. Then, for large enough $n$ we have

$$\mathbb{E}_H[X] \geq (1 + \delta + \epsilon)\mathbb{E}[X].$$

**Proof.** We start with the first item in the lemma. Assume $n^{-1} \ll p \ll n^{-1/2}$. First, note that

$$\mathbb{E}[X|F_{A,k}] \geq N_{ind}(C_4, K_{k,m_k/k})(1 - p) + \mathbb{E}[N_{ind}(C_4, G_{n-k,p})].$$

Let us compute these two quantities.

Since $K_{s,t}$ contains exactly $\binom{s}{2}\binom{t}{2}$ induced copies of $C_4$ we obtain that $K_{k,m_k/k}$ contains

$$(k - 1)m_k^2/4k - O(m_k) = (\delta + \epsilon)\mathbb{E}[X] - O(\mathbb{E}[X]^{1/2})$$

induced copies of $C_4$ as $r_k = 2\sqrt{k/(k-1)}$. Further, as $k = O(np)$ and for all nonnegative integers $a, b, c$ we have $\binom{a-c}{b}/\binom{a}{b} \geq \binom{a-b-c}{a-b}$, we deduce

$$\mathbb{E}[N_{ind}(C_4, G_{n-k,p})] = \binom{n-k}{4}\binom{n}{4}^4 \mathbb{E}[X] \geq \left(1 - \frac{k}{n-4}\right)^4 \mathbb{E}[X] = (1 - o(1))\mathbb{E}[X].$$

Combining the above we obtain,

$$\mathbb{E}[X|F_{A,k}] \geq (1 - o(1))(\delta + \epsilon)\mathbb{E}[X] - O(\mathbb{E}[X]^{1/2}) + (1 - o(1))\mathbb{E}[X] \geq (1 + \delta + 3\epsilon/4)\mathbb{E}[X].$$

For the second item assume $n^{-1/2} \ll p \ll 1$. Similar to the first case we have

$$\mathbb{E}_H[X] \geq (N_{ind}(C_4, H) + N_{ind}(K_{1,2} \cup K_{1,4}, H)p^2)(1 - p)^2 + \mathbb{E}[N_{ind}(C_4, G_{n-2m_k/n,p})].$$

We now compute these quantities.
Remark. Suppose $\gamma > 0$ is sufficiently large. Therefore, Theorem 4.1 invoked with $n = \frac{m}{\gamma}$ will be shown to be tight in the next sections for some specific sequence $c$ satisfying $n = O(np^2) = o(n)$, we deduce

$$E[N_{ind}(C_4, G_{n-2m_*/n,p})] = \left( \frac{n-2m_*/n}{4n} \right) E[X] \geq \left( 1 - \frac{2m_*}{n(n-4)} \right) E[X] \geq (1 - o(1)) E[X],$$

taking $n$ large enough we have, $E[N_{ind}(C_4, G_{n-2m_*/n,p})] \geq (1 - \epsilon/4) E[X]$. Recalling that $p \gg n^{-1/2}$ and combining all the above bounds we obtain,

$$E[H[X] \geq m_*^2/4 + m_*n^2 p^2/8 + (1 - \epsilon/4) E[X] + o(m_*^2).$$

By this inequality and the definition of $m_*$ one can check that for large enough $n$,

$$E[H[X] \geq (\delta + 3\epsilon/2) E[X] + (1 - \epsilon/4) E[X] - \epsilon/4 E[X] = (1 + \delta + \epsilon) E[X].$$

This completes the proof. \hfill \blacksquare

Now we are ready to prove Theorem 4.1. Before proving the theorem we make the following remark. Suppose $\gamma > 0$ is a real number, $k = np$ and $p \gg n^{-1}$. Then $m_0 \leq m_k \leq (1 + \gamma)m_0$ provided $n$ is sufficiently large. Therefore, Theorem 4.1 invoked with $\epsilon = \gamma$ implies:

$$P(X \geq (1 + \delta) E[X]) \geq \left( p^{m_k} \left( \frac{n}{m_k/np} \right) \right)^{(1+\gamma)} \geq p^{(1+\gamma)^3m_0}$$

for large enough $n$. Thus, by setting $\gamma = \sqrt{1 + \epsilon} - 1$, and letting $\{c_k\}_{k=2}^\infty$ be any increasing sequence satisfying $c_2 = 0$ and $\lim_{k \to \infty} c_k = 1/3$, we have the following corollary of Theorem 4.1, which will be shown to be tight in the next sections for some specific sequence $c_k$.

**Corollary 4.3.** Let $\epsilon$ and $\delta$ be positive real numbers and let $k \geq 2$ be a positive integer.

Then the following holds

$$\log P(X \geq (1 + \delta) E[X]) \geq \begin{cases} (1 + \epsilon) \left( m_k \log p + \log \left( \frac{n}{m_k/n} \right) \right) & \text{for } n^{-1+c_{k-1}} \ll p \ll n^{1+c_k}, \\ (1 + \epsilon)m_0 \log(p) & \text{for } n^{-2/3} \ll p \ll n^{-1/2}, \\ (1 + \epsilon)m_* \log(p) & \text{for } n^{-1/2} \ll p \ll 1. \end{cases}$$

for large enough $n$.

Note that in the proof of Theorem 4.1 we use a similar method as was used in [21].

**Proof of Theorem 4.1.** Throughout the proof we assume that the vertex set of $G_{n,p}$ is $[n]$.

We start with the first item. To this end fix an integer $2 \leq k \leq Cnp$ and $\epsilon$ some positive real and assume that $n^{-1} \ll p \ll n^{-1/2}$. Let $A \subset [n] \setminus [k]$ with $|A| = m_k/k$ and recall the definitions of the events $F_{A,k}$ and $F_k$:...
(I) $F_{A,k}$ is the event that $\cap_{i=1}^{k} N(i) = A$ and there are no edges between any $i,j \in [k]$. 

(II) $F_k = \cup_{A \in \binom{[n]}{m_k/k}} F_{A,k}$.

Note that for any $A,B \subseteq [n] \setminus [k]$ such that $A \neq B$ with $|A| = |B| = m_k/k$ we have $F_{A,k} \cap F_{B,k} = \emptyset$.

We will now bound $\mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X])$ by comparing it to $\mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X]|F_{A,k}) \cdot \mathbb{P}(F_{A,k})$. The advantage of this is that $\mathbb{P}(F_{A,k})$ is easy to estimate, and by using Lemma 4.2 and a uniform bound on $X$ we will derive a bound on $\mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X]|F_{A,k})$ similar to what was done in Lemma 2.2. Indeed, by the disjointness of $F_{A,k}$ we have the following.

$$\mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X]) \geq \mathbb{P}(F_k \text{ and } X \geq (1 + \delta)\mathbb{E}[X])$$

$$= \sum_{A \in \binom{[n]}{m_k/k}} \mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X]|F_{A,k}) \cdot \mathbb{P}(F_{A,k}) \geq \sum_{A \in \binom{[n]}{m_k/k}} \mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X]|F_{A,k}) \cdot p^{m_k} (1 - p)^k (1 - p^k)^{n-k}.$$  (8)

Lemma 4.2 asserts that provided $n^{-1} \ll p \ll n^{-1/2}$ we have the following for all $A \subseteq [n] \setminus [k]$ with $|A| = m_k/k$:

$$\mathbb{E}[X|F_{A,k}] \geq (1 + \delta + \varepsilon/2)\mathbb{E}[X].$$

Note that $X \leq n^4$ always, we can bound $\mathbb{E}[X|F_{A,k}]$ from above (similar to the proof of Lemma 2.2) as follows:

$$\mathbb{E}[X|F_{A,k}] \leq (1 + \delta)\mathbb{E}[X] + \mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X]|F_{A,k}) \cdot n^4.$$ 

Combining the two inequalities we obtain,

$$\mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X]|F_{A,k}) \geq \frac{\varepsilon \cdot \mathbb{E}[X]}{2n^4}.$$ 

Therefore, for large enough $n$ we also have,

$$\mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X]|F_{A,k}) \geq \varepsilon p^4 (1 - p)^2 / (2 \cdot 4^4) \geq p^5.$$  (9)

Substituting (9) into (8) gives,

$$\mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X]) \geq \sum_{A \in \binom{[n]}{m_k/k}} p^5 p^{m_k} (1 - p)^k (1 - p^k)^{n-k}$$

$$= \binom{n}{m_k/k} p^{m_k+5} (1 - p)^k (1 - p^k)^{n-k}. $$  (10)

We now show that $p^5 (1 - p)^k (1 - p^k)^{n-k} \geq p^{o(m_k)}$ and $\binom{n}{m_k/k} \geq \left( \frac{n}{m_k/k} \right)^{1+o(1)}.$
First, as \( p \ll n^{-1/2}, m_k \gg 1 \) and \( 2 \leq k \leq Cn \), we have the following for large \( n \):

\[
\begin{align*}
p^5(1 - p)^{k^2} & (1 - p^k)^{(n-k)} \geq \exp(5 \log(p) - k^2(p + p^2) - (n-k)(p^k + p^{2k})) \\
& \geq \exp(5 \log(p) - C^2n^2p^3(1 + p) - np^2(1 + p^k)) \\
& = \exp(-o(n^2p^2)) \\
& \geq p^{e^{m_k}/2}, \quad (11)
\end{align*}
\]

where the first inequality follows as \( 1 - x \geq \exp(-x - x^2) \) for all small enough \( x \). Second, for all nonnegative integers \( a, b, c \) we have \( \left( \frac{a-c}{b} \right) / \left( \frac{a}{b} \right) \geq \left( \frac{a-b-c}{a-b} \right)^b \), and thus we obtain that

\[
\left( \frac{n-k}{m_k/k} \right) / \left( \frac{n}{m_k/k} \right) \geq \left( 1 - \frac{k}{n-m_k/k} \right)^{m_k/k} = e^{-O(m_k/n)}.
\]

In addition, as \( 2 \leq k \leq Cn \) and \( n^{-1} \ll p \ll n^{-1/2} \), we have

\[
\left( \frac{n}{m_k/k} \right)^{m_k/k} \leq \left( \frac{en}{m_k} \right)^{m_k/k} \leq \left( \frac{eCn^2p}{m_k} \right)^{m_k/k} \leq \exp\left( \left( 1 - \frac{1}{k} + o\left( \frac{1}{k \log(p)} \right) \right) m_k \log(p) \right) \leq \exp((1 + o(1))m_k \log(p)).
\]

Furthermore, as \( 2 \leq k \leq Cn \) and \( n^{-1} \ll p \ll n^{-1/2} \) we have \( m_k/n = o(m_k \log(1/p)) \) and hence, for sufficiently large \( n \) we have,

\[
\left( \frac{n-k}{m_k/k} \right) / \left( \frac{n}{m_k/k} \right) \geq \left( \frac{n}{m_k/k} \right)^{e^{m_k}/2} \quad \text{(12)}.
\]

Combining (10), (11), and (12) gives,

\[
\mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X]) \geq \left( \frac{n}{m_k/k} \right)^{p^{m_k}}^{1+\epsilon}.
\]

This finishes the first part of the proof.

For the second item, define the event \( F^* \) to be the event that \( G_{n,p} \) contains \( K_{2m_s/n, n/2} \) as a subgraph on the vertex set \( [n/2 + 2m_s/n] \) with sides \( [2m_s/n] \) and \( [n/2 + 2m_s/n \setminus [2m_s/n] \). Note that \( X \leq n^4 \) always. Further, Lemma 4.2 asserts that provided \( n^{-1/2} \ll p \ll 1 \) we have

\[
\mathbb{E}[X|F^*] \geq (1 + \delta + \epsilon)\mathbb{E}[X],
\]

and thus, \( \Phi_X(\delta + \epsilon) \leq -\log \mathbb{P}(F^*) = -\log(p^{m_k}) \). Therefore, applying Lemma 2.2 gives,

\[
-\log \mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X]) \leq \Phi_X(\delta + \epsilon) + \log \left( \frac{n^4}{\epsilon \mathbb{E}[X]} \right) \leq -\log \left( \frac{p^{m_k}n^4}{\epsilon \mathbb{E}[X]} \right).
\]
Moreover, for sufficiently large \( n \) we have, \( \frac{\epsilon E[X]}{n^4} \geq p^5 \). Thus, taking negative logarithms we obtain the following for large enough \( n \):

\[
\mathbb{P}(X \geq (1 + \delta)E[X]) \geq p^{m_\ast + 5} \geq p^{(1+\epsilon)m_\ast}.
\]

This is as claimed.

5 | COUNTING THE NUMBER OF CORES

Recall that \( X \) is the random variable counting the number of induced copies of \( C_4 \) in \( G_{n,p} \). In this section we prove a general upper bound on the logarithmic probability of the upper tail event of \( X \). The main tool we use in this section is Theorem 2.7 which is a variation of [Theorem 9.1]{HarmouSam19}. There are two major parts in this section. The first is an evaluation of \( \Phi_X(\delta) \) in different regimes of \( p \). The second is an evaluation of the entropic term \( |\mathcal{C}_m| \). The main results of this section are the following lemmas.

**Lemma 5.1.** Suppose \( \epsilon, \delta \) are positive real numbers with \( \epsilon \) being small as a function of \( \delta \). Then the following hold:

(i) If \( p \ll n^{-1/2} \) then for large enough \( n \) we have,

\[
2(1 - \epsilon)\sqrt{\delta E[X]} \leq \Phi_X(\delta)/\log(1/p) \leq 2(1 + \epsilon)\sqrt{\delta E[X]}.
\]

(ii) If \( n^{-1/2} \ll p \ll 1 \) then for large enough \( n \) we have,

\[
(1 - \epsilon)\left( \sqrt{\frac{n^4p^4}{16} + 4\delta E[X]} - \frac{n^2p^2}{4} \right) \leq \frac{\Phi_X(\delta)}{\log(1/p)} \leq (1 + \epsilon)\left( \sqrt{\frac{n^4p^4}{16} + 4\delta E[X]} - \frac{n^2p^2}{4} \right).
\]

Before we present the second lemma, let us remind the reader the definition of \( \mathcal{C}_m \).

**Definition.** Let \( \epsilon, \delta, K \) be positive reals. In addition let \( p \in (0, 1) \). Then we define \( \mathcal{C}(\epsilon, \delta, K) \) to be collection of all \( G \subset K_n \) spanning subgraphs satisfying the following:

(C1) \( e(G) \leq K \cdot \Phi_X(\delta + \epsilon) \).

(C2) \( N_{ind}(C_4, G) + N_{ind}(K_{1,2} \sqcup K_1, G)p^2 \geq (\delta - \epsilon)E[X] \), and

(C3) For all \( e \in E(G) \)

\[
N_{ind}(e, C_4, G) + N_{ind}(e, K_{1,2} \sqcup K_1, G)p^2 \geq \epsilon E[X]/(2K \cdot \Phi_X(\delta + \epsilon)).
\]

We call the graphs in \( \mathcal{C}(\epsilon, \delta, K) \) core graphs. Furthermore, for every positive integer \( m \) we define \( \mathcal{C}_m(\epsilon, \delta, K) \) be the set of all cores with \( m \) edges.

**Lemma 5.2.** Suppose \( \epsilon, \delta, C, K \) are positive reals with \( \epsilon < 1 \). Furthermore, suppose \( n^{-1} \ll p \ll 1 \) as \( n \) tends to infinity. Then, there exist \( D \) and \( n_0 \) such that the following holds for all \( n > n_0 \):

Let \( m \) be a positive integer with \( Cn^2p^2 \leq m \leq K\Phi_X(\delta + \epsilon) \). Furthermore, let

\[
v_m = \max\{|\{v \in V(G) : \deg(v) \neq 0\}| : G \in \mathcal{C}_m(\epsilon, \delta, K)\}.
\]
Then,
\[ |C_m| \leq \log \left( \frac{1}{p} \right)^{D_m} \left( \frac{n}{v_m} \right). \]

We start with the first lemma. Let us give some motivation and history of the problem.

**Definition.** For every graph $H$ and any positive integer $m$ define,

\[ N(m, H) = \max \{|T \subset G : T \cong H| \} \]

where the maximum ranges over all graphs with $m$ edges.

This definition was first presented by Erdős and Hanani in [17]. In their paper they computed the asymptotic of this function where $H$ is a clique. In [1] Alon generalized this and computed the asymptotic of this function for all $H$. Later, in [18] Friedgut and Kahn reproved Alon’s result using entropy methods which they were also able to generalize for the case of hypergraphs.

In [22] Janson, Oleszkiewicz, and Ruciński found a relation between $N(m, H)$ and a related parameter $N(n, m, H)$ and the probability that the random variable counting the number of copies of a fixed graph $H$ in $G_{n,p}$ exceeds its expectation by a multiplicative factor. This result used and generalized the machinery developed by Friedgut and Kahn. This led us to generalize the definition of $N(m, H)$ to fit our setting. Let us recall some definitions from Section 3 and introduce some new ones.

**Definition.** Suppose $H$ and $G$ are graphs. Let $n$, $m$ be positive integers and let $e$ be an edge of $G$. We define:

- $N_{\text{ind}}(H, G)$ is the number of induced copies of $H$ in $G$.
- $N_{\text{ind}}(e, H, G)$ is the number of induced copies of $H$ in $G$ that contain $e$.
- $N_{\text{ind}}(n, m, H)$ is the maximum of $N_{\text{ind}}(H, G)$ over all graphs $G$ such that the number of vertices of $G$ is at most $n$ and the number of edges of $G$ is $m$.

Note that similar generalizations have been studied in [7] and [19]. The following is a simple corollary of Lemma 6.2 which we prove later in Section 6.

**Corollary 5.3.** For every $n, m$ positive integers such that $3 < m \leq n$ we have,

\[ N_{\text{ind}}(n, m, C_4) \leq \frac{m(m - n + 1)}{4} \leq \frac{m^2}{4}. \]

In [7] the authors determined the asymptotic behaviour of $\max \{N_{\text{ind}}(n, m, K_s) : n \in \mathbb{N} \}$ for any $s \in \mathbb{N}$ (and actually obtained optimal bounds), and gave tight bounds in some ranges of $m$ (in the above we consider only $K_{2,2}$). In [19] the authors considered the problem of determining the asymptotic in $n$ and $m$ of the maximum number of copies of a fixed bipartite graph $H$ in a bipartite host graph with $n$ vertices and $m$ edges. They solved this problem for some class of graphs which includes all complete bipartite graphs $K_{s,t}$. Here we prove a similar statement to the ones in [7, 19].

We wish to emphasize that there is a significant difference in the behaviour of the problem depending on whether $n^{-1} \ll p \ll n^{-1/2}$ or $n^{-1/2} \ll p \ll 1$. This can be seen for example in the first result of this section which we now start to prove. To this end we start with an observation. Recall that we denote by $K_{1,2} \cup K_1$ the disjoint union of a star with two leaves and an isolated vertex.

**Observation 5.4.** Suppose $n,m$ are positive integers with $m \leq \binom{n}{2}$. Then,

\[ N_{\text{ind}}(n, m, K_{1,2} \cup K_1) \leq mn^2/8. \]
Proof. Let $G$ be a graph achieving the maximum in the definition of $N_{ind}(n, m, K_{1,2} \cup K_1)$. Let $uv = e \in E(G)$ and denote $x = |N(u) \cup N(v) \setminus \{u, v\}|$. Then,

$$N_{ind}(e, K_{1,2} \cup K_1, G) \leq x(n - x) \leq n^2/4.$$ 

Moreover we have,

$$N_{ind}(K_{1,2} \cup K_1, G) \leq \frac{1}{2} \sum_{e \in E(G)} N_{ind}(e, K_{1,2} \cup K_1, G) \leq \frac{mn^2}{8}. \quad \blacksquare$$

Proof of Lemma 5.1. We start with the upper bounds both in (i) and in (ii).

For the upper bound in (i), assume that $n^{-1} \ll p \ll n^{-1/2}$. Since $\lim_{n \to \infty} \mathbb{E}[X] = \infty$ we may treat $m' = \sqrt{4(1 + \varepsilon)\delta \mathbb{E}[X]}$ as an integer and let $H = K_{m', m'}$. Note that $H$ has $2\sqrt{(1 + \varepsilon)\delta \mathbb{E}[X]}$ edges and provided $n$ is large enough $H$ contains at least $(1 + \varepsilon/2)\delta \mathbb{E}[X] > \delta \mathbb{E}[X]$ induced copies of $C_4$. Note further that,

$$\mathbb{E}_H[X] \geq N_{ind}(C_4, H)(1 - p)^2 + \mathbb{E}[N_{ind}(C_4, G_{n - 2m', p})] \geq (1 + \varepsilon/2)\delta \mathbb{E}[X] + \frac{(n - 2m')}{4} \mathbb{E}[X].$$

As $m' \ll n$ we also have the following for sufficiently large $n$,

$$\binom{n}{4} - \binom{n - 2m'}{4} \geq \frac{n - 2m' - 3}{n} \geq 1 - \delta \varepsilon/2.$$

Combining the above inequalities we find the following for large enough $n$,

$$\mathbb{E}_H[X] \geq (1 + \varepsilon/2)\delta \mathbb{E}[X] + (1 - \varepsilon \delta/2)\mathbb{E}[X] = (1 + \delta)\mathbb{E}[X].$$

Therefore, for large enough $n$ we have

$$\mathbb{E}_H[X] - \mathbb{E}[X] \geq \delta \mathbb{E}[X].$$

Hence, by the definition of $\Phi_\chi(\delta)$,

$$\Phi_\chi(\delta) \leq - \log(p^{\varepsilon[H]}) = 2\sqrt{(1 + \varepsilon)\delta \mathbb{E}[X]} \log(1/p) \leq 2(1 + \varepsilon)\sqrt{\delta \mathbb{E}[X]} \log(1/p).$$

For the upper bound in (ii), assume $n^{-1/2} \ll p \ll 1$. Letting

$$\tilde{m} = (1 + \varepsilon) \left( \sqrt{\frac{n^4p^4}{16} + 4\delta \mathbb{E}[X]} - \frac{n^2p^2}{4} \right),$$

and recalling that $\lim_{n \to \infty} \sqrt{\mathbb{E}[X]}/n = \infty$ we may treat $\tilde{m}/n$ as an integer and consider the following graph. Let $H$ be a graph on the vertex set $[n]$ and such that $H[[2\tilde{m}/n + n/2]] \cong K_{2\tilde{m}/n,n/2}$ and all vertices in $[n] \setminus [2\tilde{m}/n + n/2]$ are isolated. Note that $H$ contains $\tilde{m}$ edges. Let $\xi$ be a small positive real. Assuming $n$ is large enough, $H$ contains at least $(1 - \xi)\tilde{m}^2/4$
induced copies of $C_4$. Furthermore, if $n$ is large enough, $H$ contains at least $(1 - \xi)^{\tilde{m}n^2/8}$ induced copies of $K_{1,2} \sqcup K_1$. Let $\eta$ be a small positive real (that might depend on $\xi$). Then provided $n$ is large enough,

\[
\mathbb{E}_H[X] \geq (N_{ind}(C_4, H) + N_{ind}(K_{1,2} \sqcup K_1, H)p^2)(1 - p)^2 + \mathbb{E}[N_{ind}(C_4, G_{n-2\tilde{m}/n, p})]
\]

\[
\geq (1 - \eta)(1 - \xi)\left(\frac{\tilde{m}^2}{4} + \tilde{m}n^2p^2\right) + \frac{(n-2\tilde{m}/n)^4}{n^4} \mathbb{E}[X].
\]

Where the first inequality holds as the first term is at most the expected number of induced copies $C_4$ with at least one vertex in $[2\tilde{m}/n]$, and the second term is the expected number of induced copies of $C_4$ with no vertices in $[2\tilde{m}/n]$.

By the definition of $\tilde{m}$ and the choices of $\xi, \eta$ being small enough we obtain,

\[
(1 - \eta)(1 - \xi)\left(\frac{\tilde{m}^2}{4} + \tilde{m}n^2p^2\right) \geq (1 + \varepsilon/2)\delta \mathbb{E}[X].
\]

Further, for large enough $n$ we have,

\[
\left(\frac{n-2\tilde{m}/n}{4}\right) \left(\frac{n}{4}\right)^4 \geq \left(\frac{n-2\tilde{m}/n-3}{n}\right)^4 \geq (1 - \delta\varepsilon/2).
\]

Combining all of these inequalities we have,

\[
\mathbb{E}_H[X] - \mathbb{E}[X] \geq \delta \mathbb{E}[X].
\]

Hence, by the definition of $\Phi_X(\delta)$ we have,

\[
\Phi_X(\delta) \leq -\log(p^{\varepsilon(H)}) = (1 + \varepsilon)\left(\sqrt{\frac{n^4p^4}{16} + 4\varepsilon\mathbb{E}[X] - \frac{n^2p^2}{4}}\right) \log(1/p).
\]

For the lower bounds of both (i) and (ii) we start with a claim.

**Claim 5.5.** Suppose $G$ is a spanning subgraph of $K_n$ achieving the minimum in the definition of $\Phi_X(\delta)$ and let $m = e(G)$. Then for any small enough $\gamma > 0$ and large enough $n$

\[
\delta \mathbb{E}[X] \leq m^2/4 + \min\{mn^2p^2/8, m^2np^2/2\} + \gamma n^4p^4.
\]

**Proof.** Let $G$ be a spanning subgraph of $K_n$ achieving the minimum in the definition of $\Phi_X(\delta)$ and put $m = e(G)$. By the definition of $\Phi_X(\delta)$ we have $\mathbb{E}_G[X] - \mathbb{E}[X] \geq \delta \mathbb{E}[X]$. Note that we can bound $\mathbb{E}_G[X] - \mathbb{E}[X]$ from above in the following way:

\[
(\delta - \varepsilon)\mathbb{E}[X] \leq \mathbb{E}_G[X] - \mathbb{E}[X] \leq \sum_{\emptyset \neq H \subseteq C_4} \text{exn}(H, C_4)N_{ind}(H, G)p^{4-e(H)},
\]

where $\subseteq^*$ stands for spanning subgraphs, and $\text{exn}(H, C_4)$ is the number of extensions of $H$ to an induced copy of $C_4$ by adding edges. Let $P_4$ be the path with three edges, $M_2$ be a matching of size two and $K_2 \cup I_2$ be a disjoint union of an edge and an independent set of
size two. Since any matching of size two span at most one induced copy of \( P_4 \) or \( M_2 \) we have

\[
N_{\text{ind}}(P_4, G), N_{\text{ind}}(M_2, G) \leq m^2.
\]

Therefore we obtain

\[
\mathbb{E}_G[X] - \mathbb{E}[X] \leq N_{\text{ind}}(C_4, G) + N_{\text{ind}}(K_{1,2} \cup K_1, G)p^2 \\
+ \left( \frac{m^2p}{N_{\text{ind}}(P_4, G)} \right) + 2 \cdot \frac{m^2p^2}{N_{\text{ind}}(M_2, G)} + 2 \cdot \frac{mn^2p^3}{N_{\text{ind}}(K_{1,2} \cup K_1, G)}
\]

\[
\leq N_{\text{ind}}(C_4, G) + N_{\text{ind}}(K_{1,2} \cup K_1, G)p^2 + 2mp(m + mp + n^2p^2).
\]

By Observation 5.4 we have

\[
N_{\text{ind}}(K_{1,2} \cup K_1, G) \leq N_{\text{ind}}(m, n, K_{1,2} \cup K_1) \leq \frac{mn^2}{8}.
\]

Furthermore we always have,

\[
N_{\text{ind}}(K_{1,2} \cup K_1, G) \leq \left( \frac{m}{2} \right) \cdot n \leq \frac{m^2n}{2}.
\]

Combining the two we find that,

\[
N_{\text{ind}}(K_{1,2} \cup K_1, G)p^2 \leq \min\{mn^2p^2/8, m^2np^2/2\}.
\]

Hence, by the definition of \( \Phi_X(\delta) \) and Corollary 5.3, for large enough \( n \) we have,

\[
\delta \mathbb{E}[X] \leq N_{\text{ind}}(C_4, G) + N_{\text{ind}}(K_{1,2} \cup K_1, G)p^2 + 2mp(m + mp + n^2p^2) \\
\leq m^2/4 + \min\{mn^2p^2/8, m^2np^2/2\} + 2mp(m + mp + n^2p^2).
\]

(13)

As \( p \ll 1 \) and the upper bound achieved before, we have \( m = O(n^2p^2) \). Therefore, for any \( \gamma > 0 \) and sufficiently large \( n \) we obtain from (13) the following inequality

\[
\delta \mathbb{E}[X] \leq m^2/4 + \min\{mn^2p^2/8, m^2np^2/2\} + \gamma n^4p^4.
\]

We now prove the lower bound in (i). For this assume that \( n^{-1} \ll p \ll n^{-1/2} \) and note that this implies that \( \min\{mn^2p^2/8, m^2np^2/2\} = o(m^2) \). Thus, for any \( \gamma > 0 \) and large enough \( n \) we deduce the following from Claim 5.5,

\[
\delta \mathbb{E}[X] \leq m^2/4 + \min\{mn^2p^2/8, m^2np^2/2\} + \gamma n^4p^4 \leq (1/4) + (\gamma)m^2 + \gamma n^4p^4.
\]

Taking \( \gamma \) sufficiently small we obtain the following bound on \( m \),

\[
(1 - \epsilon)\delta \mathbb{E}[X] \leq \frac{1}{4(1 - \epsilon)}m^2,
\]
implying
\[ m \geq 2(1 - \epsilon)\sqrt{\delta \mathbb{E}[X]}. \]

This proves the lower bound in (i) as for sufficiently large \( n \) we obtain,
\[ \Phi_X(\delta) = -\log(p^m) \geq 2(1 - \epsilon)\sqrt{\delta \mathbb{E}[X]} \log(1/p). \]

Lastly, we prove the lower bound in (ii). For this assume \( n^{-1/2} \ll p \ll 1 \). Let \( \gamma > 0 \) be some small constant, then for large enough \( n \) Claim 5.5 implies,
\[ \delta \mathbb{E}[X] \leq m^2/4 + mn^2p^2/8 + \gamma \mathbb{E}[X]. \]

We conclude the following as the above is a quadratic inequality in \( m \) and the fact that \( \mathbb{E}[X] = (1 + o(1))n^4p^4/8 \),
\[ m \geq \sqrt{\frac{n^4p^4}{16} + 4(\delta - \gamma) \mathbb{E}[X]} - \frac{n^2p^2}{4} \geq (1 - \epsilon) \left( \sqrt{\frac{n^4p^4}{16} + 4\delta \mathbb{E}[X]} - \frac{n^2p^2}{4} \right). \]

This implies (ii) as follows,
\[ \Phi_X(\delta) = -\log(p^m) \geq (1 - \epsilon) \left( \sqrt{\frac{n^4p^4}{16} + 4\delta \mathbb{E}[X]} - \frac{n^2p^2}{4} \right) \log(1/p). \]

This finishes the first evaluation we prove in this section. The second evaluation is the evaluation of the entropic term \(|C_m|\) given by Lemma 5.2. Roughly speaking Lemma 5.2 shows that the number of core graphs with \( m \) edges is determined by \( v_m \) the maximum number of non-isolated vertices over all graphs in \( C_m \). As seen already, there is a big difference in the behaviour of the problem depending on whether \( p \ll n^{-1/2} \) or \( p \gg n^{-1/2} \). Later, we will derive two corollaries from Lemma 5.2, corresponding to these regimes, these corollaries will be very important in Section 6.

**Proof of Lemma 5.2.** Let \( G \in C_m \) be a core graph with \( m \) edges and let \( uv \) be an edge in \( G \). Then by the definition of a core graph we have,
\[ N_{ind}(uv, C_4, G) + N_{ind}(uv, K_{1,2}, G)np^2 \geq \epsilon \mathbb{E}[X]/(2K \cdot \Phi_X(\delta + \epsilon)). \]

Therefore,
\[ \deg(u) \deg(v) \geq N_{ind}(uv, C_4, G) \geq \epsilon \mathbb{E}[X]/(4K \cdot \Phi_X(\delta + \epsilon)), \quad (14) \]

or
\[ \deg(u) + \deg(v) \geq N_{ind}(uv, K_{1,2}, G) \geq \epsilon \mathbb{E}[X]/(4K \cdot \Phi_X(\delta + \epsilon)np^2), \quad (15) \]
call this property \((\ast)\). To bound \(|C_m|\) it is enough to bound the number of subgraphs of \( K_n \) with \( m \) edges satisfying property \((\ast)\). This can be bounded from above by multiplying the following quantities:

1. The number of possible ways to choose the set of non-isolated vertices of such graph.
(2) The number of possible choices of the edges of the graph such that property (*) is being satisfied.

The first item can be bounded from above by the number of ways to choose a set of at most \( v_m \) vertices which is at most \( 2^{v_m} \binom{n}{v_m} \).

We now bound the second item. Let \( H \) be a graph with \( v \) vertices and \( m \) edges that satisfies property (*). For every integer \( 0 \leq t \leq \log_2(m) \) define \( V_t(H) = \{ v \in V(H) : 2^t \leq \deg(v) < 2^{t+1} \} \). Furthermore, for every integer \( 0 \leq t \leq \log_2(m) \) define \( U_t(H) = \bigcup_{\ell \geq t} V_\ell \).

Using a standard double counting argument and the fact that \( U_t(H) \subseteq V(H) \) we have

\[
|V_t(H)| \leq |U_t(H)| \leq \min \left\{ \frac{m}{2^t-1}, v_m \right\}
\]

for all \( t \geq 0 \). By property (*) edges can be placed between the sets \( V_t \) in two ways.

The first option is when the edges satisfy (14). Since, the degree of each vertex is always bounded from above by \( n \), we obtain that the degrees of the endpoints of each edge satisfying (14) are bounded from below by \( d^* \) defined as

\[
d^* = \frac{\epsilon \mathbb{E}[X]}{4K \cdot n \Phi_X(\delta + \epsilon)} = \frac{C_0 np^2}{\log(1/p)}
\]

where \( C_0 \) is some positive real that might depend on \( \epsilon, \delta \) and \( K \). Thus, such edges can only connect a vertex in \( V_t \) and a vertex in \( U_{\lfloor \log_2(d^*) \rfloor} \) for integers \( \lfloor \log_2(d^*) \rfloor \leq t \leq \log_2(n) \) and \( \ell(t) = \frac{\epsilon \mathbb{E}[X]}{(4K \cdot \Phi_X(\delta + \epsilon) 2^t)} \). We denote the number of such options by \( T_m \).

The second option is that the edge has an endpoint in \( U_{\lfloor \log_2(r) \rfloor} \) where \( r \) is defined as

\[
r = \frac{\epsilon \mathbb{E}[X]}{(8K \cdot \Phi_X(\delta + \epsilon) np^2)}.
\]

That happens when the edge satisfies (15). We denote the number of such options by \( S_m \).

Furthermore, note that provided \( n \) is large enough there are at most

\[
\left( \log_2(n) - \log_2(d^*) + 2 \right)^{v_m} = \left( \log_2 \left( \frac{\log(1/p)}{C_0 p^2} \right) + 2 \right)^{v_m} \leq 3^{v_m} \log(1/p)^{v_m}
\]

partitions of the non-isolated vertices of our graph into sets \( V_t \) where \( \lfloor \log_2(d^*) \rfloor \leq t \leq \log_2(n) \) is an integer and \( \bigcup_{t=0}^{\lfloor \log_2(d^*) \rfloor} V_t \). We will think of the vertices in \( V_t \) as vertices which will have degrees between \( 2^t \) and \( 2^{t+1} \) where \( \lfloor \log_2(d^*) \rfloor \leq t \leq \log_2(n) \) is an integer.

We conclude that given the vertex set and a partition as explained above, there are at most \( T_m + S_m \) pairs of vertices that can be edges of \( H \). Hence, the number of graphs with property (*) is bounded from above by

\[
(3 \log(1/p))^{v_m} 2^{v_m} \binom{n}{v_m} \left( \frac{S_m + T_m}{m} \right) \leq (6 \log(1/p))^{v_m} \binom{n}{v_m} \left( \frac{S_m + T_m}{m} \right).
\]

Let us now estimate \( S_m \) and \( T_m \).

\[
T_m = \sum_{t = \lfloor \log_2(d^*) \rfloor}^{\log_2(n)} |V_t| |U_{\lfloor \log_2(\ell(t)) \rfloor}| \leq \sum_{t = \lfloor \log_2(d^*) \rfloor}^{\log_2(n)} \frac{m}{2^t-1} \ell(t) = \sum_{t = \lfloor \log_2(d^*) \rfloor}^{\log_2(n)} \frac{m^2}{2^t-1} \frac{4K \cdot \Phi_X(\delta + \epsilon) 2^t}{\epsilon \mathbb{E}[X]}
\]

\[
\leq (\log_2(n) - \lfloor \log_2(d^*) \rfloor + 1) \frac{64K \Phi_X(\delta + \epsilon) m^2}{\epsilon \mathbb{E}[X]} \leq \frac{192K \Phi_X(\delta + \epsilon) m^2 \log(1/p)}{\epsilon \mathbb{E}[X]}.
\]
Recalling the assumption that \( m \leq K \Phi_X(\delta + \epsilon) \) we obtain,

\[
T_m \leq \frac{192K^3 \Phi_X^2(\delta + \epsilon) \log(1/p)}{\epsilon E[X]}.
\]

Applying Lemma 5.1 we find that there is \( C_1 > 0 \) such that,

\[
\Phi_X(\delta + \epsilon) \leq \frac{C_1 E[X]}{n^2p^2} \log(1/p).
\]

By our assumptions, \( Cn^2p^2 \leq m \leq \Phi_X(\delta + \epsilon) = O(n^2p^2 \log(1/p)) \). This implies there are \( C_2, C_3 > 0 \) such that for large enough \( n_0 \) we have,

\[
T_m \leq C_2 n^2p^2 \log^4(1/p) \\
\leq C_3 m \log^5(1/p).
\]

For \( S_m \) recall that there are at most \( 4m/r \) vertices in \( U_{\lfloor \log_2(r) \rfloor}(H) \) and further recall that \( \Phi_X(\delta + \epsilon) = O(n^2p^2 \log(1/p)) \). Hence,

\[
S_m \leq \frac{n 4m \log(1/p)}{r} \leq C_4 m \log^2(1/p),
\]

where \( C_4 = C_4(\epsilon, \delta, K) > 0 \). This implies that for large enough \( n_0 \) we have \( S_m, T_m = O(m \log^5(1/p)) \). Putting it all together there exists \( C_5 > 0 \) such that provided \( n_0 \) is large enough we have,

\[
\log(\lfloor C_m \rfloor) \leq (6 \log(1/p))^{v_m} \left( T_m + S_m \right) \\
\leq (6 \log(1/p))^{2m} \left( \frac{C_5 m \log^5(1/p)}{m} \right)^m \\
\leq \log(1/p)^{O(m)}.
\]

This proves the lemma. \( \blacksquare \)

We now deduce two corollaries and discuss how one would use them in order to obtain an upper bound on the upper tail probability of \( X \). The first corollary will be used when \( n^{-1} \ll p \ll n^{-1/2} \) and the second when \( n^{-1/2} \ll p \ll 1 \).

**Corollary 5.6.** Suppose \( \epsilon, \delta, \eta, C, K \) are positive reals with \( \epsilon < 1 \). Furthermore, suppose \( n^{-1} \ll p \ll n^{-1/2} \) as \( n \) tends to infinity. Then, there exist \( D > 0 \) and \( n_0 \) such that the following holds for all \( n > n_0 \):

Let \( m \) be a positive integer with \( Cn^2p^2 \leq m \leq K \Phi_X(\delta + \epsilon) \). Furthermore, let

\[
v_m = \max \{|\{v \in V(G) : \deg(v) \neq 0\} : G \in C_m(\epsilon, \delta, K)\}.
\]

Then,

\[
\log(|C_m|) \leq -v_m \log(np^2) + \eta m \log(n).
\]
\textbf{Proof.} By Lemma 5.2 and the assumption that $p \gg n^{-1}$ we have the following provided $n_0$ is sufficiently large,

$$\log(|C_m|) \leq \log\left(\frac{n}{v_m}\right) + Dm \log(1/p) \leq v_m \log(en/v_m) + \eta m \log(n)/3.$$ 

Since, $v_m \leq 2m$ we deduce that

$$\log(|C_m|) \leq v_m \log(n/v_m) + 2\eta m \log(n)/3.$$ 

We now split into two cases. First, assume $v_m \leq \eta m/3$. This assumption implies that

$$v_m \log(n/v_m) \leq \eta m \log(n)/3.$$ 

On the other hand, if we assume $v_m \geq \eta m/3$ we obtain

$$v_m \log(n/v_m) \leq v_m \log(3n/\eta m) = -v_m \log(np^2) + v_m \log(O(1)),$$

where the equality is due to the assumption that $m = \Omega(n^2p^2)$. Since $v_m \leq 2m$, in both cases we obtain that, provided $n$ is large enough, the following holds:

$$\log(|C_m|) \leq -v_m \log(np^2) + \eta m \log(n).$$

\[\blacksquare\]

Corollary 5.6 suggests the following ‘plan of attack’ which we implement in Section 6. Assuming $p \ll n^{-1/2}$ the term $\log(np^2)$ is negative. Hence, by Corollary 5.6, to give an upper bound on $|C_m|$ it is enough to estimate $v_m$, the maximum number of vertices that a core with $m$ edges can have. This estimation and Corollary 5.6 then yield an upper bound on the number of cores with $m$ edges. Denote this bound by $\beta_m$. We then compare $p^m \beta_m$ for every $m_0 \leq m \leq K \Phi_X(\delta)$, where $m_0$ is the minimum number of edges in a core graph and denote this maximum by $\beta^*$. This implies,

$$\sum_{m=m_0}^{K \Phi_X(\delta)} p^m |C_m| \leq K \Phi_X(\delta) \beta^*.$$ 

We also show that $K \Phi_X(\delta)$ is negligible compared to $\beta^*$ and therefore we obtain,

$$\log\left(\sum_{m=m_0}^{K \Phi_X(\delta)} p^m |C_m|\right) \leq (1 - \eta) \log(\beta^*).$$

Combining this with Theorem 3.2 we obtain the desired upper bound on the logarithm of the upper tail probability,

$$\log \mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X]) \leq (1 - \epsilon) \log(\beta^*).$$

We wish to emphasize that $\beta^*$ depends on $p$, and that we will also show that the above upper bounds are matched to the lower bounds that we obtained in Section 4.

Next we present a corollary of Lemma 5.2 which will be used in the range where $p \gg n^{-1/2}$. 
Corollary 5.7. Suppose $\epsilon, \delta, C, K$ are positive reals with $\epsilon < 1$. Furthermore, suppose $n^{-1/2} \ll p \ll 1$ as $n$ tends to infinity. Then, there exists $n_0$ such that the following holds for all $n > n_0$:

Let $m$ be a positive integer with $Cn^2p^2 \leq m \leq K\Phi_X(\delta + \epsilon)$. Then,

$$|C_m| \leq \left(\frac{1}{p}\right)^{\epsilon m}.$$ 

This is also an immediate corollary of Lemma 5.2. To see this note that in this range of $p$ we have $n \ll m$ and thus $\left(\frac{n}{v_m}\right) \leq 2^n \leq \log (1/p)^m$. Hence Lemma 5.2 implies

$$\log |C_m| \leq O(m \log \log (1/p)) = o(m \log (1/p)).$$

Similar to before $\Phi_X(\delta + \epsilon)$ will be shown to be negligible and therefore Corollary 5.7 will yield the following provided $p \gg n^{-1/2}$,

$$\log \left(\sum_{m=m_*}^{K\Phi_X(\delta)} p^m |C_m|\right) \leq (1 - \epsilon)m_* \log (p)$$

where $m_*$ is the minimum number of edges in a core graph. Which then by Theorem 3.2 implies

$$\log \mathbb{P}(X \geq (1+\delta)\mathbb{E}[X]) \leq (1 - \epsilon)m_* \log (p).$$

In the next section we will compute this $m_*$ and show that the above matches the lower bounds given in Section 4.

6 | UPPER BOUNDS

As can be seen in previous sections there is a big difference in the behaviour of the problem depending on the regime $p$ lies at. As explained in Section 5, in order to obtain quantitative bounds on the upper tail probability we need to bound the entropic term $|C_m|$ when $\frac{\log^d(n)}{n} \ll p \ll \frac{1}{\sqrt{n \log(n)}}$, and when $p \gg n^{-1/2}$ we only need to compute the minimum number of edges in a core graph. Actually, the situation is more involved in the sparse regime where we see a surprising change in the behaviour of the problem. To present the main result of this section let us define a sequence $c_k$ for $k \geq 2$:

$$c_k = \begin{cases} 0 & \text{for } k = 1, \\ \frac{1}{2+\sqrt{\frac{1}{k-1}}} & \text{for } k \geq 2. \end{cases}$$

We note that $c_k$ is an increasing sequence and $\lim_{k \to \infty} c_k = 1/3$. This is the sequence promised in the our main theorem 1.2 and in Section 4. Furthermore, recall the definitions of $m_k$ and $m_*$ which were given at the beginning of Section 4. For the convenience of the reader we give these definitions here as well. First, we (re)define $m_k = r_k \sqrt{(\delta - \epsilon)\mathbb{E}[X]}$, where

$$r_k = \begin{cases} 2 & \text{for } k = 0, \\ 2\sqrt{k/(k-1)} & \text{for } k \geq 2. \end{cases}$$
Second, we (re)define 
\[ m_* = \left( \sqrt{r + d^2} - d \right) \sqrt{\mathbb{E}[X]} / 2, \]
where \( r = 16(\delta - 3\epsilon / 2) \) and \( d = \sqrt{2}/(1 - \epsilon) \). We are now ready to state the main theorem of this section which is the following.

**Theorem 6.1.** Suppose \( k \) is an integer greater than 1 and suppose that \( \epsilon, \delta \) are positive reals with \( \epsilon \) small enough. Then there exists \( n_0 \) such that for any \( n > n_0 \) the following holds:

1. If \( \max \left\{ \frac{\log^2(n)}{n}, n^{-1+\epsilon} \right\} \leq p \leq n^{-1+\epsilon} \), then
   \[ \log \mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X]) \leq (1 - \epsilon) \log \left( p^{m_0} \left( \frac{n}{m_0/k} \right) \right). \]

2. If \( n^{-2/3} \leq p \leq \frac{1}{\sqrt{n\log(n)}} \), then
   \[ \log \mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X]) \leq (1 - \epsilon)m_0 \log(p). \]

3. If \( n^{-1/2} \ll p \ll 1 \), then
   \[ \log \mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X]) \leq (1 - \epsilon)m_* \log(p). \]

The proof naturally splits into three parts. We will prove each part in a different subsection. Before we split into cases let us prove a lemma which makes a connection between the number of induced copies of \( C_4 \) in a graph to the numbers of edges and vertices of that graph. This lemma will be important for us both in the sparse regime and the dense regime.

**Lemma 6.2.** Suppose \( n, m \) are positive integers such that \( m > 3 \). Let \( G \) be a graph with \( n \) vertices and \( m \) edges. Then,
\[ N_{ind}(C_4, G) \leq \frac{m^2}{4}. \]
Furthermore, if we assume that \( \delta(G) \geq 2 \) then
\[ N_{ind}(C_4, G) \leq \frac{m(m - n + 1)}{4}. \]

Remark. The bound from the lemma is sharp when \( n = \frac{m}{k} + k \) for some \( k \in \mathbb{N} \) as the graph \( K_{k, \frac{n}{k}} \) has \( n \) vertices and \( m \) edges and contains \( \binom{k}{2} \left( \frac{n}{2} \right) = \frac{m}{4}(m - k - \frac{m}{k} + 1) \) induced copies of \( C_4 \).

Let us mention that from the above lemma we find that \( m_0 \) is a lower bound on the number of edges of a core. This is because if \( m < m_0^2 \) then \( N_{ind}(C_4, G) \leq m^2/4 < (\delta - \epsilon)\mathbb{E}[X] \) contradicting \( G \) being a structured seed and in particular a core.

**Proof of Lemma 6.2.** Let \( G \) be a graph on \( n \) vertices and \( m \) edges. We state with the following observation:
\[ N_{ind}(C_4, G) = \frac{1}{4} \sum_{e \in E(G)} N_{ind}(e, C_4, G) \leq \frac{m}{4} \max_{e \in E(G)} N_{ind}(e, C_4, G). \]
Clearly, \( N_{\text{ind}}(e, C_4, G) \leq m \), which implies the first assertion of the lemma. For the furthermore part, assume that \( \delta(G) \geq 2 \). By the assumption on the minimum degree we may assume that \( n \leq m \leq \left( \frac{n}{2} \right) \). By the above observation, we only need to prove that \( N_{\text{ind}}(uv, C_4, G) \leq m - n + 1 \) for all \( uv \in E(G) \). Note that every \( C_4 \) containing \( uv \) is determined by the ‘parallel’ edge to \( uv \) meaning the edge between the two other vertices in that \( C_4 \). This is because we are considering induced copies. In other words, the number of induced copies containing \( uv \) is bounded by the number of edges between \( N(u) \setminus \{v\} \) and \( N(v) \setminus \{u\} \). Let \( X \) be the set of vertices not in \( N(u) \cup N(v) \) and note that \( |X| \geq n - \deg(u) - \deg(v) \). The number of edges between \( N(u) \setminus \{v\} \) and \( N(v) \setminus \{u\} \) is bounded from above by \( m - e(X) - (\deg(u) + \deg(v) - 1) \) where \( e(X) \) is the number of edges with an endpoint in \( X \). Since \( \delta(G) \geq 2 \) we may estimate \( e(X) \) as follows,

\[
e(X) \geq \frac{1}{2} \sum_{x \in X} \deg(x) \geq |X|.
\]

Thus, the number of edges between \( N(u) \setminus \{v\} \) and \( N(v) \setminus \{u\} \) is bounded from above by

\[
m - (|X| + \deg(v) + \deg(u) - 1) \leq m - n + 1
\]

and the lemma follows. \( \blacksquare \)

In order to prove Theorem 6.1 we now split into cases, each dealing with each item of the theorem.

6.1 The dense regime

Assume that \( n^{-1/2} \ll p \ll 1 \) and assume that \( \varepsilon \) is any positive real which is small enough as a function of \( \delta \). Corollary 5.7 asserts that in this regime of \( p \) the number of core graphs is negligible. Therefore, as explained at the end of Section 5, to bound the upper tail probability using Theorem 3.2 we are only left with showing that \( m_* \) is a lower bound on the minimum number of edges in a core. This is given in the following lemma.

**Lemma 6.3.** Suppose \( \varepsilon, \delta \) are positive real numbers with \( \varepsilon \) small enough as a function of \( \delta \). Further suppose \( p \gg n^{-1/2} \). Then, there exists \( n_0 \) such that for all \( n > n_0 \)

\[
\min\{e(G) : G \in C\} \geq m_*.
\]

Before proving the claim let us recall Observation 5.4. For the convenience of the reader we state it here as well:

**Observation 6.4.** Suppose \( n, m \) are positive integers with \( m \leq \left( \frac{n}{2} \right) \). Then,

\[
N_{\text{ind}}(m, n, K_{1,2} \sqcup K_1) \leq mn^2 / 8.
\]

**Proof of Lemma 6.3.** Suppose \( G \in C_m \) is a core graph with \( m \) edges. Then \( G \) is also a structured seed and thus

\[
N_{\text{ind}}(C_4, G) + N_{\text{ind}}(K_{1,2} \sqcup K_1, G)p^2 \geq (\delta - \varepsilon)\mathbb{E}[X].
\]
Lemma 6.2 and Observation 6.4 assert that

\[ N_{ind}(C_4, G) \leq N_{ind}(m, C_4) \leq m^2/4, \]

and

\[ N_{ind}(K_{1,2} \cup K_1, G) \leq N_{ind}(m, n, K_{1,2} \cup K_1) \leq mn^2/8. \]

Therefore,

\[ m^2 + mn^2p^2/2 - 4(\delta - \epsilon)E[X] \geq 0. \]

Taking \( n_0 \) large enough we have \( E[X] \geq (1 - \epsilon)^2n^4p^4/8 \) and thus,

\[ m^2 + m\sqrt{2E[X]} \geq 4(\delta - \epsilon)E[X] \geq 0. \]

This implies that for large enough \( n_0 \) and small enough \( \epsilon \) we have

\[ m \geq \left( \sqrt{\frac{2}{(1-\epsilon)^2}} + 16(\delta - \epsilon) - \frac{\sqrt{2}}{1-\epsilon} \right) \sqrt{E[X]}/2 = m^*. \]

As explained after Corollary 5.7, this implies the following corollary which is the third item in Theorem 6.1.

**Corollary 6.5.** Suppose \( \epsilon, \delta \) are positive real numbers with \( \epsilon \) small enough as a function of \( \delta \). Further suppose \( n^{-1/2} \ll p \ll 1 \). Then, for large enough \( n \) we have

\[ \log \mathbb{P}(X \geq (1 + \delta)E[X]) \leq (1 - \epsilon)m^* \log (p). \]
Proof. Let $G \in C_m$ be a core graph with $C n^2 p^2 \leq m \leq K \Phi_X(\delta + \epsilon)$ edges and let $uv \in E(G)$ be an edge of $G$. By the definition of $C_m$ we have the following:

$$N_{ind}(uv, C_4, G) + N_{ind}(uv, K_{1,2}, G)n p^2 \geq \frac{\epsilon [E[X]]}{2K \Phi_X(\delta + \epsilon)}.$$ 

Note that $N_{ind}(uv, K_{1,2}, G) \leq m \leq K \Phi_X(\delta + \epsilon)$. Lemma 5.1 asserts that there exists $D > 0$ such that provided $n_0$ is large enough we have

$$\Phi_X(\delta + \epsilon) \leq D \sqrt{E[X]} \log(1/p).$$ 

Hence as $p \ll \frac{1}{\sqrt{n \log(n)}}$ we deduce the following for large enough $n_0$:

$$N_{ind}(uv, K_{1,2}, G)n p^2 \leq KD \sqrt{E[X]} n p^2 \log(1/p) \ll \sqrt{E[X]} / \log(1/p),$$

and

$$N_{ind}(uv, C_4, G) \geq \frac{\epsilon \sqrt{E[X]}}{2DK \log(1/p)} - N_{ind}(uv, K_{1,2}, G)n p^2 \geq \frac{\epsilon \sqrt{E[X]}}{4DK \log(1/p)}. \quad (16)$$

Note also that $\deg(u) \deg(v) \geq N_{ind}(C_4, uv, G)$ and thus the assertion follows for small enough $\xi$,

$$\deg(u) \deg(v) \geq \frac{\epsilon \sqrt{E[X]}}{4DK \log(1/p)} \geq \frac{\xi n^2 p^2}{\log(1/p)}$$

and if $\deg(u) = 1$ or $\deg(v) = 1$ then $N_{ind}(C_4, uv, G) = 0$ which is a contradiction to (16).

The above lemma will be used many times throughout this subsection, mostly the fact that when $\frac{\log^3(n)}{n} \leq p \ll \frac{1}{\sqrt{n \log(n)}}$ the minimum degree of a core graph is at least 2. We will omit the reference to this lemma and keep in mind that the minimum degree of all graphs considered is at least 2.

The following is a corollary which bounds $v_m = \max |\{v \in V(G) : G \in C_m \text{ and } \deg(v) > 0\}|$. This will be used afterwards to formalize the discussion after Corollary 5.6.

**Lemma 6.7.** Suppose $\epsilon, \delta, K, \gamma$ are positive reals. Suppose further that $\frac{\log^3(n)}{n} \leq p \ll \frac{1}{\sqrt{n \log(n)}}$. Then, there exists an integer $n_0$ such that for all integers $n > n_0$ the following holds:

Suppose $m$ is an integer with $m_0 \leq m \leq K \Phi_X(\delta + \epsilon)$. Then,

$$v_m \leq \frac{m}{2} + m^{3/4}.$$ 

**Proof.** Fix an arbitrary $G \in C_m$ and define, $A = \{v \in V(G) : \deg(v) < \sqrt{m/\log^2(n)}\}$. We have

$$2m = \sum_{v \in V(G)} \deg(v) = \sum_{v \in A} \deg(v) + \sum_{v \in A^c} \deg(v) \geq |A^c| \sqrt{m/\log^2(n)}.$$ 

Thus, $|A^c| \leq 2 \sqrt{m} \cdot \log^2(n)$. 

By Theorem 5.1 we have $\Phi_X(\delta + \epsilon) = O(n^2p^2 \log(1/p))$ and hence,

$$\left( \frac{\sqrt{m}}{\log^2(n)} \right)^2 \leq \frac{O(n^2p^2 \log(1/p))}{\log^4(n)} \ll \frac{n^2p^2}{\log(1/p)}.$$  

Thus, by Lemma 6.6 for sufficiently large $n_0$ no two vertices of $A$ can be connected by an edge. Therefore,

$$m \geq \sum_{v \in A} \deg(v).$$

Since $\deg(v) \geq 2$ for all of the vertices of $G$, we obtain $|A| \leq \frac{m}{2}$.

Therefore for large enough $n_0$,

$$|V(G)| = |A \cup A^c| = |A| + |A^c| \leq \frac{m}{2} + 2\sqrt{m} \cdot \log^2(n).$$

Provided $n_0$ is large enough, $2\sqrt{m}\log^2(n) \leq m^{3/4}$. Hence, for large enough $n_0$ the assertion of the lemma holds.

We now formalize the discussion after Corollary 5.6 in the following proposition.

**Proposition 6.8.** Suppose $\delta$ is a positive real and $\epsilon$, $\eta$ are sufficiently small positive reals. Furthermore, suppose $\frac{\log^2(n)}{n} \leq p \ll \frac{1}{\sqrt{n \log(n)}}$. Then, there exists $n_0$ such that for all $n > n_0$ we have,

$$\log \mathbb{P}(X \geq (1 + \delta)E[X]) \leq (1 - \eta) \max_{m_0 \leq m \leq m_2} \left\{ m \log(p) - v_m \log(np^2) \right\}.$$  

To prove this proposition we start with a lemma reducing the problem of estimating $m \log(p) - v_m \log(np^2)$ only when $m_0 \leq m \leq K\Phi_X(\delta + \epsilon)$.

**Lemma 6.9.** Suppose $\delta$ is a positive real and $\epsilon$, $\eta$ are sufficiently small positive reals. Furthermore, suppose $n^{-1} \ll p \ll n^{-1/2} / \log(n)$. Then, there exists $K, n_0 > 0$ such that for all $n > n_0$ we have,

$$\log \mathbb{P}(X \geq (1 + \delta)E[X]) \leq (1 - \eta) \max_{m_0 \leq m \leq K\Phi_X(\delta + \epsilon)} \left\{ m \log(p) - v_m \log(np^2) \right\}.$$  

**Proof.** For all $\eta_1, K' > 0$ Corollary 5.6 asserts that provided $n_0$ is large enough we have the following for all $m_0 \leq m \leq K'\Phi_X(\delta + \epsilon)$:

$$\log(p^m|C_m|) \leq m \log(p) - v_m \log(np^2) + \eta_1 m \log(n),$$

where $v_m = \max \{|v \in V(G) : G \in C_m \text{ and } \deg(v) > 0\}|$.

Since $n^{-1} \ll p \ll \frac{1}{\sqrt{n \log(n)}}$, provided $n$ is sufficiently large, we have $v_m \leq m/2 + m^{3/4}$ for all $m_0 \leq m \leq K'\Phi_X(\delta + \epsilon)$ by Lemma 6.7. Thus,

$$m \log(p) - v_m \log(np^2) \leq -\Omega(m \log(n)).$$
For any positive \( \eta_2 \) we may choose \( \eta_1 \) small enough such that we have the following for sufficiently large \( n \),

\[
m \log(p) - v_m \log(np^2) + \eta_1 m \log(n) \leq (1 - \eta_2)(m \log(p) - v_m \log(np^2)).
\]

Therefore, provided \( \eta_2 \) is sufficiently small Theorems 3.2 and 5.1 then yield that there exist \( K, n_0 > 0 \) such that for all \( n > n_0 \):

\[
\log \mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X]) \leq (1 - \eta) \max_{m_0 \leq m \leq K \Phi_\delta(\delta + \varepsilon)} \left\{ m \log(p) - v_m \log(np^2) \right\}.
\]

We now use the above lemma to prove Proposition 6.8.

**Proof of Proposition 6.8.** By Lemma 6.9 it is sufficient to prove that for all \( m_2 \leq m \leq K \Phi_\delta(\delta + \varepsilon) \) we have the following provided \( \eta' \) is sufficiently small

\[
m \log(p) - v_m \log(np^2) \leq (1 - \eta')(m_2 \log(p) - v_{m_2} \log(np^2)).
\]

Provided \( n_0 \) is large enough Lemma 6.7 gives \( v_m \leq m/2 + m^{3/4} \leq m/2 + \eta'm \) for all \( m_2 \leq m \leq K \Phi_\delta(\delta + \varepsilon) \). Therefore,

\[
m \log(p) - v_m \log(np^2) \leq m \log(p) - (1/2 + \eta')m \log(np^2) \\
\leq (-1/2 - \eta)m \log(n) \leq (-1/2 - \eta')m_2 \log(n).
\]

Noting that each copy of \( K_{m_2/2,2} \) in \( K_n \) is a core graph with \( m_2 \) edges we find that \( v_{m_2} \geq m_2/2 \). Therefore,

\[
m_2 \log(p) - v_{m_2} \log(np^2) \geq m_2 \log(p) - \frac{m_2}{2} \log(np^2) = -\frac{m_2}{2} \log(n).
\]

These inequalities imply the following for all \( m_2 \leq m \leq K \Phi_\delta(\delta + \varepsilon) \) and sufficiently small \( \eta'' \)

\[
m \log(p) - v_m \log(np^2) \leq (1 - \eta'')(m_2 \log(p) - v_{m_2} \log(np^2)).
\]

This establishes the proposition provided \( \eta'' \) is small enough.

Proposition 6.8 and Corollary 5.6 show that to prove Theorem 6.1 when \( \frac{\log^\alpha(n)}{n} \leq p \ll \frac{1}{\sqrt{n \log(n)}} \) we only need to bound \( m \log(p) - v_m \log(np^2) \) for \( m_0 \leq m \leq m_2 \). Therefore, to prove the first item in Theorem 6.1 it is enough to prove that, when \( n^{-1+\varepsilon-1} \leq p \leq n^{-1+\varepsilon} \) we have

\[
m \log(p) - v_m \log(np^2) \leq (1 - \eta)(m_2 \log(p) - v_{m_2} \log(np^2)),
\]

for all \( m_0 \leq m \leq m_2 \). Moreover, to prove the second item in Theorem 6.1 it is enough to prove that, when \( n^{-2/3} \leq p \ll \frac{1}{\sqrt{n \log(n)}} \) we have

\[
m \log(p) - v_m \log(np^2) \leq (1 - \eta)m_0 \log(p),
\]

for all \( m_0 \leq m \leq m_2 \). We now split into these two cases mentioned above and prove them. First we deal with the denser case where \( n^{-2/3} \leq p \ll \frac{1}{\sqrt{n \log(n)}} \).
6.2.1  The denser case in the sparse regime

Assume $n^{-2/3} \leq p \ll \frac{1}{\sqrt{n \log(n)}}$. Proposition 6.8 implies that we need to evaluate the term $m \log(p) - v_m \log(np^2)$ only for $m_0 \leq m \leq m_2$. To do this we use Lemma 6.2 which implies a connection between the number of induced copies of $C_4$ in a graph and the number of edges and vertices in it. We start with a simple corollary of Lemma 6.2.

**Corollary 6.10.** Suppose $\varepsilon, \delta, K$ are positive reals. Suppose further that $\frac{\log(n)}{n} \leq p \ll 1$ and $m$ is an integer. Then, for any $G \in C_m$ the number of vertices of $G$ is at most

$$m - 4(\delta - \varepsilon)E[X]/m + 1.$$  

*Proof.* Since the minimum degree of $G$ is at least 2, we may apply Lemma 6.2 and obtain,

$$(\delta - \varepsilon)E[X] \leq N(C_4, G) \leq N_{ind}(v(G), m, C_4) \leq \frac{m(m - v(G) + 1)}{4} \quad (17)$$

By algebraic manipulations we see that, $v(G) \leq m - 4(\delta - \varepsilon)E[X]/m + 1$ which is the assertion of the corollary. □

Now we obtain the desired bound for $m \log(p) - v_m \log(np^2)$ in this regime of $p$.

**Lemma 6.11.** Suppose that $\varepsilon, \delta, \gamma$ are positive reals with $\varepsilon$ small enough, and $\gamma$ small enough as a function of $\varepsilon$. Furthermore, suppose $n^{-2/3} \leq p \ll \frac{1}{\sqrt{n \log(n)}}$ and $K(\varepsilon, \delta)$ is some constant that might depends on $\varepsilon$ and $\delta$. Then there exists $n_0$ such that for any $n > n_0$ and any $m_0 \leq m \leq m_2$ we have,

$$m \log(p) - v_m \log(np^2) \leq (1 - \gamma)m_0 \log(p).$$

*Proof.* Let $c \in [1/3, 1/2]$ be such that $p = n^{-1+c}$ and recall that $m = \Theta(n^2p^2)$. Applying Corollary 6.10 we obtain the following for any $m_0 \leq m \leq m_2$ and sufficiently large $n_0$:

$$m \log(p) - v_m \log(np^2) \leq (-1 + c)m \log(n) + (1 - 2c)(m - 4(\delta - \varepsilon)E[X]/m + \eta m) \log(n)$$

$$\leq -cm \log(n) - 4(1 - 2c)(\delta - \varepsilon)E[X] \log(n)/m + \eta m \log(n)/3.$$  

Recalling that $m_0 = 2\sqrt{(\delta - \varepsilon)E[X]}$ we may rewrite the above as follows provided $n_0$ is large enough:

$$m \log(p) - v_m \log(np^2) \leq (-cm - (1 - 2c)m_0^2/m) \log(n) + \eta m \log(n)/3.$$  

We claim that $g_c(m) := -cm - (1 - 2c)m_0^2/m \leq (-1 + c)m_0$ for all $m_0 \leq m \leq m_2$. Indeed, $g_c(m) = -c + (1 - 2c)(m_0/m)^2 \leq 1 - 3c \leq 0$ for $m \geq m_0$ and $c \geq 1/3$. This concludes the proof as now provided $\gamma$ is small enough we have the following for all $m_0 \leq m \leq m_2$:

$$m \log(p) - v_m \log(np^2) \leq (-1 + c)m_0 \log(n) + \eta m \log(n)/3$$

$$= m_0 \log(p) + \eta m \log(n)/3 \leq m_0 \log(p) + \eta m_2 \log(n)/3$$

$$\leq (1 - \gamma)m_0 \log(p).$$  

This together with Proposition 6.8 conclude the proof for the second item in Theorem 6.1.
6.2.2 The sparser case in the sparse regime

Assume \( \frac{\log^2(n)}{n} \leq p \ll n^{-2/3-\epsilon} \). It was already seen that there is a big difference in the behaviour of the upper tail probability depending on the regime \( p \) lies at. In this section we present a surprising change of the behaviour in the sparser regime. This phenomenon comes from the fact that the number of core graphs is significant and matters a lot in the evaluation of the upper tail probability. This is completely different from the previous cases where the number of cores was negligible. This will be explained in more details later on.

The essence of the proofs in this subsection is still to make use of the connection between the number of induced copies of \( C_4 \) in a core graph and the number of vertices in it. To state the claims in this section it is useful to introduce the following notations.

**Notation.** Suppose \( G \) is a graph and \( k \) is a nonnegative integer. Then we denote the following sets accordingly:

- Denote by \( X_k(G) \) the set of all edges of \( G \) with an endpoint of degree \( k \) and denote its cardinality by \( x_k(G) \).
- Denote by \( X_{>k}(G) \) the set of all edges of \( G \) whose both endpoints have degree greater than \( k \); similarly we denote by \( x_{>k}(G) \) the cardinality of \( X_{>k}(G) \).

In most cases we omit ‘\( G \)’ for brevity.

The following lemma uses the above notations to bound from above the number of induced copies of \( C_4 \) in any core graph with \( m \) edges where \( p \ll \frac{1}{\sqrt{n \log(n)}} \). The motivation of giving this upper bound is to reduce the problem to an optimization problem involving \( x_i \).

**Lemma 6.12.** Suppose \( \epsilon, \delta, R, K \) are positive reals and \( \frac{\log(n)}{n} \leq p \ll \frac{1}{\sqrt{n \log(n)}} \). Then there exists \( n_0 \) such that for any \( n > n_0 \) the following holds:

Let \( G \) be a core graph with \( m \) edges. Then

\[
N_{ind}(C_4, G) \leq \sum_{i=2}^{R} \frac{x_i (i)}{2} \left( \sum_{i < j} \frac{x_j}{i} + \frac{x_j}{2i} \right) + \sum_{i=2}^{R} \frac{x_i e(G)(i-1)}{R} + \frac{x_{>R}^2}{4}, \tag{18}
\]

and

\[
v(G) \leq \sum_{i=2}^{R} \frac{x_i}{i} + \frac{2e(G)}{R}. \tag{19}
\]

**Proof.** Let \( G \) be a core graph. First, let us make some notations for the proof. For every nonnegative integer \( k \) let \( V_k(G) \) be set of all vertices of \( G \) with degree \( k \) and denote its cardinality by \( v_k(G) \). Furthermore, let \( V_{>k}(G) \) be the set of all vertices of \( G \) with degree greater than \( k \) and denote its cardinality by \( v_{>k}(G) \).

We now proceed with the proof. Note that taking \( n_0 \) large enough according to Lemma 6.6 applied with \( s = R^2 + 1 \), we are guaranteed that no two vertices of degree less or equal to \( R \) can be connected by an edge. This implies that \( \{X_i : i \leq R\} \cup \{X_{>R}\} \) is a partition of the edges of \( G \). Furthermore, \( v_{>R}(G) \leq 2e(G)/R \) and for any \( i \leq R \) we have \( v_i = x_i/i \). This follows by a standard double counting argument and as the set of all vertices of degree at most \( R \) is an independent set. We thus obtain (19) as

\[
v(G) = \sum_{i=2}^{R} v_i(G) + v_{>R}(G) \leq \sum_{i=2}^{R} \frac{x_i}{i} + \frac{2e(G)}{R}.
\]
To prove (18) we note that there are three types of induced copies of $C_4$ in core graphs:

(i) Induced copies of $C_4$ with exactly two vertices in $\bigcup_{i=2}^{R} V_i$.
(ii) Induced copies of $C_4$ with exactly one vertex in $\bigcup_{i=2}^{R} V_i$.
(iii) Induced copies of $C_4$ with no vertices in $\bigcup_{i=2}^{R} V_i$ that is, all vertices in $V_{R}$.

Let us bound from above the number of induced copies of $C_4$ of each type. We claim that the number of induced copies of $C_4$ of the first type is at most

$$\sum_{i=2}^{R} v_i \left( \sum_{i<j} v_j + \frac{v_i}{2} \right) = \sum_{i=2}^{R} \frac{v_i}{i} \left( \sum_{i<j} \frac{v_j}{j} + \frac{x_j}{2i} \right).$$

Indeed, for every $i$, there are at most $\frac{v_i}{2} \binom{1}{i}$ copies of $C_4$ with two vertices in $V_i$ and, for all $i < j$ there are at most $v_i v_j \binom{i}{2}$ copies of $C_4$ with one vertex in $V_i$ and one vertex in $V_j$.

We now bound the number of induced copies of $C_4$ of the second kind. Note that each such induced copy of $C_4$ is determined by choosing its vertex of degree less or equal to $R$, two of its neighbours and another vertex of degree larger than $R$. Therefore, the number of induced copies of $C_4$ of the second kind is at most

$$\sum_{i=2}^{R} v_{R} v_i \binom{i}{2} = \sum_{i=2}^{R} x_i v_{R} (i-1) \leq \sum_{i=2}^{R} x_i e(G)(i-1) \frac{R}{R}.$$ 

To bound from above the number of induced copies of $C_4$ of the third kind we observe that each such induced copy of $C_4$ is determined by each of the two perfect matchings in it. Since each induced copy of $C_4$ of the third type contains only edges of $X_{R}$, there are at most $\frac{1}{2} \binom{x_{R}}{2} \leq \frac{x_{R}^2}{4}$ such copies. Summing all of these bounds we obtain the assertion of the lemma.

Our only use of Lemma 6.12 is when $R = \lceil 1/\epsilon \rceil$. Therefore, from now on we set $R = \lceil 1/\epsilon \rceil$. We will only consider core graphs $G$ with at most $m_2$ edges, thus we may bound from above the right-hand side of (18) by

$$f(x_1, x_2, \ldots, x_R, x_{R}) = \sum_{i=2}^{R} \binom{i}{2} x_i \left( \sum_{i<j} \frac{x_j}{j} + \frac{x_i}{2i} \right) + \epsilon m_2 \sum_{i=2}^{R} (i-1)x_i + \frac{x_{R}^2}{4}$$

$$= \sum_{i=2}^{R} (i-1)x_i \left( \sum_{i<j} \frac{x_j}{2j} + \frac{x_i}{4i} + \epsilon m_2 \right) + \frac{x_{R}^2}{4}.$$ 

In particular, for every core graph $G$ with at most $m_2$ edges,

$$(\delta - \epsilon)E[X] \leq N_{ind}(C_4, G) \leq f(x_1(G), x_2(G), \ldots, x_R(G), x_{R}(G)).$$

Recall that in this range of $p$ we have $\log(np^2) < 0$ and note that by Lemma 6.12 for any core graph $G$ the term $e(G) \log(p) - v(G) \log(np^2)$ can be bounded from above using $x_i(G)$ by

$$\sum_{i=2}^{R} \left( x_i(G) \log(p) - x_i(G) \frac{\log(np^2)}{i} + x_{R}(G) \log(p) - \frac{2e(G) \log(np^2)}{R} \right).$$
Furthermore, for $\varepsilon < \eta/2$ we have,

$$2e(G)/R \leq 2e(G)\varepsilon \leq \eta e(G).$$

Therefore, letting $n_0$ sufficiently large we find that,

$$\max_{m_0 \leq m \leq m_2} m \log(p) - v_m \log(np^2) \leq \max_{G \in \bigcup_{m \in m_0} C_m} \sum_{i=2}^{R} \left( x_i(G) \log(p) - x_i(G) \log(np^2)/i \right)$$

$$+ x_{>R}(G) \log(p) + \eta m_2 \log(n). \tag{20}$$

We now restate this bound in a more compact way. To this end we introduce the following notations.

- $x(G) \in \mathbb{R}^{R+1}$ is the vector defined by $x_i(G)$ in its $i$-th coordinate for $1 \leq i \leq R$ and $x_{>R}(G)$ in the last coordinate.
- $u \in \mathbb{R}^{R+1}$ is defined by 0 as the first coordinate, $\log(p) - \log(np^2)/i$ as the $i$-th coordinate for $2 \leq i \leq R$, and $\log(p)$ in the last coordinate.

Using these notations we may rewrite (20) as follows,

$$\max_{m_0 \leq m \leq m_2} m \log(p) - v_m \log(np^2) \leq \max_{G \in \bigcup_{m \in m_0} C_m} \left( x(G), u \right) + \eta m_2 \log(n). \tag{21}$$

From now on we let $t = (\delta - \varepsilon)\mathbb{E}[X]$. Since $t \leq f(x(G))$ for every $G \in \bigcup_{m \in m_0} C_m$,

$$\max_{G \in \bigcup_{m \in m_0} C_m} \left( x(G), u \right) \leq \max \{ \left( x, u \right) : x \in \mathbb{R}^{R+1}_{\geq 0} \wedge f(x) \geq t \}.$$ 

Note that in the above we dropped the condition that $x$ came from a graph. By this we only consider more possible cases and thus obtain an upper bound on the maximization problem.

Now we are ready to state the main technical proposition in order to bound from above the term $m \log(p) - v_m \log(np^2)$ when \(\frac{\log^k(n)}{n} \leq p \leq n^{-2/3-\varepsilon}\).

**Proposition 6.13.** Suppose $k > 1$ is an integer and suppose that $\eta, \varepsilon, \delta$ are positive reals with $\varepsilon$ small enough as a function of $\eta$ and $k$. Then, there exists $n_0$ such that for any $n > n_0$ and any $n^{-1+c_{\delta-1}} \leq p \leq n^{-1+c_k}$ we have

$$\max \{ \left( x, u \right) : x \in \mathbb{R}^{R+1}_{\geq 0} \wedge f(x) \geq t \} \leq (1 - \eta) m_k u_k.$$ 

This proposition and Proposition 6.8 imply the following corollary. This corollary yields the first item in Theorem 6.1 as explained before.

**Corollary 6.14.** Suppose $k > 1$ is an integer and suppose that $\eta, \varepsilon, \delta, K$ are positive reals with $\varepsilon$ small as a function of $\eta$ and $k$. Then, there exists $n_0$ such that for any $n > n_0$ and any $\max \left\{ \frac{\log^k(n)}{n}, n^{-1+c_{\delta-1}} \right\} \leq p \leq n^{-1+c_k}$ we have the following:

$$m \log(p) - v_m \log(np^2) \leq (1 - \eta)m_k u_k = (1 - \eta)m_k (\log(p) - \log(np^2)/k)$$

for every $m_0 \leq m \leq K\Phi_X(\delta + \varepsilon)$. 


We now turn to the proof of Proposition 6.13. We show that for \( \max \left\{ \frac{\log^p(n)}{n}, n^{-1+\epsilon_k} \right\} \leq p \leq n^{-1+\epsilon_k} \) the maximum of \( \langle x, u \rangle \) is achieved by a vector \( x \) of the form \( \alpha \cdot e_k \) where \( e_k \) is the \( k \)-th element in the standard basis. The strategy of the proof is to think of the vector \( x \) as a distribution vector and push its mass towards the ‘center of mass’ while keeping \( \langle x, u \rangle \) constant and ensuring that \( f \) evaluated on the vector after the transformation is still greater than \( t \). To be more precise, the center of mass will be the \( k \)-th coordinate such that \( k \) satisfies \( \max \left\{ \frac{\log^p(n)}{n}, n^{-1+\epsilon_k} \right\} \leq p \leq n^{-1+\epsilon_k} \). This will be done iteratively by showing that if \( x \) is a vector achieving the maximum of \( \langle x, u \rangle \) and \( x_i = 0 \) for all \( i \leq j < k \) then we may define a vector \( \tilde{x} \) which also achieves this maximum and satisfies \( \tilde{x}_i = 0 \) for all \( i \leq j+1 \). A similar statement will be shown for the other direction i.e. pushing the mass from the right to the left.

Let us introduce the following notations of pushing mass. Let \( 2 \leq k \leq R \) be some integer (should be thought of as the center of mass) and suppose that \( x \in \mathbb{R}^{R+1} \). Then for any integers \( 2 \leq i < k < j \leq R+1 \) define the following vectors,

\[
x^{(i,k)} = x - x_i e_i + \frac{u_i x_j}{u_{i+1}} e_{i+1},
\]

\[
x^{(j,k)} = x - x_j e_j + \frac{u_j x_i}{u_{j-1}} e_{j-1}.
\]

One should think of these operations as pushing the mass towards the \( k \)-th coordinate while staying on the hyperplane defined by \( \langle x, u \rangle = s \) for some constant \( s \).

**Proof of Proposition 6.13.** As explained earlier we prove this proposition iteratively. We do so in several claims. First, we show that we may assume that the distribution vector \( x \) satisfies \( x_{R+1} = 0 \). This is given in the following claim.

**Claim 6.15.** If \( p \leq n^{-2/3-\eta} \) for some positive \( \eta \), then provided \( \epsilon \) is small enough there exists \( n_0 \) such that for any \( n > n_0 \) the following holds:

\[
\max \{ \langle x, u \rangle : x \in \mathbb{R}^{R+1}_{\geq 0} \land f(x) \geq t \} = \max \{ \langle x, u \rangle : x \in \mathbb{R}^{R+1}_{\geq 0} \land x_{R+1} = 0 \land f(x) \geq t \}.
\]

**Proof.** Let

\[
s = \max \{ \langle x, u \rangle : x \in \mathbb{R}^{R+1}_{\geq 0} \land f(x) \geq t \}.
\]

Furthermore, let \( x \in \mathbb{R}^{R+1}_{\geq 0} \) be such that \( s = \langle x, u \rangle \) and \( f(x) \geq t \). By the definition of \( x^{(R+1,k)} \) we have,

\[
\langle x, u \rangle = \langle x^{(R+1,k)}, u \rangle
\]

and therefore, \( s = \langle x^{(R+1,k)}, u \rangle \). In order to prove the lemma it is sufficient to prove \( f(x^{(R+1,k)}) \geq f(x) \). To do so let

\[
f_1(x_1, x_2, \ldots, x_{R+1}) = \sum_{i=2}^{R-1} (i - 1) x_i \left( \sum_{i \leq j \leq R-1} \frac{x_j}{2j} + \frac{x_i}{4i} + \epsilon m_2 \right),
\]

which depends only on \( x_1, x_2, \ldots, x_{R-1} \), and

\[
f_2(x_1, x_2, \ldots, x_{R+1}) = \sum_{i=2}^{R-1} \frac{(i - 1) x_i x_R}{2R} + (R - 1) x_R \left( \frac{x_R}{4R} + \epsilon m_2 \right) + \frac{x_{R+1}^2}{4}.
\]
Note that for all \( y \in \mathbb{R}^{R+1} \) we have
\[
f(y) = f_1(y) + f_2(y).
\]

Note also that for all \( 1 \leq i \leq R - 1 \) we have \( x_i^{(R+1,k)} = x_i \). Therefore, we obtain that
\[
f(x^{(R+1,k)}) - f(x) = f_2(x^{(R+1,k)}) - f_2(x).
\]
As \( x_i^{(R+1,k)} = x_R + \frac{u_{R+1}}{u_R} x_{R+1} \) and \( x_i^{(R+1,k)} = 0 \) a straightforward computation gives
\[
f(x^{(R+1,k)}) - f(x) = \left( \sum_{i=2}^{R} \frac{(i-1)x_i}{2R} + \epsilon m_2(R-1) \right) \frac{u_{R+1}}{u_R} x_{R+1}
+ \left( \frac{(R-1)u_{R+1}^2}{Ru_R^2} - 1 \right) \frac{x_{R+1}^2}{4}.
\]

Note that \( u_{R+1} = \log(p) < 0 \) and \( u_R < 0 \). This holds as \( u_R = \log(p) - \log(np^2)/R \) and we assume \( \epsilon \) is small enough so that \( p^{R-2} \leq n \), as then the left-hand side decays to zero and the right-hand side approaches infinity. This implies that
\[
\frac{u_{R+1}}{u_R} > 0.
\]

Therefore,
\[
f(x^{(R+1,k)}) - f(x) \geq \left( \frac{(R-1)u_{R+1}^2}{Ru_R^2} - 1 \right) \frac{x_{R+1}^2}{4}.
\]

Proving that the above is nonnegative provided \( n_0 \) is sufficiently large will conclude the proof. Indeed, letting \( y = \frac{\log(np^2)}{R \log(p)} \)
\[
\frac{(R-1)u_{R+1}^2}{Ru_R^2} = \frac{(R-1)\log^2(p)}{R(\log(p) - \log(np^2)/R)^2} \geq 1
\]
is equivalent to,
\[
-y^2 + 2y - 1/R \geq 0,
\]
which is also equivalent to
\[
1 - \sqrt{1 - 1/R} \leq y \leq 1 + \sqrt{1 - 1/R}.
\]

Since \( 1 - \sqrt{1-x} \leq x/2 + x^2 \) and \( 1 + \sqrt{1-x} \geq 2 - x \) for all \( 0 \leq x \leq 1 \), it is enough to prove that
\[
2/R + 1/R^2 \leq \frac{\log(np^2)}{R \log(p)} \leq 2 - 1/R.
\]

First, as \( \log(p) < 0 \), the right inequality is equivalent to
\[
\log(np^2) \geq (2R - 1) \log(p).
\]
Equivalently, \( \log(n) \geq (2R - 3) \log(p) \), which holds for \( \varepsilon \) satisfying \( 2R - 3 \geq 0 \) as then the left-hand side is positive for \( n > 1 \) and the right-hand side tends to minus infinity as \( p \) tends to infinity. Second, as \( np^2 \ll 1 \), the left inequality is equivalent to

\[
(1/2 + 1/R) \log(p) \geq \log(np^2).
\]

This holds if and only if \( p \leq n^{-1/(3/2-1/R)} \). Finally, assuming that \( \varepsilon \) is small enough so that \( 1/R \leq 3/2 - 1/(2/3 + \eta) \) implying the claim. ■

We now continue to the more technical part of the proof which is to push the mass to the \( k \)-th coordinate provided that the last coordinate is 0, which we may assume due to the above claim.

**Claim 6.16.** If \( n^{-1+\varepsilon_{i-1}} \leq p \leq n^{-1+\varepsilon_i} \) then there exists \( n_0 \) such that for any \( n > n_0 \) the following holds:

\[
\max\{\langle x, u \rangle : x \in \mathbb{R}_{\geq 0}^{R+1} \land f(x) \geq t\} = \max\{au_k : f(au_k) \geq t\}.
\]

**Proof.** Let

\[
s = \max\{\langle x, u \rangle : x \in \mathbb{R}_{\geq 0}^{R+1} \land f(x) \geq t\}.
\]

Furthermore, let \( x \in \mathbb{R}_{\geq 0}^{R+1} \) be such that \( s = \langle x, u \rangle \) and \( f(x) \geq t \). By Claim 6.15 we may assume \( x_{R+1} = 0 \). As \( x \) satisfies \( \langle x, u \rangle = s \) we also have \( \langle x^{(i,k)}, u \rangle = \langle x^{(i,k)}, u \rangle = s \) for any \( 2 \leq i < k < j \leq R \).

We will start with the push of the mass from the left to the right. To this end we prove iteratively the following: Let \( x \in \mathbb{R}_{\geq 0}^{R+1} \) and suppose that \( x_r = 0 \) for all \( 1 \leq r < i < k \). Then \( f(x^{(i,k)}) - f(x) \geq 0 \) provided \( n_0 \) is large enough. To see this let

\[
f_i(x_1, x_2, \ldots, x_{R+1}) = \sum_{\ell = i+2}^{R} (\ell - 1)x_{\ell} \left( \sum_{\ell < j \leq R} \frac{x_j}{2j} + \frac{x_\ell}{4\ell} + \varepsilon m_2 \right),
\]

which depends only on \( x_{i+2}, x_{i+3}, \ldots, x_R \), and let

\[
f_2(x_1, x_2, \ldots, x_{R+1}) = \sum_{\ell = i}^{i+1} (\ell - 1)x_{\ell} \left( \sum_{\ell < j \leq R} \frac{x_j}{2j} + \frac{x_\ell}{4\ell} + \varepsilon m_2 \right),
\]

which depends only on \( x_i, x_{i+1}, \ldots, x_R \). Note that for all \( y \) with \( y_r = 0 \) for \( r < i \) we have \( f(y) = f_1(y) + f_2(y) \). This implies that \( f(x^{(i,k)}) - f(x) = f_2(x^{(i,k)}) - f_2(x) \). By the definition of \( x^{(i,k)} \) we have

\[
f(x^{(i,k)}) - f(x) = \left( \sum_{i+1 \leq j} \frac{x_ix_j}{2j} + \varepsilon m_2x_i \right) \left( \frac{iu_i}{u_{i+1}} - (i - 1) \right) + \frac{x_i^2}{4} \left( \frac{iu_i^2}{(i + 1)u_{i+1}^2} - \frac{i - 1}{i} \right).
\]

Thus, to show that \( f(x^{(i,k)}) - f(x) \geq 0 \) it is suffices to show the following:

\[
\frac{iu_i}{u_{i+1}} - (i - 1) \geq 0,
\]

(22)
As $np^2 = o(1)$ we have $0 > u_i > u_{i+1}$. Therefore, inequality (22) is equivalent to

$$i \log(p) - \log(np^2) \leq (i - 1) \log(p) - \frac{i - 1}{i + 1} \log(np^2).$$

By algebraic manipulation the above is equivalent to

$$\log(p) \leq \frac{2}{i + 1} \log(np^2).$$

This always holds in our settings provided $n$ is sufficiently large. Indeed, the above is equivalent to $p^{i-3} \leq n^2$ which holds as $n^{-1} \leq p \ll 1$. It is left to show inequality (23). We prove this in the following claim.

Claim 6.17.

$$i^2 u_i^2 \geq (i + 1)(i - 1)u_{i+1}^2 \iff p \geq n^{-1+c_i}.$$

Proof. The inequality in the left-hand side is equivalent to $i u_i \leq \sqrt{(i + 1)(i - 1)} u_{i+1}$ as $0 > u_i > u_{i+1}$. Recall that $u_\ell = \log(p) - \log(np^2)/\ell$ and plug this in the above inequality to obtain

$$i \log(p) - \log(np^2) \leq \sqrt{(i - 1)(i + 1)} \log(p) - \sqrt{\frac{i - 1}{i + 1}} \log(np^2).$$

Equivalently,

$$\left(i - 2 - \sqrt{(i + 1)(i - 1)} + 2 \sqrt{\frac{i - 1}{i + 1}}\right) \log(p) \leq \left(1 - \sqrt{\frac{i - 1}{i + 1}}\right) \log(n).$$

Thus we conclude that our assumption is equivalent to the following:

$$\log(p) \geq \frac{\sqrt{i + 1} - \sqrt{i - 1}}{(i - 2) \sqrt{i + 1} - (i - 3)^{3/2}} \log(n) = \left(\frac{1}{2 + \sqrt{\frac{i + 1}{i - 1}}} - 1\right) \log(n) = (-1 + c_i) \log(n).$$

Since we assumed $p \geq n^{-1+c_{k-1}}$ and for all $i < k$ we have $c_i \leq c_{k-1}$ we also have $p \geq n^{-1+c_{k-1}} \geq n^{-1+c_i}$. Therefore, for any $x \in \mathbb{R}^{R+1}$ we may iterate this process for $2 \leq i \leq k-1$ and obtain that $f(x^{(2,k)(3,k)\ldots(k-1,k)}) \geq f(x)$. This conclude the push of the mass of $x$ from the $i$-th coordinates for $i < k$ to the $k$-th coordinate.

Now provided what we showed so far we prove a similar statement for $x^{(i,k)}$. More specifically, we prove iteratively the following: Let $x \in \mathbb{R}^{R+1}$ and suppose that $x_r = 0$ for all $k < j < r \leq R + 1$ and $x_\ell = 0$ for all $\ell < k$. Then $f(x^{(i,k)}) - f(x) \geq 0$ provided $n_0$ is large enough. To see this let

$$f_3(x_1, x_2, \ldots, x_{R+1}) = \sum_{i=k}^{j-2} (i - 1) x_i \left(\sum_{i<\ell \leq j-2} \frac{x_\ell}{2\ell} + \frac{x_j}{4i} + \varepsilon m_2\right).$$
which depends only on \(x_i\) for \(k \leq i \leq j - 2\), and let

\[
f_4(x_1, x_2, \ldots, x_{R+1}) = \sum_{i=k}^{j-2} (i-1)x_i \sum_{\ell=j-1}^j \frac{x_{\ell}}{2\ell} + \sum_{i=j-1}^j (i-1)x_i \left( \sum_{i<\ell \leq j} \frac{x_{\ell}}{2\ell} + \frac{x_i}{4i} + \epsilon m_2 \right),
\]

which depends only on \(x_i\) for \(k \leq i \leq j\).

Note that for all \(y\) with \(y_i = 0\) for all \(i \leq k\) or \(i \geq j + 1\) we have \(f(y) = f_3(y) + f_4(y)\). Note also that \(x^{(i,k)}_i = x_i\) for all \(i \leq j - 2\) thus, \(f(x^{(i,k)}) - f(x) = f_4(x^{(i,k)}) - f_4(x)\). This implies that

\[
f(x^{(i,k)}) - f(x) = \sum_{i=k}^{j-1} \frac{(i-1)x_ix_j}{2} \left( \frac{u_j}{u_{j-1}(j-1)} - \frac{1}{j} \right) + \frac{x_j^2}{4} \left( \frac{(j-2)u_j^2}{(j-1)u_{j-1}^2} - \frac{j-1}{j} \right).
\]

Thus, to show that \(f(x^{(i,k)}) - f(x) \geq 0\) it is sufficient to show the following:

\[
\frac{u_j}{u_{j-1}} \geq \frac{j-1}{j},
\]

and

\[
\frac{u_j^2}{u_{j-1}^2} \geq \frac{(j-1)^2}{j(j-2)}.
\]

Note that these two inequalities are respectively equivalent to the following inequalities:

\[
j \log(p) - \log(np^2) \leq (j-1) \log(p) - \log(np^2), \tag{24}
\]

\[
\frac{u_j^2}{u_{j-1}^2} \leq \frac{j(j-2)}{(j-1)^2}. \tag{25}
\]

Inequality (24) holds always as \(p < 1\). Moreover, by Claim 6.17 inequality (25) holds if and only if \(p \leq n^{-1+c_{j-1}}\). This indeed holds as we assume \(p \leq n^{-1+c_i}\) and we also have \(c_k \leq c_{j-1}\) for all \(j > k\). This conclude the proof of the lemma.

Now we deduce the proposition from the above claims. Assuming \(n_0\) is large enough and using Claim 6.16 it is sufficient to bound from above

\[
s := \max \{ au_k : f(\alpha e_k) \geq t \}.
\]

Let \(\alpha \in \mathbb{R}\) be a witness for that. We have,

\[
t \leq f(\alpha e_k) = \alpha^2 \frac{k - 1}{4k} + \epsilon \alpha (k - 1)m_2
\]

\[
= \left( \frac{\delta + \epsilon}{m_k^2} \right) \mathbb{E}[X] \alpha^2 + 2\epsilon \alpha (k - 1) \sqrt{\frac{\delta + \epsilon}{m_k^2} \mathbb{E}[X]}
\]

\[
= \frac{\delta + \epsilon}{\delta - \epsilon} \cdot \left( \frac{\alpha}{m_k} \right)^2 \left( t + 4\epsilon \sqrt{k(k-1)} \left( \frac{\alpha}{m_k} \right) t \right).
\]
Denoting $x = \alpha/m_k$ implies

$$x \geq -2\varepsilon\sqrt{k(k-1)} + 2\sqrt{\varepsilon^2 k(k-1) + \frac{\delta - \varepsilon}{\delta + \varepsilon}} \quad \text{or} \quad x \leq -2\varepsilon\sqrt{k(k-1)} - 2\sqrt{\varepsilon^2 k(k-1) + \frac{\delta - \varepsilon}{\delta + \varepsilon}}.$$ 

Hence, for every positive $\eta$ and sufficiently small $\varepsilon$ we obtain

$$x \geq 1 - \eta \quad \text{or} \quad x \leq -1 + \eta.$$ 

The second option is not possible as $\alpha$ is nonnegative. Thus we have, $\alpha/m_k \geq 1 - \eta$ and hence

$$s = au_k \leq (1 - \eta)u_k.$$ 

This is as claimed.

\[\Box\]

### 7 THE SOLUTION TO THE VARIATIONAL PROBLEM

In this section we solve the naive mean field variational problem (when $\sqrt{\log(n)/n} \ll p \ll n^{-1/2-\varepsilon}$) for the upper tail of the number of induced copies of $C_4$ in $G_{n,p}$. Let us recall the definition of the variational problem.

Let $N$ be a positive integer and let $Y = (Y_1, Y_2, \ldots, Y_N)$ be a sequence of independent Bernoulli random variables with mean $p$. Further, let $X$ be a function from the hypercube to $\mathbb{R}$. Then the naive mean field variational problem associated to the upper tail of $X(Y)$ is the function $\Psi^\star_X : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ such that for every $\delta \geq 0$:

$$\Psi^\star_X(\delta) = \inf \left\{ \sum_{i \in N} I_p(z_i) : \mathbb{E}[X(Z)] \geq (1 + \delta)\mathbb{E}[X(Y)] \right\},$$

where $I_p(q) = D_{KL}(\text{Ber}(q) || \text{Ber}(p)) = q \log\left(\frac{q}{p}\right) + (1 - q) \log\left(\frac{1-q}{1-p}\right)$, and the infimum is taken over all $Z = (Z_1, Z_2, \ldots, Z_N)$ sequences of $N$ Bernoulli random variables with means $z_i$ respectively.

We will be interested in the case where $N = \left(\frac{n}{2}\right)$, the random variables $Y_i$ correspond to the edges of $G_{n,p}$, and $X$ is the random variable counting the number of induced copies of $C_4$. The main proposition of this section is the following.

**Proposition 7.1.** Suppose $\varepsilon, \delta$ are positive reals. Further, suppose that $\sqrt{\log(n)/n} \ll p \ll n^{-1/2-\varepsilon}$. Then, for large enough $n$, we have

$$(1 - \varepsilon)\sqrt{\frac{\delta}{2}} \leq \frac{\Psi^\star_X(\delta)}{n^2p^2 \log(1/p)} \leq (1 + \varepsilon)\sqrt{\frac{\delta}{2}}.$$ 

The upper bound follows immediately from

$$\Psi^\star_X(\delta) \leq \log(1/p) \min\{e(C) : \mathbb{E}[X|C \subseteq G_{n,p}] \geq (1 + \delta)\mathbb{E}[X]\}.$$ 

In Section 4 we gave an upper bound for this minimum. It is attained when $C$ is the complete bipartite graph $K_{m_0,m_0}$ where $m_0^2 \approx \sqrt{\frac{2}{\varepsilon} n^2p^2}$. For more details the reader is referred to Section 4.
For the lower bound we start with the following reduction.

**Lemma 7.2.** Suppose \( \epsilon, \delta \) are positive reals. Furthermore, suppose \( p \ll 1 \). Then, for large enough \( n \), we have

\[
\Psi_X(\delta) \geq \inf_{\tilde{\eta}} \left\{ \sum_{i \in \binom{[n]}{2}} I_p(q_i) : \mathbb{E} \left[ N_{\text{ind}}(C_4, G_{n, \tilde{\eta}}) \right] \geq (1 + \delta - \epsilon) \mathbb{E}[X] \text{ and } \forall i \in \binom{[n]}{2} q_i \geq p \right\}.
\]

To prove the lemma we define the set \( C_4 \) as follows: Let \( C_4 \) be a set of representatives for the set of all tuples \((x, y, z, w) \in [n]^4\) with distinct coordinates and the equivalence relation, \((x, y, z, w) \sim (x', y', z', w')\) if and only if \( \{xy, yz, zw, wx\} = \{x'y', y'z', z'w', w'x'\} \); that is, both \((x, y, z, w)\) and \((x', y', z', w')\) induce the same 4-cycle in \( K_n \).

**Proof.** Let \( \tilde{\eta} \in [0, 1] \binom{n}{2} \) be such that \( \mathbb{E} \left[ N_{\text{ind}}(C_4, G_{n, \tilde{\eta}}) \right] \geq (1 + \delta) \mathbb{E}[X] \). Define \( \bar{r} \) by setting \( r_i = \max\{p, q_i\} \). Since \( I_p(x) \) is monotone decreasing between 0 and \( p \), we obtain that

\[
\sum I_p(q_i) \geq \sum I_p(r_i).
\]

To conclude the proof, we prove that \( \mathbb{E} \left[ N_{\text{ind}}(C_4, G_{n, \tilde{\eta}}) \right] \geq (1 + \delta - \epsilon) \mathbb{E}[X] \). Indeed, provided \( n \) is sufficiently large, we have

\[
\mathbb{E} \left[ N_{\text{ind}}(C_4, G_{n, \tilde{\eta}}) \right] = \sum_{(x, y, z, w) \in C_4} r_{xy}r_{yz}r_{zw}(1 - r_{xz})(1 - r_{yw}) \\
\geq \sum_{(x, y, z, w) \in C_4} q_{xy}q_{yz}q_{zw}(1 - q_{xz})(1 - q_{yw})(1 - p)^2 \\
\geq (1 + \delta)(1 - p)^2 \mathbb{E}[X] \geq (1 + \delta - \epsilon) \mathbb{E}[X].
\]

Next we present a lemma about the structure of \( G_{n, \tilde{\eta}} \) when \( \sum I_p(q_i) \) is close to the infimum in Lemma 7.2.

**Lemma 7.3.** Suppose \( \epsilon, \delta \) are positive reals with \( \epsilon \) sufficiently small. Suppose also that \( \sqrt{\log(n)/n} \ll p \ll n^{-1/2} \). Then, for any sequence \( \tilde{\eta} \) satisfying:

1. \( q_i = p + u_i \) for some \( u_i \geq 0 \),
2. \( \sum_{i \in \binom{[n]}{2}} I_p(q_i) \leq 2\delta n^2 p^2 \log(1/p), \) and
3. \( \mathbb{E} \left[ N_{\text{ind}}(C_4, G_{n, \tilde{\eta}}) \right] \geq (1 + \delta - \epsilon) \mathbb{E}[X], \)

we have \( \mathbb{E} \left[ N_{\text{ind}}(C_4, G_{n, \tilde{\eta}}) \right] \geq (\delta - 2\epsilon) \mathbb{E}[X] \) provided \( n \) is large enough.

This lemma is a variant of the methods in the papers of Lubetzky and Zhao [27] and in Bhattacharya, Ganguly, Lubetzky and Zhao [Sec. 5.2] [5]. In these papers the authors use the language of graph limits or ‘graphons’; we recreate their proof in the language of graphs in the Appendix. Further, Lubetzky and Zhao [27] and Bhattacharya, Ganguly, Lubetzky and Zhao [5], analysed the function \( I_p(p + x) \) and proved several lemmas in the language of graphons. We use these lemmas in the proof of Proposition 7.1, once again in the language of graphs and not graphons. For a proof of the first lemma we refer the reader to [27] and [5]; the second lemma will be proven in the Appendix with some extensions.
Lemma 7.4 \((27)\). The following holds,

\[
I_p(p + x) = \begin{cases} 
(1 + o(1))\frac{x^2}{2p}, & 0 \leq x \ll p, \\
(1 + o(1))x \log(x/p), & p \ll x \leq 1 - p.
\end{cases}
\]

Lemma 7.5 \((27)\). Suppose \(\epsilon, \delta, C\) are positive reals and \(\sqrt{\log(n)/n} \ll p \ll n^{-1/2}\). Suppose also that \(\bar{u} \in [0, 1 - p]^{(n \choose 2)}\) such that \(\sum_{i \in [n \choose 2]} I_p(p + u_i) \leq Cn^2p^2 \log(1/p)\). Let \(b = b(n)\) be such that \(\max\{np^2, \sqrt{p \log(1/p)}\} \ll b \ll 1 - \epsilon\). Then, provided \(n\) is large enough there is a constant \(D > 0\) such that the following holds:

(i) \(\sum_{x \in [n]} \left(\sum_{y \neq x} u_{xy}\right)^2 \leq Dn^3p^2b\) and

(ii) \(\sum_{i \in [n \choose 2]} u_i \leq Dn^2p^{3/2} \sqrt{\log(1/p)}\).

Now we are ready to prove Proposition 7.1.

Proof of Proposition 7.1. We wish to prove the following:

\[
\sum_{i \in [n \choose 2]} I_p(q_i) \geq (1 - \eta)\sqrt{\frac{\delta}{2} n^2p^2 \log(1/p)}
\]

for sufficiently small \(\eta > 0\) and for \(\bar{q} \in [0, 1]^{(n \choose 2)}\) such that

\[
\mathbb{E}[N_{ind}(C_4, G_{n, \bar{u}})] \geq (1 + \delta)\mathbb{E}[X].
\]

We may assume that

\[
\sum_{i \in [n \choose 2]} I_p(q_i) \leq Cn^2p^2 \log(1/p),
\]

for some absolute constant \(C > 0\) as otherwise there is nothing to prove. By Lemma 7.2 and Lemma 7.3 we may assume that \(q_i = p + u_i\) for some \(u_i \geq 0\) which satisfies

\[
\mathbb{E}[N(C_4, G_{n, \bar{u}})] \geq \mathbb{E}[N_{ind}(C_4, G_{n, \bar{u}})] \geq (\delta - \epsilon)\mathbb{E}[X]
\]

where \(N(C_4, G_{n, \bar{u}})\) is the number of copies of \(C_4\) in \(G_{n, \bar{u}}\). Note that

\[
\mathbb{E}[N(C_4, G_{n, \bar{u}})] = \sum_{(x,y,z,w) \in C_4} u_{xy}u_{yz}u_{zw}u_{wx} \geq (\delta - \epsilon)\mathbb{E}[X].
\]

In the next three claims we collect some properties of \((x, y, z, w) \in C_4\) with non-negligible contribution to the sum above. Afterwards, we use these properties to show that the set of all 4-cycles with non-negligible contribution to the sum above is a subset of

\[
C_4' := \{(a_0, a_1, a_2, a_3) \in C_4 : u_{a_ia_{i+1}} > \gamma \text{ for all } i = 0, 1, 2, 3\},
\]

where the summation is modulo 4 and \(\gamma\) is some positive constant depending on \(\epsilon\). We now explain briefly the proof strategy: First, we show that to have a non-negligible
contribution a 4-cycle \((a_0, a_1, a_2, a_3) \in C_4\) must contain a \(K_{1,2}\) with both edge weights at least \(\sqrt{p}/\log(1/p)\); for convenience assume that these edges are \(a_0a_1, a_1a_2\). Afterwards, we show that, among such 4-cycles, the only ones contributing a non-negligible amount must additionally satisfy \(ua_2a_3 \geq p^{1-2\varepsilon}\) or \(ua_3a_0 \geq p^{1-2\varepsilon}\). Then, we prove that a 4-cycle with the above properties contributes a non-negligible amount only if \(ua_0a_1, ua_2a_3 \geq \gamma\) (in the complementary case we show \(ua_1a_2, ua_3a_1 \geq \gamma\)). Using symmetries of the 4-cycle, we may use this argument again from a different point of view. We then obtain that the only 4-cycles contributing a non-negligible amount to the sum are ones with \(ua_ia_{i+1} \geq \gamma\) for all \(i\). This explanation can also be viewed in Figure 1.

We now execute this plan rigorously.

**Claim 7.6.** For all \(\eta > 0\) the following holds for large enough \(n\)

\[
\sum_{(x,y,z,w) \in A} u_{xy}u_{yz}u_{zw}u_{wx} \leq \eta^2 n^4 p^4,
\]

where \(A\) is the set of all tuples \((x, y, z, w) \in C_4\) with

\[
u_{xy}, u_{zw} \leq \sqrt{p}/\log(1/p) \quad \text{or} \quad u_{xy}, u_{yz} \leq \sqrt{p}/\log(1/p).
\]

**Proof.** Suppose otherwise, meaning,

\[
\sum_{(x,y,z,w) \in A} u_{xy}u_{yz}u_{zw}u_{wx} > \eta^2 n^4 p^4.
\]

By the definition of \(A\) we have

\[
u_{xy}, u_{zw} \leq \sqrt{p}/\log(1/p) \quad \text{or} \quad u_{xz}, u_{yw} \leq \sqrt{p}/\log(1/p).
\]

This implies that

\[
\eta^2 n^4 p^4 < \sum_{(x,y,z,w) \in A} u_{xy}u_{yz}u_{zw}u_{wx} \leq \left(\sum_{x,y} u_{xy}\right)^2 \frac{p}{\log^2(1/p)},
\]

and therefore,

\[
\eta n^2 p^{3/2} \log(1/p) \leq \sum_{x,y} u_{xy}.
\]
By Lemma 7.5 we have $\sum_{x,y} u_{xy} = O\left(n^2 p^{3/2} \sqrt{\log(1/p)}\right)$. This implies,

$$\eta n^2 p^{3/2} \log(1/p) \leq O\left(n^2 p^{3/2} \sqrt{\log(1/p)}\right),$$

contradicting the assumption that $p = o(1)$.

\begin{center}■\end{center}

**Claim 7.7.** For all $\eta > 0$ the following holds for large enough $n$

$$\sum_{(a_0, a_1, a_2, a_3) \in B \setminus A} u_{a_0 a_1} u_{a_1 a_2} u_{a_2 a_3} u_{a_3 a_0} \leq \eta^2 n^4 p^4,$$

where $B = \bigcup_{i=0}^3 B_i$, and $B_i$ is the set of all $(a_0, a_1, a_2, a_3) \in C_4$ with

$$u_{a_0 a_1} u_{a_1 a_2} u_{a_2 a_3} u_{a_3 a_0} \leq p^{1-2\epsilon}.$$

**Proof.** Assume towards contradiction that

$$\sum_{(a_0, a_1, a_2, a_3) \in B \setminus A} u_{a_0 a_1} u_{a_1 a_2} u_{a_2 a_3} u_{a_3 a_0} > \eta^2 n^4 p^4.$$

We claim that for any $(a_0, a_1, a_2, a_3) \in B \setminus A$ there is $i$ satisfying

$$u_{a_0 a_1} u_{a_1 a_2} v_{a_2 a_3} u_{a_3 a_0} > \sqrt{p} / \log(1/p) \quad \text{and} \quad u_{a_0 a_1} u_{a_1 a_2} u_{a_2 a_3} u_{a_3 a_0} \leq p^{1-2\epsilon},$$

where summation is taken modulo 4. Indeed, let $(a_0, a_1, a_2, a_3) \in B \setminus A$. Since $(a_0, a_1, a_2, a_3) \notin A$ we have

$$u_{a_0 a_1} > \sqrt{p} / \log(1/p) \quad \text{or} \quad u_{a_0 a_1} > \sqrt{p} / \log(1/p)$$

and

$$u_{a_0 a_1} > \sqrt{p} / \log(1/p) \quad \text{or} \quad u_{a_0 a_1} > \sqrt{p} / \log(1/p).$$

Without loss of generality, assume that $u_{a_0 a_1} u_{a_1 a_2} > \sqrt{p} / \log(1/p)$. Note that as $(a_0, a_1, a_2, a_3) \in B$ there is $i$ such that $u_{a_0 a_1} u_{a_1 a_2} u_{a_2 a_3} u_{a_3 a_0} \leq p^{1-2\epsilon}$. For $\epsilon < 1/4$ we have $\sqrt{p} / \log(1/p) \geq p^{1-2\epsilon}$ implying that $i \neq 0, 1, 3$ and therefore, $i = 2$. We continue by letting $L_1$ be the set of all $\{x, y\} \in \binom{[n]}{2}$ such that $u_{xy} > \sqrt{p} / \log(1/p)$, and we claim that

$$\sum_{(a_0, a_1, a_2, a_3) \in B \setminus A} u_{a_0 a_1} u_{a_1 a_2} u_{a_2 a_3} u_{a_3 a_0} \leq np^{2(1-2\epsilon)} \left( \sum_{\{x, y\} \in L_1} u_{xy} \right)^2.$$ 

This follows from the fact that any $K_{1,2}$ in $K_n$ can be extended to a $C_4$ in at most $n$ ways. Therefore, we obtain the following:

$$\sum_{\{x, y\} \in L_1} u_{xy} > \eta n^{3/2} p^{1+2\epsilon}.\]
By Lemma 7.4 we have,
\[ \sum_{x,y} I_p(p + u_{xy}) \geq \sum_{\{x,y\} \in \mathcal{L}_1} I_p(p + u_{xy}) = \sum_{\{x,y\} \in \mathcal{L}_1} (1 + o(1))u_{xy} \log(u_{xy}/p) = \Omega(n^{3/2}p^{1+2\varepsilon} \log(1/p)). \]

This is a contradiction to our assumption that \( \sum_{x,y} I_p(p + u_{xy}) = O(n^2p^2 \log(1/p)) \) as \( p \ll n^{-1/2-\varepsilon} \) and therefore, \( n^{3/2}p^{1+2\varepsilon} = o(n^2p^2) \). \hfill \square

**Claim 7.8.** For any \( \eta > 0 \) there exists \( \gamma > 0 \) such that the following holds provided \( n \) is large enough
\[ \sum_{(a_0, a_1, a_2, a_3) \in D(\gamma)} u_{a_0a_1}u_{a_1a_2}u_{a_2a_3}u_{a_3a_0} \leq \eta^2n^4p^4, \]
where \( D(\gamma) = D_0(\gamma) \cup D_1(\gamma) \), and \( D_i(\gamma) \) is the set of all tuples \( (a_0, a_1, a_2, a_3) \in C_4 \) with
\[ u_{a_ia_{i+1}} \leq \gamma \quad \text{or} \quad u_{a_ia_{i+2}a_{i+3}} \leq \gamma, \]
and
\[ u_{a_ia_{i+1}a_{i+2}}, u_{a_ia_{i+3}} \geq p^{1-2\varepsilon}. \]

**Proof.** Let \( \gamma = (\eta \varepsilon / C)^2 \) and assume towards contradiction that
\[ \sum_{(a_0, a_1, a_2, a_3) \in D(\gamma)} u_{a_0a_1}u_{a_1a_2}u_{a_2a_3}u_{a_3a_0} > \eta^2n^4p^4. \]

Note that any \( (a_0, a_1, a_2, a_3) \in D(\gamma) \) satisfies
\[ u_{a_ia_{i+1}} \leq \gamma \quad \text{or} \quad u_{a_ia_{i+2}a_{i+3}} \leq \gamma \]
and
\[ u_{a_ia_{i+3}}, u_{a_{i+1}a_{i+2}} \geq p^{1-2\varepsilon}, \]
for some \( i \) and where summation is taken modulo 4. Letting \( \mathcal{L}_2 \) be the set of all \( \{x, y\} \in \binom{[n]}{2} \) such that \( u_{xy} \geq p^{1-2\varepsilon} \) we have the following by the definition of \( D(\gamma) \):
\[ \sum_{(a_0, a_1, a_2, a_3) \in D(\gamma)} u_{a_0a_1}u_{a_1a_2}u_{a_2a_3}u_{a_3a_0} \leq \gamma \left( \sum_{\{x, y\} \in \mathcal{L}_2} u_{xy} \right)^2. \]

This implies that
\[ \sum_{\{x, y\} \in \mathcal{L}_2} u_{xy} > \eta n^2p^2 / \sqrt{\gamma}. \]

Therefore, by the definition of \( \mathcal{L}_2 \) and by Lemma 7.4 we have,
\[ \sum_{x,y} I_p(p + u_{xy}) \geq \sum_{\{x,y\} \in \mathcal{L}_2} I_p(p + u_{xy}) \geq \sum_{\{x,y\} \in \mathcal{L}_2} u_{xy} \log(u_{xy}/p) \geq \frac{2\eta \varepsilon n^2p^2 \log(1/p)}{\sqrt{\gamma}}. \]
This is a contradiction to the choice of $\gamma$ as by our assumptions we have $\sum_{x,y} I_p(p + u_{xy}) \leq Cn^2 p^2 \log(1/p)$.

We now proceed with the proof of Proposition 7.1. By the our assumptions and the above three claims we obtain the following for any $\eta > 0$ and small enough $\gamma > 0$:

\[ \sum_{(a_0,a_1,a_2,a_3) \in \mathcal{C}_4 \setminus (A \cup B \cup D(\gamma))} u_{a_0,a_1} u_{a_1,a_2} u_{a_2,a_3} u_{a_3,a_0} (1 - u_{a_0,a_2})(1 - u_{a_1,a_3}) \geq (\delta - \eta) \mathbb{E}[X]. \]

We claim that provided $n$ is large enough we have:

\[ \mathcal{C}_4 \setminus (A \cup B \cup D(\gamma)) \subseteq \mathcal{C}_4'. \]

Indeed, as $(a_0, a_1, a_2, a_3) \notin A$, there exists $i$ such that provided $\varepsilon < 1/4$ and $n$ is large enough we have:

\[ u_{a_0,a_{i+1}} u_{a_{i+1}a_{i+2}} > \sqrt{p} / \log(1/p) > p^{1-2\varepsilon}. \]

Further, as $(a_0, a_1, a_2, a_3) \notin B_{i+2}$ we have

\[ u_{a_{i+2}a_{i+3}} > p^{1-2\varepsilon} \quad \text{or} \quad u_{a_{i}a_{i+3}} > p^{1-2\varepsilon}. \]

Without loss of generality, assume the latter holds. In particular we have,

\[ u_{a_{i}a_{i+3}}, u_{a_{i+1}a_{i+2}} > p^{1-2\varepsilon}. \]

Since $(a_0, a_1, a_2, a_3) \notin D_i(\gamma)$, we have

\[ u_{a_{i}a_{i+1}}, u_{a_{i+2}a_{i+3}} > \gamma. \]

Finally, as $(a_0, a_1, a_2, a_3) \notin D_{i+1}(\gamma)$, for large enough $n$ we have $p^{1-2\varepsilon} < \gamma$ implying that

\[ u_{a_{i}a_{i+3}}, u_{a_{i+2}a_{i+3}} > \gamma. \]

Letting $\mathcal{L}_3$ be the set of all \( \{x,y\} \in \left( \left\lfloor \frac{n}{2} \right\rfloor \right) \) with $u_{xy} > \gamma$, we claim the following holds:

\[ \sum_{(a_0,a_1,a_2,a_3) \in \mathcal{C}_4'} u_{a_0,a_1} u_{a_1,a_2} u_{a_2,a_3} u_{a_3,a_0} (1 - u_{a_0,a_2})(1 - u_{a_1,a_3}) \leq \left( \sum_{(a_0,a_1) \in \mathcal{L}_3} u_{a_0,a_1} \right)^2 / 4. \quad (26) \]

Indeed,

\[ 2 \sum_{(a_0,a_1,a_2,a_3) \in \mathcal{C}_4'} u_{a_0,a_1} u_{a_1,a_2} u_{a_2,a_3} u_{a_3,a_0} (1 - u_{a_0,a_2})(1 - u_{a_1,a_3}) \]

is at most

\[ \sum_{(a_0,a_1,a_2,a_3) \in \mathcal{C}_4'} u_{a_0,a_1} u_{a_2,a_3} (u_{a_1,a_2} u_{a_1,a_3} (1 - u_{a_0,a_2})(1 - u_{a_1,a_3}) + u_{a_0,a_2} u_{a_1,a_3} (1 - u_{a_0,a_2})(1 - u_{a_1,a_3})), \quad (\star) \]
where the sum ranges over unordered pairs, \(\{\{a_0, a_1\}, \{a_2, a_3\}\} \subseteq \mathcal{L}_3\), such that \(a_0, a_1, a_2, a_3\) are all distinct. We also have

\[
(\star) \leq (1 - u_{a_0a_2}) + u_{a_0a_2} = 1.
\]

This implies that

\[
(\delta - \eta)\mathbb{E}[X] \leq \sum_{(a_0, a_1, a_2, a_3) \in \mathcal{L}_3} u_{a_0a_1} u_{a_1a_2} u_{a_2a_3} u_{a_3a_0}(1 - u_{a_0a_2})(1 - u_{a_1a_3}) \leq \frac{\sum u_{a_0a_1} u_{a_1a_2}}{2},
\]

where the summation in the right-hand side is over unordered pairs, \(\{\{a_0, a_1\}, \{a_2, a_3\}\} \subseteq \mathcal{L}_3\), such that \(a_0, a_1, a_2, a_3\) are all distinct. This implies (26) as

\[
\frac{\left(\sum_{(a_0, a_1) \in \mathcal{L}_3} u_{a_0a_1}\right)^2}{2} = \frac{\sum_{(a_0, a_1) \in \mathcal{L}_3} u_{a_0a_1}^2 + 2 \sum u_{a_0a_1} u_{a_2a_3} }{2} \geq \sum u_{a_0a_1} u_{a_2a_3},
\]

where the unlabeled summations are over unordered pairs \(\{\{a_0, a_1\}, \{a_2, a_3\}\} \subseteq \mathcal{L}_3\). We conclude that:

\[
2\sqrt{\delta - \eta} \mathbb{E}[X] \leq \sum_{\{a_0, a_1\} \in \mathcal{L}_3} u_{a_0a_1}.
\]

Now we can bound from below the ‘cost’ using Lemma 7.4 as follows:

\[
\sum_{\{x, y\} \in \mathcal{L}_3} I_{p}(p + u_{xy}) \geq \sum_{\{x, y\} \in \mathcal{L}_3} I_{p}(p + u_{xy}) = (1 + o(1)) \sum_{\{x, y\} \in \mathcal{L}_3} u_{xy} \log(u_{xy}/p) \\
\geq (1 + o(1)) \sum_{\{x, y\} \in \mathcal{L}_3} u_{xy} \log(1/p) \geq (1 + o(1)) 2\sqrt{\delta - \eta} \mathbb{E}[X] \log(1/p).
\]

This finishes the proof of the proposition as \(\mathbb{E}[X] = (1 + o(1)) \frac{\delta^2 p^2}{8}\).

ACKNOWLEDGMENTS

The author thanks his advisor Wojciech Samotij for his guidance through the write-up of this paper, as well as helpful and fruitful discussions. The author also thanks Arnon Chor and Dor Elboim for helpful discussions. Lastly, the author thanks Eden Kuperwasser for fruitful discussions and for creating the figure in this paper. This research was supported in part by the Israel Science Foundation grants 1145/18 and 2110/22.

REFERENCES

1. N. Alon, On the number of subgraphs of prescribed type of graphs with a given number of edges, Israel J. Math. 38 (1981), no. 1-2, 116–130.
2. F. Augeri, Nonlinear large deviation bounds with applications to Wigner matrices and sparse Erdős-Rényi graphs, Ann. Probab. 48 (2020), no. 5, 2404–2448.
3. F. Augeri, A transportation approach to the mean-field approximation, Probab. Theory Relat. Fields 180 (2021), no. 1, 1–32.
4. A. Basak and R. Basu. *Upper tail large deviations of regular subgraph counts in Erdős–Rényi graphs in the full localized regime*, arXiv preprint arXiv:1912.11410, 2019.

5. B. B. Bhattacharya, S. Ganguly, E. Lubetzky, and Y. Zhao, *Upper tails and independence polynomials in random graphs*, Adv. Math. 319 (2017), 313–347.

6. B. Bollobás, “Random graphs,” Combinatorics (Swansea, 1981), London Math. Soc. Lecture Note Ser, Vol 52, Cambridge Univ. Press, Cambridge-New York, 1981, pp. 80–102.

7. B. Bollobás, C. Nara, and S.-i. Tachibana, *The maximal number of induced complete bipartite graphs*, Discret. Math. 62 (1986), no. 3, 271–275.

8. S. Chatterjee, *The missing log in large deviations for triangle counts*, Random Struct. Algoritm. 40 (2012), no. 4, 437–451.

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

10. S. Chatterjee and S. R. Srinivas Varadhan, *The large deviation principle for the Erdős–Rényi random graph*, Eur. J. Comb. 32 (2011), no. 7, 1000–1017.

11. N. Cook and A. Dembo, *Large deviations of subgraph counts for sparse Erdős–Rényi graphs*, Adv. Math. 373 (2020), 107289.

12. N. A. Cook, A. Dembo, and H. T. Pham, *Regularity method and large deviation principles for the Erdős–Rényi hypergraph*, arXiv preprint arXiv:2102.09100, 2021.

13. R. DeMarco and J. Kahn, *Tight upper tail bounds for cliques*, Random Struct. Algoritm. 41 (2012), no. 4, 469–487.

14. R. DeMarco and J. Kahn, *Upper tails for triangles*, Random Struct. Algoritm. 40 (2012), no. 4, 452–459.

15. A. Dembo and O. Zeitouni, “Large deviations techniques and applications,” Stochastic modelling and applied probability, Corrected reprint of the second (1998) ed., Vol 38, Springer-Verlag, Berlin, 2010.

16. R. Eldan, *Gaussian-width gradient complexity, reverse log-Sobolev inequalities and nonlinear large deviations*, Geom. Funct. Anal. 28 (2018), no. 6, 1548–1596.

17. P. Erdős, *On the number of complete subgraphs contained in certain graphs*, Magyar Tud Akad. Mat. Kutató Int. Közl 7 (1962), no. 3, 459–464.

18. E. Friedgut and J. Kahn, *On the number of copies of one hypergraph in another*, Israel J. Math. 105 (1998), 251–256.

19. D. Gerbner, D. T. Nagy, B. Patkós, and M. Vizer, *On the maximum number of copies of H in graphs with given size and order*, J. Graph Theory 96 (2021), no. 1, 34–43.

20. B. Gunby. *Upper tails of subgraph counts in sparse regular graphs*, arXiv preprint arXiv:2010.00658, 2020.

21. M. Harel, F. Mousset, and W. Samotij, *Upper tails via high moments and entropic stability*, arXiv:1904.08212, 2019.

22. S. Janson, K. Oleszkiewicz, and A. Ruciński, *Upper tails for subgraph counts in random graphs*, Israel J. Math. 142 (2004), 61–92.

23. J. H. Kim and V. H. Vu, *Divide and conquer martingales and the number of triangles in a random graph*, Random Struct. Algoritm. 24 (2004), no. 2, 166–174.

24. Y. P. Liu and Y. Zhao, *On the upper tail problem for random hypergraphs*, Random Struct. Algoritm. 58 (2021), no. 2, 179–220.

25. E. Lubetzky and Y. Zhao, *On replica symmetry of large deviations in random graphs*, Random Struct. Algoritm. 47 (2015), no. 1, 109–146.

26. E. Lubetzky and Y. Zhao, *On the variational problem for upper tails in sparse random graphs*, Random Struct. Algoritm. 50 (2017), no. 3, 420–436.

27. A. Ruciński, *When are small subgraphs of a random graph normally distributed?*, Probab. Theory Relat. Fields 78 (1988), no. 1, 1–10.

28. V. H. Vu, *A large deviation result on the number of small subgraphs of a random graph*, Combin Probab. Comput. 10 (2001), no. 1, 79–94.

How to cite this article: A. Cohen Antonir, *The upper tail problem for induced 4-cycles in sparse random graphs*, Random Struct. Alg. 64 (2024), 401−459. https://doi.org/10.1002/rsa.21187
APPENDIX A: TRANSLATING GRAPHONS’ LANGUAGE INTO GRAPHS’ LANGUAGE

The aim of this Appendix is to add a proof (in the language of graphs instead of graphons) for Lemmas 7.3 and 7.5 which were proven by Lubetzky–Zhao [27] and Bhattacharya, Ganguly, Lubetzky and Zhao [5] in the language of graphons.

We start with a discussion of how one would prove Lemma 7.3, then we state some lemmas and an extension of Lemma 7.5. To prove Lemma 7.3 we will use Lemma 7.5, and therefore it will be proven before Lemma 7.3. Recall Lemma 7.3:

**Lemma A.1.** Suppose \( \varepsilon, \delta \) are positive reals with \( \varepsilon \) sufficiently small. Suppose also that \( \sqrt{\log(n)}/n \ll p \ll n^{-1/2} \). Then, for any sequence \( q \) satisfying:

1. \( q_i = p + u_i \) for some \( u_i \geq 0 \),
2. \( \sum_{i \in \{1, \ldots, n\}} I_p(q_i) \leq \sqrt{2n^2p^3 \log(1/p)} \), and
3. \( \mathbb{E}[N_{\text{ind}}(C_4, G_{n,q})] \geq (1 + \delta - \varepsilon)\mathbb{E}[X] \),

we have \( \mathbb{E}[N_{\text{ind}}(C_4, G_{n,q})] \geq (\delta - 2 \varepsilon)\mathbb{E}[X] \) provided \( n \) is large enough.

We start with by analysing the term \( \mathbb{E}[N_{\text{ind}}(C_4, G_{n,q})] \). Suppose \( q \in [0,1]\{\} \) satisfies the conditions of Lemma 7.3. Note that

\[
\mathbb{E}[N_{\text{ind}}(C_4, G_{n,q})] = \sum_{(x,y,z,w) \in C_4} q_{xy}q_{yz}q_{zw}q_{wx}(1-q_{xz})(1-q_{yw}).
\]

Since \( q_i = p + u_i \geq u_i \) we have the following for large enough \( n \):

\[
\mathbb{E}[N_{\text{ind}}(C_4, G_{n,q})] \leq \sum_{(x,y,z,w) \in C_4} q_{xy}q_{yz}q_{zw}q_{wx}(1-u_{xz})(1-u_{yw})
\]

\[
= \sum_{(x,y,z,w) \in C_4} (p + u_{xy})(p + u_{yz})(p + u_{zw})(p + u_{wx})(1-u_{xz})(1-u_{yw})
\]

\[
= \mathbb{E}[N(C_4, G_{n,p})] + \sum_{H \leq C_4} \mathbb{E}[N(H, G_{n,\vec{q}})]p^{4-\varepsilon_H}
\]

\[
= \mathbb{E}[X]/(1-p)^2 + \sum_{H \leq C_4} \mathbb{E}[N(H, G_{n,\vec{q}})]p^{4-\varepsilon_H}
\]

\[
= (1 + \varepsilon/2)\mathbb{E}[X] + \sum_{H \leq C_4} \mathbb{E}[N(H, G_{n,\vec{q}})]p^{4-\varepsilon_H},
\]

where \( C \) stands for a spanning subgraph which is not equal to the host graph. Recalling that \( \mathbb{E}[N_{\text{ind}}(C_4, G_{n,q})] - \mathbb{E}[X] \geq (\delta - \varepsilon)\mathbb{E}[X] \), we obtain,

\[
(\delta - 3\varepsilon/2)\mathbb{E}[X] \leq \mathbb{E}[N_{\text{ind}}(C_4, G_{n,q})] + \mathbb{E}[N(P_4, G_{n,a})]p + \mathbb{E}[N(M_2, G_{n,a})]p^2 + \mathbb{E}[N(K_{1,2} \sqcup K_1, G_{n,a})]p^2 + \mathbb{E}[N(K_2 \sqcup 2K_1, G_{n,a})]p^3.
\]

\[
\leq \mathbb{E}[N_{\text{ind}}(C_4, G_{n,a})] + \mathbb{E}[N(P_4, G_{n,a})]p + \mathbb{E}[N(M_2, G_{n,a})]p^2 + \mathbb{E}[N(K_{1,2}, G_{n,a})]np^2 + \mathbb{E}[N(K_2, G_{n,a})]n^2p^3,
\]

where \( P_4 \) is the path with three edges, \( M_2 \) is the matching of size two, \( K_{1,2} \sqcup K_1 \) is the complete bipartite graph with sides of size 1 and 2 and an extra isolated vertex and \( K_2 \sqcup 2K_1 \) is the disjoint union of an
edge and an independent set of size two. Therefore, to prove the lemma, it is enough to show that all of the terms in the right-hand side of the above inequality except for $\mathbb{E}[N_{\text{ind}}(C_4, G_{n, \overline{b}})]$ are negligible in comparison to $n^4 p^4$.

For this we remind the reader Lemma 7.4 and cite two useful lemmas from [27] and [5]:

**Lemma A.2** ([27]). The following holds,

$$I_p(p + x) = \begin{cases} (1 + o(1))x^2/p, & 0 \leq x \ll p, \\ (1 + o(1))x \log(x/p), & p \ll x \leq 1 - p. \end{cases}$$

**Lemma A.3** ([27]). There exists $p_0 > 0$ such that for all $0 < p \leq p_0$ and $0 \leq x \leq b \leq 1 - p - 1/\log(1/p)$ we have,

$$I_p(p + x) \geq \frac{x^2I_p(p + b)}{b^2}.$$  

**Corollary A.4** ([27]). There exists $p_0 > 0$ such that for all $0 < p \leq p_0$ and all $0 \leq x \leq 1 - p$ we have,

$$I_p(p + x) \geq x^2I_p(1 - 1/\log(1/p)) = (1 + o(1))x^2I_p(1)$$

where the $o(1)$-term goes to zero as $p \to 0$.

Another useful tool we use is the generalized Hölder inequality:

**Lemma A.5** (Generalized Hölder inequality [5]). Let $\mu_1, \mu_2, \ldots, \mu_n$ be probability measures on $\Omega_1, \Omega_2, \ldots, \Omega_n$ respectively, and let $\mu = \prod_{i=1}^n \mu_i$. Let $A_1, A_2, \ldots, A_m$ be non-empty subsets of $[n]$ and for $A \subseteq [n]$ put $\mu_A = \prod_{j \in A} \mu_j$ and $\Omega_A = \prod_{j \in A} \Omega_j$. Let $f_i \in L^p(\Omega_{A_i}, \mu_{A_i})$ for each $i \in [m]$, and suppose that $\sum_{i:j \neq j} \frac{1}{p_i} \leq 1$ for all $j \in [n]$. Then,

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

In particular, if every element of $[n]$ is contained in at most two sets $A_j$, then we can take all $p_i = 2$ and obtain:

$$\int \prod_{i=1}^m |f_i| d\mu \leq \prod_{i=1}^m \left( \int |f_i|^2 d\mu_{A_i} \right)^{1/2}.$$  

Next, we recreate a proof of Lubetzky–Zhao [27] and Bhattacharya, Ganguly, Lubetzky and Zhao [5] of an extension of Lemma 7.5. Then, we will derive the required bounds.

**Lemma A.6.** Suppose $\epsilon, \delta, C$ are positive reals and $\sqrt{\log(n)/n} \ll p \ll n^{-1/2}$. Suppose also that $\tilde{u} \in [0, 1 - p]^{(\epsilon)}$ such that $\sum_{i \in [n]} I_p(p + u_i) \leq Cn^2p^2 \log(1/p)$. Let $b = b(n)$ be such that $\max\{np^2, \sqrt{p \log(1/p)}\} \ll b \leq 1 - \epsilon$. Then, provided $n$ is large enough there is a constant $D > 0$ such that the following holds:

1. $\sum_{i \in [n]} \left( \sum_{j \neq i} u_{ij} \right)^2 \leq Dn^3p^2b$,
2. $\sum_{i \in [n]} u_i \leq Dn^2p^{3/2} \sqrt{\log(1/p)}$, and
(iii) \( \sum_{i \in \binom{[n]}{2}} u_i^2 \leq Dn^2p^2. \)

**Proof.** We first show that the set \( B = \{ x \in [n] : \sum_{y \neq x} u_{xy} \geq bn \} \) is empty, provided \( n \) is large enough. If this were not true, then, as \( I_p(p + x) \) is a convex function of \( x \), we have the following by Jensen’s inequality

\[
\sum_{x,y} I_p(p + u_{xy}) \geq (n - 1) \sum_x I_p \left( p + \frac{\sum_{y \neq x} u_{xy}}{n - 1} \right) \geq (n - 1) \sum_{x \in B} I_p \left( p + \frac{\sum_{y \neq x} u_{xy}}{n - 1} \right)
\]

\[
\geq (n - 1) \sum_{x \in B} I_p(p + b) \geq (1 + o(1))(n - 1)|B|b \log(b/p)
\]

where the last inequality follows from Lemma 7.4. Since \( b \gg \sqrt{p \log(1/p)} \) and \( \sum_{i \in \binom{[n]}{2}} I_p(p + u_i) \leq Cp^2n^2 \log(1/p) \) we obtain the following:

\[
|B| \leq (1 + o(1)) \frac{Cp^2n^2 \log(1/p)}{(n - 1)b \log(\sqrt{\log(1/p)})} \ll 1,
\]

and therefore, \( B \) is empty. To prove (i), we use the convexity of \( I_p(p + x) \) and Lemma A.3,

\[
\sum_{x,y} I_p(p + u_{xy}) \geq (n - 1) \sum_x I_p \left( p + \frac{\sum_{y \neq x} u_{xy}}{n - 1} \right) = (n - 1) \sum_{x \notin B} I_p \left( p + \frac{\sum_{y \neq x} u_{xy}}{n - 1} \right)
\]

\[
\geq (n - 1)I_p(p + b)\sum_{x \notin B} \left( \frac{\sum_{y \neq x} u_{xy}}{b^2 n^2} \right)^2.
\]

Since, \( \sum_{i \in \binom{[n]}{2}} I_p(p + u_i) \leq Cp^2n^2 \log(1/p) \) we obtain:

\[
\sum_x \left( \sum_{y \neq x} u_{xy} \right)^2 \leq \frac{Cp^2n^4 b^2 \log(1/p)}{(n - 1)I_p(p + b)}
\]

As \( b \gg \sqrt{p \log(1/p)} \gg p \) we may use Lemma 7.4 and we obtain:

\[
\sum_{x \in [n]} \left( \sum_{y \neq x} u_{xy} \right)^2 \leq \frac{Cn^4 p^2 b^2 \log(1/p)}{(1 + o(1))(n - 1)b \log(b/p)} \leq 2Cn^3 p^2 b.
\]

Now we prove (ii). By convexity we have,

\[
\binom{n}{2} I_p \left( p + \frac{\sum_{i \in \binom{[n]}{2}} u_i}{\binom{n}{2}} \right) \leq \sum_{x,y} I_p(p + u_{xy}) \leq Cn^2 p^2 \log(1/p).
\]
Since \( p \gg \sqrt{p^3 \log(1/p)} \) we may use Lemma 7.4, and obtain the following for large enough \( n \):

\[
I_p\left(p + \sqrt{12 Cp^3 \log(1/p)}\right) \geq \frac{12Cp^3 \log(1/p)}{3p} = 4Cp^2 \log(1/p).
\]

This implies the following provided \( n \) is large enough,

\[
I_p\left(p + \frac{\sum_{i \in \binom{n}{2}} u_i}{\binom{n}{2}}\right) \leq 4Cp^2 \log(1/p) \leq I_p\left(p + \sqrt{12 Cp^3 \log(1/p)}\right).
\]

By the monotonicity of \( I_p(p + x) \) for \( x > 0 \) we obtain (ii). That is

\[
\sum_{i \in \binom{n}{2}} u_i \leq \sqrt{3Cn^4 p^3 \log(1/p)}.
\]

Lastly, we prove (iii). By Corollary A.4,

\[
\sum_{i \in \binom{n}{2}} I_p(p + u_i) \geq (1 + o(1)) \sum_{i \in \binom{n}{2}} u_i^2 \log(1/p).
\]

This and the assumption of the lemma implies (iii).

Now we can finish the proof of Lemma 7.3.

**Lemma A.7.** Suppose \( \delta, C \) are positive reals and \( \sqrt{\log(n)/n} \ll p \ll n^{-1/2} \). Suppose also that \( \bar{u} \) is a sequence of \( \binom{n}{2} \) reals between 0 and 1 such that \( \sum_{i \in \binom{n}{2}} I_p(p + u_i) \leq Cp^2 n^2 \log(1/p) \). Then,

\[
\mathbb{E}[N(M_2, G_{n, \bar{u}})]p^2 = o(n^4 p^4),
\]

\[
\mathbb{E}[N(K_2, G_{n, \bar{u}})]n^2 p^3 = o(n^4 p^4),
\]

\[
\mathbb{E}[N(P_4, G_{n, \bar{u}})]p = o(n^4 p^4),
\]

\[
\mathbb{E}[N(K_{1,2}, G_{n, \bar{u}})]np^2 = o(n^4 p^4).
\]

**Proof.** By item (ii) in Lemma A.6 we have \( \mathbb{E}[e(G_{n, \bar{u}})] = o(n^2 p^2) \), and therefore,

\[
\mathbb{E}[N_{\text{ind}}(M_2, G_{n, \bar{u}})]p^2 \leq \mathbb{E}[N(M_2, G_{n, \bar{u}})]p^2 \leq \mathbb{E}[e(G_{n, \bar{u}})]p^2 = o(n^4 p^4),
\]

\[
\mathbb{E}[N_{\text{ind}}(K_2, G_{n, \bar{u}})]n^2 p^3 = \mathbb{E}[e(G_{n, \bar{u}})]n^2 p^3 = o(n^4 p^4),
\]

where \( N(M_2, G_{n, \bar{u}}) \) is the number of \( M_2 \) in \( G_{n, \bar{u}} \). Let \( \max \{np^2, \sqrt{p \log(1/p)}\} \ll b = b(n) \ll 1 \). Note that each \( P_4 \subseteq K_n \) with vertices \( a_0, a_1, a_2, a_3 \) can be represented in exactly two
ways as a tuple \((a_0, a_1, a_2, a_3)\) where edges are consecutive vertices in this tuple. Let \(P_4\) be a collection of exactly one such representative for each \(P_4\) in \(K_n\). Observe the following:

\[
\mathbb{E}[N(P_4, G_{n,\bar{u}})] = \sum_{(x,y,z,w)\in P_4} u_{xy} u_{yz} u_{zw} \leq \sum_{y,z,w\in[n]} \left( \sum_{x\neq y} u_{xy} \right) u_{yz} u_{zw},
\]

where \(N(P_4, G_{n,\bar{u}})\) is the number of \(P_4\) in \(G_{n,\bar{u}}\). By the generalized Hölder inequality we have:

\[
\sum_{y,z,w\in[n]} \left( \sum_{x\neq y} u_{xy} \right) u_{yz} u_{zw} \leq \left( \sum_{y} \mathbb{E}[\text{deg}(y)]^2 \right)^{1/2} \left( \sum_{x,y} u_{xy}^2 \right).
\]

Applying items (i) and (iii) in Lemma A.6 we obtain

\[
\mathbb{E}[N(P_4, G_{n,\bar{u}})] \leq D^{3/2} n^{7/2} p^4 \sqrt{b} = o(n^4 p^4).
\]

This establishes the third assertion of the lemma. The fourth assertion of the lemma follows immediately from Lemma A.6 item (i) and the definition of \(b\):

\[
\mathbb{E}[N(K_{1,2}, G_{n,\bar{u}})] \leq \sum_{x,y,z\in[n]} u_{xy} u_{yz} \leq \sum_{y\in[n]} \left( \sum_{x\in[n]} u_{xy} \right)^2 \leq D n^2 p^{3/2} b = o(n^3 p^2).
\]