UPPER TAIL BEHAVIOR OF THE NUMBER OF TRIANGLES IN RANDOM GRAPHS WITH CONSTANT AVERAGE DEGREE

SHIRSHENDU GANGULY, ELLA HIESMAYR, AND KYEONGSIK NAM

Abstract. Let \( N \) be the number of triangles in an Erdős–Rényi graph \( G(n, p) \) on \( n \) vertices with edge density \( p = d/n \), where \( d > 0 \) is a fixed constant. It is well known that \( N \) weakly converges to the Poisson distribution with mean \( d^3/6 \) as \( n \to \infty \). We address the upper tail problem for \( N \), namely, we investigate how fast \( k \) must grow, so that \( P(N \geq k) \) is not well approximated anymore by the tail of the corresponding Poisson variable. Proving that the tail exhibits a sharp phase transition, we essentially show that the upper tail is governed by Poisson behavior only when \( k^{1/3} \log k < \left( \frac{3}{2\sqrt{2}} \right)^{2/3} \log n \) (sub-critical regime) as well as pin down the tail behavior when \( k^{1/3} \log k > \left( \frac{3}{2\sqrt{2}} \right)^{2/3} \log n \) (super-critical regime). We further prove a structure theorem, showing that the sub-critical upper tail behavior is dictated by the appearance of almost \( k \) vertex-disjoint triangles whereas in the supercritical regime, the excess triangles arise from a clique like structure of size approximately \( (6k)^{1/3} \). This settles the long-standing upper-tail problem in this case, answering a question of Aldous, complementing a long sequence of works, spanning multiple decades and culminating in [19], which analyzed the problem only in the regime \( p \gg \frac{1}{n} \). The proofs rely on several novel graph theoretical results which could have other applications.

Contents

1. Introduction and main results 1
2. Lower bound 7
3. Graph theoretic results 10
4. Rarity of triangle induced graphs 17
5. Proof of Theorem 1.1 via graph decomposition into TISGs. 20
6. Proof of Theorem 1.2 24
References 30

1. Introduction and main results

The statistical properties of the number of triangles or other local structures in random graphs have been a major topic of study for the last few decades witnessing several important advances. Postponing a somewhat detailed overview of the literature to Section 1.2, let us first move towards stating the basic set up and the main results of this article.

Throughout the paper, let \( G_n = G(n, \frac{d}{n}) \) be the Erdős–Rényi graph on \( n \) vertices with edge density \( d/n \) where \( d > 0 \) is a fixed constant. We will denote by \( V_n := V(G_n) \) and \( E_n := E(G_n) \), the corresponding vertex and edge sets respectively (note that \( |V_n| = n \)). Further, let \( N \) be the number of triangles in \( G_n \), i.e., the number of unordered triples \((u, v, w)\) of vertices \( u, v, w \in V_n \), where all the possible edges \((u, v), (v, w), (u, w)\) are elements of \( E_n \). It is known [8] that \( N \) weakly converges to the Poisson distribution with mean \( \frac{d^3}{6} \), to be denoted \( \text{Poi}(\frac{d^3}{6}) \). This implies that, for
any bounded sequence \( \{k_n\}_{n \in \mathbb{N}} \),

\[
\left| \mathbb{P}(N \geq k_n) - \mathbb{P}\left( \text{Poi}\left( \frac{d^3}{6} \right) \geq k_n \right) \right| \to 0
\tag{1}
\]
as \( n \to \infty \), where throughout the paper, \( \mathbb{P} \) and \( \mathbb{E} \) will denote the underlying probability and the corresponding expectation.

The purpose of the article is to investigate at what depth into the tail, the Poisson behavior no longer holds, i.e., how fast \( \{k_n\}_{n \in \mathbb{N}} \) must grow such that no version of (1) is true, and further, how does \( G_n \) look like, conditional on \( N \geq k \).

1.1. Main results. Throughout the paper, we assume that \( k_n \) is increasing in \( n \) and for the sake of simplicity, we will suppress the \( n \) dependence and just use \( k \).

Theorem 1.1. The following describes the “upper-tail” probabilities of \( N \).

1. There exists a constant \( c > 0 \) depending only on \( d \) such that the following holds. Let \( \delta > 0 \) be a sufficiently small constant. Suppose that \( k^{1/3} \log k < \left( \frac{3}{\sqrt{2}} \right)^{2/3} - \delta \log n \). Then, for sufficiently large \( n \),

\[
e^{-k \log k - ck} \leq \mathbb{P}(N \geq k) \leq e^{-k \log k + ck}.
\tag{2}
\]

2. Let \( \delta > 0 \) and \( 0 < \mu < \frac{1}{10} \) be constants. Suppose that \( k^{1/3} \log k > \left( \frac{3}{\sqrt{2}} \right)^{2/3} + \delta \log n \) and \( k \leq n^{\frac{4}{3} \mu} - \mu \). Then, for any \( \varepsilon > 0 \) and sufficiently large \( n \),

\[
n^{-\left( \frac{\delta}{\sqrt{2}} \right)^{2/3} + \varepsilon} k^{2/3} \leq \mathbb{P}(N \geq k) \leq n^{-\left( \frac{\delta}{\sqrt{2}} \right)^{2/3} - \varepsilon} k^{2/3}.
\tag{3}
\]

Thus Theorem 1.1 establishes a tight bound on the upper tail behavior of \( N \) and provides a precise critical location at which the same changes. It shows that the Poisson tail holds when \( k^{1/3} \log k \leq \left( \frac{3}{\sqrt{2}} \right)^{2/3} \log n \) (to be called the sub-critical regime from now on) and not beyond, the super-critical regime.

The next result answers a question of David Aldous, who had asked about the structure of the graph \( G_n \), conditioned on \( N \geq k \). We start with the following notation. For a graph \( G = (V, E) \), let \( \Delta(G) \) (simply \( \Delta \) when \( G \) is clear) be the number of triangles in \( G \) and for any subset \( W \subseteq V \), define \( \Delta_G(W) \) to be the number of triangles in \( G \) consisting of vertices in \( W \). If it is clear what the underlying graph \( G \) is, then we suppress the \( G \) dependence in \( \Delta_G(W) \) and just use the notation \( \Delta(W) \).

Theorem 1.2 (Structure theorem). Let \( 0 < \mu < \frac{1}{10} \) be fixed and \( k \leq n^{\frac{4}{3} \mu} - \mu \). Now for any constant \( 0 < \varepsilon < 1 \), define the following two events

\[
D_\varepsilon := \{ \text{There exist at least } (1 - \varepsilon)k \text{ vertex-disjoint triangles} \}
\]
and

\[
C_\varepsilon := \{ \text{There exists } V' \subseteq V_n \text{ such that } |V'| \leq (1 + \varepsilon)^{6^{1/3} k^{1/3}}, \Delta(V') \geq (1 - \varepsilon)k \}.
\]

In words, \( C_\varepsilon \) denotes the presence of an “almost” clique accounting for most triangles.

Then,

\[
\lim_{n \to \infty} \mathbb{P}(D_\varepsilon \cup C_\varepsilon | N \geq k) = 1.
\tag{4}
\]
This, in conjunction with Theorem 1.1, implies the following structure theorem. Let $\delta > 0$ be a sufficiently small constant. If $k^{1/3} \log k < \left( \frac{3}{\sqrt{2}} \right)^{2/3} - \delta \log n$, then
\[
\lim_{n \to \infty} \mathbb{P}(D_\varepsilon | N \geq k) = 1,
\] and if $k^{1/3} \log k > \left( \frac{3}{\sqrt{2}} \right)^{2/3} + \delta \log n$, then
\[
\lim_{n \to \infty} \mathbb{P}(C_\varepsilon | N \geq k) = 1.
\]

Thus, the above states that in the sub-critical regime, it is likely that, conditional on $\{N \geq k\}$, one sees the appearance of almost $k$ disjoint triangles, while beyond that, an almost clique structure accounts for most of the excess triangles.

Remark 1.3. A few remarks are in order at this point. Note that the results for the super-critical regime in the above theorems essentially assume $k \leq n^{1/10}$, whereas it should ideally work for $k = O(n^3)$. This technical condition is an artifact of our proof and allows certain union bound arguments to work. While it is possible to push this with a more efficient union bound scheme, for expository purposes, we made no such attempt, considering the central problem we were aiming to solve was to find the threshold until which the Poisson statistics remain valid.

However, during the writing of the paper, a preprint [10] was posted on the arXiv, which proves (see Theorem 1.2 therein), among other things, one half of our Theorem 1.1 for a slightly different range of $k$. In particular, it is shown that the supercritical behavior, i.e., (3), holds for $k$ growing faster than $(\log n)^3$ (see [10, Remark 1.3] for details). Note that the critical location in Theorem 1.1 is given by $k^{1/3} \log k \approx \left( \frac{3}{\sqrt{2}} \right)^{2/3} \log n$ which implies $k \approx (\frac{\log n}{\log \log n})^3 \ll (\log n)^3$, and so the result in [10] falls short of going down to the critical location and further does not provide any insight into the subcritical behavior. However, on the other hand, unlike our technical condition, which is essentially $k \leq n^{1/10}$, there is no restriction on the upper bound for $k$ in their results. It is worth mentioning that the methods in [10] are rather different from ours and rely more on refining the approach in [19].

Summarizing, while our result covers the sub-critical region and the super-critical region up to $k \leq n^{1/10}$, the result in [10] covers the super-critical region from $(\log n)^3$ onwards.

Thus, the two results together settle the long-standing “upper tail problem” for the number of triangles in the case where the average degree is a constant.

Before discussing the key proof ideas, we now provide a broad overview of the advances made on the upper tail problem in various settings over the years.

1.2. Previous work. While the present paper considers the Erdős-Rényi random graph with constant average degree, the problem was initially considered for a general density $p$, i.e., for the random graph $G(n, p)$. While, eventually, counts for various subgraphs were studied quite closely, the initial works in this direction tried to pin down the tail probabilities for $N$, the triangle count, as in the present paper. Thus, formally, the upper tail problem for $N$ asks to estimate the large deviation rate function given by
\[
R(n, p, \delta) := -\log \mathbb{P}(N \geq (1 + \delta)\mathbb{E}[N]) \quad \text{for fixed } \delta > 0.
\]

This simple to state problem turned out to be fundamental and extremely challenging, leading to intense research for over two decades, (see e.g., [22, 29, 25, 21, 23, 11, 17, 18] and [9, 20] and the
\[
1\approx \text{will be informally used throughout the article to denote ‘close to’ in a sense whose meaning might change across locations depending on the context. We will refrain from being more precise.}
references therein). It followed from the works [29],[25] that \(^2\)
\[ n^2p^2 \lesssim R(n,p,\delta) \lesssim n^2p^2\log(1/p) \]
(the harder direction, the lower bound, relied on the “polynomial concentration” machinery, whereas the upper bound is obtained by planting a clique, and observing that a set of \(s \sim \delta^{1/3}np\) vertices can form a clique with probability \(p^s = p^{O(n^2p^2)}\), thus contributing roughly \(\delta^{(n/3)p^3}\) extra triangles). In parallel work, Chatterjee [11], and DeMarco and Kahn [18] settled the sharp order of the rate function showing
\[ R(n,p,\delta) \asymp n^2p^2\log(1/p) \quad \text{for } p \geq \frac{\log n}{n}. \]  
(7)

However, the methods in [11, 18, 17] were not strong enough to recover the exact asymptotics of this rate function leading one to wonder if planting a clique is always the best strategy.

Progress in this front has subsequently witnessed an explosion kickstarted by the pioneering work of Chatterjee and Varadhan [15] that introduced a large deviation framework for \(\mathcal{G}_{n,p}\) in the dense regime \((0 < p < 1\) fixed) via the theory of graphons (see the survey by Chatterjee [12]). However, the above framework is not equipped to handle the sparse regime \((p \rightarrow 0)\), where the understanding still remained rather limited until another breakthrough by Chatterjee and Dembo [13]. This reduced it to a natural variational problem in a certain range of \(p\). A significant amount of activity was then devoted to, at least asymptotically, solving the variational problem. This was initiated in [27] (who also considered the case when the triangle was replaced by a general clique), thereby yielding the following conclusion: for fixed \(\delta > 0\), if \(n^{-1/42}\log n \leq p = o(1)\), then
\[ R(n,p,\delta) \sim I(\delta)n^2p^2\log(1/p) \quad \text{where } I(\delta) = \min\{\frac{1}{2}\delta^{2/3}, \frac{1}{3}\delta\}. \]  
(8)

Thus the aforementioned clique construction gives the correct leading order constant if \(\delta \geq 27/8\). This was subsequently extended to general subgraphs beyond the case of the clique in [7]. However, as was indicated in (7) proved in [11, 18]. (8) is expected to hold until the much lower threshold of \(p \gg \frac{\log n}{n}\). Several important attempts were made subsequently in this direction. [5, 16] pushed the result down to \(p \gg \frac{1}{\sqrt{n}}\) where the problem undergoes a natural transition. More recently, in a breakthrough paper [19], using a different approach going back to the classical work of Janson, Oleszkiewicz and Ruciński [21], the upper tail problem was essentially completely solved in the regime \(p \gg \frac{1}{n}\). This includes in particular the window \(\frac{1}{n} \ll p \ll \frac{\log n}{n}\) beyond the ‘mean-field’ regime covered in (7). It is known that in the latter regime, the triangle count \(N\) behaves as a Poisson variable. Further, it is straightforward to verify that for a Poisson variable \(X\) with mean \(\mu\),
\[ -\log[\mathbb{P}(X \geq (1 + \delta)\mu)] = (1 + o(1))\left((1 + \delta)\log(1 + \delta) - \delta\right)\mu \]
as \(\mu \rightarrow \infty\). Now, when \(\frac{1}{n} \ll p \ll \frac{\log n}{n}\), the mean triangle count \(\mathbb{E}(N)\) does diverge to infinity as \(n \rightarrow \infty\) falling in the above setup and it was indeed shown in [19] that in this case,
\[ -\log[\mathbb{P}(N \geq (1 + \delta)\mathbb{E}(N))] = (1 + o(1))\left((1 + \delta)\log(1 + \delta) - \delta\right)\mathbb{E}(N), \]
(for a more detailed statement and related results, the reader is encouraged to refer to [19, Section 8]).

However, none of the methods described above are equipped to handle the case of constant average degree, where \(N\) behaves like a Poisson random variable with bounded mean, the object of study in the present paper.

\(^2\)We write \(f \lesssim g\) to denote \(f = O(g); f \asymp g\) means \(f = \Theta(g); f \sim g\) means \(f = (1+o(1))g\) and \(f \ll g\) means \(f = o(g)\).
Finally, it might be worth pointing out that while the previous works were only looking at the large deviation regime, i.e., deviation from the mean by a constant multiplicative factor, a regime which is not interesting when the mean is bounded, we focus on the “entire” tail of the random variable $N$.

1.3. Idea of proof. In this section, we highlight some of the key ideas going into our arguments. Broadly, there are two distinct kinds of ingredients that go into the proofs: probabilistic and graph theoretical ones. While the probabilistic ideas mostly pertain to the results of this article, the graph theoretical results could be of independent interest and find other applications.

To get started, we notice that there are two natural ways for a graph to have $k$ triangles: either possessing $k$ vertex-disjoint triangles or a clique of size approximately $(6k)^{1/3}$. Using a standard second moment bound, one can obtain rather sharp estimates on the probability of such events, which yields a lower bound for the probability $P(N \geq k)$.

Most of the article is devoted to the highly delicate task of obtaining matching upper bounds and thereby establishing that these are the ‘only’ ways, probabilistically speaking.

To accomplish this, on the event that $G_n$ has $k$ triangles, we focus on the subgraph induced by the triangles and analyze the connected components. We call such a component a triangle induced graph, more generally, a graph $G = (V, E)$ is called a triangle-induced graph (to be called a TIG from here on) if it is a connected graph and can be obtained by taking a union of triangles (see Definition 1).

The main goal is now to obtain a sharp upper bound on the probability that $G_n$ contains such a TIG. Equipped with the same, the van den Berg-Kesten (BK) inequality (which bounds the probability of disjoint occurrences of several events), then allows us to bound the probability of several disjoint TIGs, thereby establishing the desired upper bound for $P(N \geq k)$. More precisely, note that the event $\{N \geq k\}$ implies the occurrence of vertex-disjoint TIGs each induced by $\ell_i$, say for $i = 1, \cdots, m$, triangles with $1 = \ell_1 = \cdots = \ell_j < \ell_{j+1} \leq \cdots \leq \ell_m$ and $\ell_1 + \cdots + \ell_m = k$ (we denote this event by $E_{\ell_1, \cdots, \ell_m}$). Also, let $F_\ell$ be the event that there exists a subgraph which is a TIG induced by $\ell$ triangles. Then, by BK inequality,

$$P(E_{\ell_1, \cdots, \ell_m}) \leq P(E_{1, \cdots, 1}) P(F_{\ell_{j+1}}) \cdots P(F_{\ell_m}) \quad (9)$$

(taking $j$ many 1s indexing $E_{1, \cdots, 1}$).

First, it is straightforward to upper bound the probability $P(E_{1, \cdots, 1})$ (i.e. existence of $j$ vertex-disjoint triangles). It turns out that the upper bound obtained by a naive first moment method is indeed sharp, which can be established by computing its second moment:

$$P(E_{1, \cdots, 1}) \approx e^{-j \log j}. \quad (10)$$

Most of the new ideas go into the proof of the upper bound of $P(F_\ell)$, which is one of our key results. In other words, we estimate the probability that there exists a connected subgraph, say $H$, induced by $\ell$ triangles. If $H$ has $v$ vertices and $e$ edges, then the probability that $G_n$ contains such an $H$ is bounded by

$$n^v v^{2e} \cdot \left( \frac{d}{n} \right)^e = v^{2e} d^e \frac{1}{n^{v-e}},$$

since the number of such subgraphs in $G_n$ is bounded by $n^v \binom{v}{2}^e \leq n^v v^{2e}$ (the $n^v$ term coming from the choices of the vertices, while $\binom{v}{2}^e$ bounds the number of choices for edges).
The proof now hinges on a crucial graph-theoretical ingredient which we prove in Section 3, which states that \( e - v \) can be lower bounded in terms of the number of triangles. In other words, we deduce that there exists an ‘almost’ (we will not make this notion precise in this discussion) concave function \( h \) with \( h(y) \approx ay^{2/3} \) (\( a \) is an explicit constant) such that

\[
e - v \geq h(\ell)
\]

(The function \( h \) is essentially the inverse of the function appearing on the RHS of the inequality in Lemma 3.3. Several analytic properties of the function \( h \) which are exploited crucially are established in Section 4.) The next step implements an efficient union bound scheme over all possible subgraphs \( H \) induced by \( \ell \) triangles. Without going into details, let us just mention that this step relies on establishing that any such \( H \) must admit a further ‘dense’ subgraph (where the number of edges is proportional to the square of the number of vertices) which accounts for most of the triangles.

Putting the above together, we essentially establish that

\[
\mathbb{P}(F_{\ell}) \leq \left( \frac{1}{n} \right)^{h(\ell)}
\]  

(see Lemma 4.4 for a precise statement). It might be worth mentioning that the recent preprints [3, 2] study the somewhat related notion of large deviation properties of the connected components in sparse random graphs.

Therefore, applying (10) and (11) to (9), we approximately have an upper bound

\[
\mathbb{P}(E_{\ell_1, \ldots, \ell_m}) \leq e^{-j \log j \left( \frac{1}{n} \right)^{h(\ell_{j+1} + \cdots + \ell_m)}}.
\]  

By concavity properties of \( h \), this quantity is bounded by

\[
e^{-j \log j \left( \frac{1}{n} \right)^{h(\ell_{j+1} + \cdots + \ell_m)}} = e^{-j \log j \left( \frac{1}{n} \right)^{h(k-j)}} \approx e^{-j \log j \left( \frac{1}{n} \right)^{a(k-j)^{2/3}}},
\]  

where the last approximate equality is obtained using the aforementioned asymptotics of the function \( h \).

Towards analyzing the bound in (13), let us define a function \( \phi(j) := e^{-j \log j n^{-a(k-j)^{2/3}}} \). Then, by the above,

\[
\mathbb{P}(E_{\ell_1, \ldots, \ell_m}) \leq \phi(j)
\]  

(recall that there are \( j \) 1s in \( \ell_1, \ldots, \ell_m \)). It turns out, and is not difficult to check, that the function \( \phi \) exhibits the following transition.

\[
\begin{cases}
k^{1/3} \log k < a \log n \Rightarrow \phi(j) \text{ attains the maximum at } j \approx k \\
k^{1/3} \log k > a \log n \Rightarrow \phi(j) \text{ attains the maximum at } j \approx 0.
\end{cases}
\]

Using (14), we obtain an upper bound for \( \mathbb{P}(E_{\ell_1, \ldots, \ell_m}) \) by taking a union bound over all possible tuples \( (\ell_1, \ldots, \ell_m) \) such that \( \ell_1 \leq \cdots \leq \ell_m \) and \( \ell_1 + \cdots + \ell_m = k \). Since by [4] the number of such partitions grows like \( e^{\sqrt{k}} \), essentially the bound in (14) prevails.

The aforementioned argument also allows us to prove Theorem 1.2. In the sub-critical regime, by (15), conditioned on possessing \( k \) triangles, it is likely to have ‘almost’ \( k \) vertex-disjoint triangles, whereas, in the super-critical regime, with high probability, \( j \approx 0 \) and thus \( \ell_{j+1} + \cdots + \ell_m \approx k \). Further, the application of Jensen’s inequality in (13) is sharp if and only if \( \ell_m \approx k \) and other \( \ell_i \)'s are negligible. This implies that with high probability conditioned on \( \{N \geq k\} \), there exists a TIG induced by \( k' \approx k \) triangles.
At this point another crucial graph theoretic result is proven. Namely, that conditioned on a TIG having \( k' \) triangles, it is likely that this component contains an approximate clique accounting for almost \( k' \) triangles. We will not elaborate on the proof in this discussion beyond mentioning that this involves proving a structure theorem quantifying when the following well known inequality (known as Kruskal-Katona bound, see e.g. [1]), bounding the number of triangles in any graph in terms of the number of edges,

\[
\Delta \leq \frac{\sqrt{2}}{3}|E|^{3/2},
\]

is almost sharp (it is a straightforward computation to check that, ignoring lower order terms, equality holds when the graph is a clique).

Though it has been pointed out to us, after the completion of the paper, by Noga Alon that a similar structural result had already been proved quite a few years back in [24], we decided to keep the proof, since it seems to be somewhat different from the previous one, borrows ideas from spectral graph theory and could be of independent interest.

1.4. Organization of paper. The rest of the paper is organized as follows. In Section 2, we deduce lower bounds for \( \mathbb{P}(N \geq k) \). In Section 3, we introduce and prove several graph theoretical results that have been already alluded to in Section 1.3. In Section 4, we establish a sharp upper bound on the probability that \( G_n \) contains a subgraph which is a TIG induced by \( \ell \) triangles (i.e. a precise version of (11)). Using this combined with a technical convexity argument, in Section 5 we proceed with the details of (12)-(15) and deduce a matching upper bound on \( \mathbb{P}(N \geq k) \). Finally, in Section 6, we prove Theorem 1.2.

Before embarking on the proofs, we remark that, adopting standard practice, often in the proofs, we will use the same constant, say \( C \), whose value might change from line to line.

1.5. Acknowledgement. S.G. thanks Sourav Chatterjee for mentioning to him the question of studying the upper tail of \( N \), which he had learnt from David Aldous. S.G. also thanks Noga Alon for pointing him to [24]. His research was partially supported by NSF grant DMS-1855688, NSF CAREER grant DMS-1945172, and a Sloan Fellowship. E.H. was partially supported by NSF grant DMS-1855688.

2. Lower bound

We start by establishing lower bounds for the tail probability of \( N \) by analyzing two specific events, namely having \( k \) disjoint triangles and having a clique of size approximately \((6k)^{1/3}\). Let us introduce the following parameters which will recur throughout the article,

\[
d_0 := \frac{d^3}{6}, \quad d_1 := \frac{ed^3}{6}.
\]

To get started, let \( H_k \) be the graph of size \( 3k \) vertices formed by \( k \) disjoint triangles.

**Lemma 2.1.** There is a constant \( c > 0 \) depending only on \( d \), such that the following holds. Suppose that \( k \to \infty \) and \( \frac{k^2}{n} \to 0 \) as \( n \to \infty \). Then, for large enough \( n \),

\[
\mathbb{P}(G_n \text{ contains } H_k) \geq e^{-k \log k - ck}.
\]

**Proof.** Let \( X \) be the number of subgraphs \( H_k \) in \( G_n \), which in addition satisfy that vertices in \( H_k \) have no other edges. In other words,

\[
X = \sum_{T_1, \ldots, T_k \text{ disjoint triangles}} 1(\text{edges are present in } T_1, \ldots, T_k),
\]
all vertices in $T_1 \cup \cdots \cup T_k$ have no other edges).

Note that there are $\frac{1}{k!} \binom{n}{3} \cdots \binom{n-3k+3}{3}$ terms in the summation. Thus, for any $\varepsilon > 0$, for large enough $k$,

$$
\mathbb{E}X = \frac{1}{k!} \binom{n}{3} \cdots \binom{n-3k+3}{3} \left( \frac{d}{n} \right)^{3k} \left( 1 - \frac{d}{n} \right)^{3k(n-3k) + \frac{3k(3k-1)}{2} - 3kk} \quad (17)
$$

$$
\geq \frac{1}{k!} \left( \frac{d^3}{6} \right)^k \left( 1 - \frac{3k}{n} \right)^{3n^3} \left( 1 - \frac{d}{n} \right)^{3kn} \geq \frac{1}{k!} \left( \frac{d^3}{6} \right)^k e^{-(3+\varepsilon)dk}.
$$

(We used $\lim_{n \to \infty} (1 - \frac{1}{n})^n = e^{-1}$ and the condition $\frac{k^2}{n} \to 0$ to deduce the last inequality.) Recall that by Stirling’s formula,

$$
\left( \frac{k}{e} \right)^k < k! < C \sqrt{k} \left( \frac{k}{e} \right)^k \quad (18)
$$

Thus, further using that $d_1 = \frac{ed^3}{6}$ from (16),

$$
\mathbb{E}X \geq C \frac{1}{\sqrt{k}} \frac{d^k}{k^k} e^{-(3+\varepsilon)dk} = C e^{-k \log k + (\log d_1 - (3+\varepsilon)d)k - \frac{1}{2} \log k}. \quad (19)
$$

Let us now compute the second moment of $X$. Since we require in the definition of $X$ that the vertices of the $k$ triangles have no other edges, computing the second moment of $X$ involves considering the following situation: two sets of $k$ triangles share some triangles and are otherwise vertex-disjoint. The contribution from such pairs of subgraphs $H_k$s sharing $k - \ell$ triangles for $\ell = 0, 1, \cdots, k$ is given by

$$
\sum_{\ell=0}^{k} \frac{1}{(k+\ell)!} \binom{n}{3} \cdots \binom{n-3(k+\ell)+3}{3} \left( \frac{d}{n} \right)^{3k+\ell} \left( 1 - \frac{d}{n} \right)^{3k(n-3(k+\ell)+\frac{3(k+\ell)(3k+\ell-1)}{2} - 3(k+\ell))} \cdot \left( \frac{d}{n} \right)^{3k+\ell} \left( 1 - \frac{d}{n} \right)^{3k(n-3(k+\ell)+\frac{3k+\ell)(3k+\ell-1)}{2} - 3(k+\ell)}.
$$

Above, the product of the $k + \ell$ many binomial $\binom{n-3(k+\ell)}{3}$ terms, along with the $\frac{1}{(k+\ell)!}$, comes from choosing $k + \ell$ disjoint triangles. The $\binom{k+\ell}{k}$ term is exactly the number of possible ways to choose two subgraphs $H_k$s sharing $k - \ell$ triangles. Since we require that vertices in $k + \ell$ triangles cannot have other edges, we obtain the last two terms.

Note that the $\ell = 0$ term is simply $\mathbb{E}X$. To bound each term, since $\ell \leq k$ and $\frac{k^2}{n} \to 0$, for any $\varepsilon > 0$, for large enough $n$,

$$
\left( 1 - \frac{d}{n} \right)^{3(k+\ell)(n-3(k+\ell)+\frac{3(k+\ell)(3k+\ell-1)}{2} - 3(k+\ell))} \leq \left( 1 - \frac{d}{n} \right)^{3(k+\ell)n-O(k^2)} \leq e^{-(3-\varepsilon)d(k+\ell)}.
$$

Also, note that $\binom{k+\ell}{k} \leq 2^{k+\ell}$ and $\binom{k+\ell}{\ell} \leq 2^k$. Thus, for $\ell \geq 1$, using Stirling’s formula (18), each term in the summation above is bounded by

$$
\frac{1}{(k+\ell)!} \left( \frac{d^3}{6} \right)^{k+\ell} e^{-(3-\varepsilon)d(k+\ell)} 2^{k+\ell} \leq \left( \frac{d_1}{k+\ell} \right)^{k+\ell} e^{-(3-\varepsilon)d(k+\ell)} 2^{2k+\ell}
$$

(recall that $d_1 = \frac{ed^3}{6}$). Using (19), the above quantity is bounded by $a_\ell \mathbb{E}X$, where

$$
a_\ell = \left( \frac{d_1}{k+\ell} \right)^{k+\ell} e^{-(3-\varepsilon)d(k+\ell)} 2^{2k+\ell} C \left( \frac{1}{\sqrt{k}} \frac{d^k}{k^k} e^{-(3+\varepsilon)dk} \right)^{-1}
$$
\[ C^{-1} \frac{k^k}{(k + \ell)^{k+\ell}} \frac{d!}{n^\ell} e^{2c\varepsilon dk} e^{-(3-\varepsilon)\varepsilon d^2 \ell^2} 2^{2k+\ell} \sqrt{k} \]

\[ \leq C \frac{1}{(k + \ell)^\ell} \left( d! e^{2c\varepsilon d^2} \right)^{k+\ell} \sqrt{k} \leq C^{k+\ell} \frac{\sqrt{k}}{(k + \ell)^\ell} \leq C^k \left( \frac{C}{k + \ell} \right)^\ell \]

\( (C > 1 \text{ is a constant depending on } d \text{ whose value above changes from line to line}) \). Thus, for a large enough \( k \),

\[ \sum_{\ell=0}^k a_\ell \leq (k + 1)C^k < C^{2k}. \]

Hence,

\[ \mathbb{E}X^2 \leq C^{2k} \mathbb{E}X. \]

Therefore, since

\[ \mathbb{P}(X \geq 1) \geq \frac{(\mathbb{E}X)^2}{\mathbb{E}X^2} \geq C^{-2k} \mathbb{E}X \]

combined with (19), we conclude the proof.

\[ \square \]

\textbf{Remark 2.2.} For later purposes we record an analogous upper bound using the first moment method. The expectation of the number of subgraphs \( H_k \) is

\[ \frac{1}{K!} \left( \frac{n}{3} \right) \left( \frac{n-3}{3} \right) \cdots \left( \frac{n-3k+3}{3} \right) \left( \frac{d}{n} \right)^{3k} \]

(20)

Note that this is nothing other than the quantity (17) in the previous lemma, without the power of \( 1 - \frac{d}{n} \) term. This is because we no longer require vertices in \( H_k \) to have no other edges. The quantity (20) is bounded by

\[ \frac{1}{K!} \left( \frac{d^3}{6} \right)^k n(n-1) \cdots (n-3k+1) \leq \frac{1}{K!} \left( \frac{d^3}{6} \right)^k \leq e^{-k \log k + c'k} \]

for some constant \( c' > 0 \). Hence,

\[ \mathbb{P}(G_n \text{ contains } H_k) \leq e^{-k \log k + c'k}. \]  

(21)

The second lower bound is obtained by planting a clique. For a positive integer \( k \), let \( K_k \) be a clique of size \( k \). The following lemma is rather straightforward and we omit its proof.

\textbf{Lemma 2.3.} For any \( n \) and \( k \),

\[ \mathbb{P}(G_n \text{ contains } K_k) \geq \left( \frac{d}{n} \right)^{\binom{k}{2}}. \]

In particular, for any \( \varepsilon > 0 \), for large enough \( n \) and \( k \),

\[ \mathbb{P}(G_n \text{ contains a clique of size at least } (6k)^{1/3}) \geq n^{-(1+\varepsilon)\left( \frac{3}{32} \right)^2 k^{2/3}}. \]

Therefore, combined with Lemma 2.1, we have the following result.

\textbf{Proposition 2.4.} There is a constant \( c > 0 \), depending only on \( d \), such that the following holds. Suppose that \( k \to \infty \) and \( \frac{k^2}{n} \to 0 \) as \( n \to \infty \). Then, for any \( \varepsilon > 0 \), for large enough \( n \),

\[ \mathbb{P}(N \geq k) \geq \max(e^{-k \log k - ck}, n^{-(1+\varepsilon)\left( \frac{3}{32} \right)^2 k^{2/3}}). \]
A simple calculation now shows that if \( k^{1/3} \log k < ((\frac{3}{\sqrt{2}})^{2/3} - \delta) \log n \), then
\[
e^{-k \log k - ck} > n^{-(1+\varepsilon)(\frac{3}{\sqrt{2}})k^{2/3}/},
\]
whereas, if \( k^{1/3} \log k > ((\frac{3}{\sqrt{2}})^{2/3} + \delta) \log n \) instead, then for small enough \( \varepsilon > 0 \),
\[
e^{-k \log k - ck} < n^{-(1+\varepsilon)(\frac{3}{\sqrt{2}})k^{2/3}/}.
\]

3. Graph theoretic results

As indicated in Section 1.3, the proofs throughout the article relies on key graph theoretic results, some of which we believe to be of independent interest, potentially having other applications. In this section, we record and prove all such results.

From now on, for any subgraph \( H \) of \( G \), define \( V(H) \) and \( E(H) \) to be the set of vertices and edges of \( H \) respectively. Recall also that \( \Delta(G) \) denotes the number of triangles in \( G \). To alleviate the notation, we simply use the notation \( \Delta \) when \( G \) is clear from the context. Also, for any graph (which will be clear from the context), we denote the corresponding graph metric by \( d(\cdot, \cdot) \).

We start with the following well known Kruskal-Katona bound on the number of triangles in a graph in terms of the number of edges. The proof is an application of Hölder’s inequality.

**Lemma 3.1.** [6, Lemma 2.2], [1] For any graph \( G = (V, E) \),
\[
\Delta \leq \frac{\sqrt{2}}{3}|E|^{3/2}. \tag{22}
\]

We next restate the following crucial definition from Section 1.

**Definition 1.** A graph \( G = (V, E) \) is called a triangle-induced graph (to be called a TIG from here on) if it is a connected graph and can be obtained by taking a union of triangles. We say that a TIG \( G = (V, E) \) is spanned by \( \ell \) triangles if there exist (distinct) triangles \( T_1, \ldots, T_\ell \) such that \( V = \cup_i V(T_i) \) and \( E = \cup_i E(T_i) \).

\[\text{Figure 1. Example of a triangle-induced graph, TIG, with 8 vertices spanned by 5 triangles. Here the graph is sequentially constructed as a union of triangles, and at each step the newly added triangle is highlighted using bold edges. Note that this graph is also spanned by 6 triangles.}\]

Note that, e.g., although a clique of size 4 contains four triangles, it can be spanned by three triangles as well.

In the remainder of this section we prove three main results. First, we prove Lemma 3.2 which states that TIGs are far from trees, namely, it lower bounds the number of tree-excess edges in a TIG in terms of the total number of edges. Then, we prove Lemma 3.3, which upper bounds the number of triangles in a connected graph in terms of the number of tree-excess edges. The next result we prove is Lemma 3.4 which states that any connected graph having not too many tree-excess edges contains a dense sub-graph which accounts for most triangles. We finish with the proof of a structural result, Proposition 3.6, which states that a graph where the Kruskal-Katona
bound is almost sharp, necessarily contains a subgraph that is almost a clique, in the sense of the relation between the number of vertices, edges and triangles.

**Lemma 3.2.** Suppose that $G = (V, E)$ is a TIG. Then,

$$|E| \leq 5(|E| - |V|) + 5.$$  

**Proof.** Let $T$ be a spanning tree of $G$ (recall that $G$ is connected), and pick an arbitrary vertex as the root and denote it by $\rho$. We construct a map $\phi : E(T) \to E \setminus E(T)$ which is an at most 4 to 1 mapping. From now on, we call any element in $E \setminus E(T)$ an excess edge.

For any edge $e \in E(T)$, let $\Delta(e)$ be a triangle containing $e$; there may be several such in which case we choose an arbitrary one (such a triangle always exists since $G$ is a TIG). Note that since $T$ is a tree, not all edges in $\Delta(e)$ can belong to $E(T)$. This leads us to consider the following two cases.

**Type 1.** $\Delta(e)$ contains only one excess edge $E \setminus E(T)$: We denote this excess edge by $\phi(e)$.

**Type 2.** $\Delta(e)$ contains two excess edges $E \setminus E(T)$: Let $v$ and $w$ be the endpoints of the edge $e$. Since $d(\rho, v) \neq d(\rho, w)$ (recall that $e$ is an edge in a tree $T$), without loss of generality, we assume that $d(\rho, v) > d(\rho, w)$. Then, define $\phi(e)$ to be the excess edge in $\Delta(e)$ which has $v$ as an endpoint.

![Figure 2](image.png)

**Figure 2.** The solid edges in graph (A) form a spanning tree and the excess edges are dashed. We take node 1 as the root of the spanning tree. For any edge of the spanning tree, the node with the higher label is further away from the root.

**Claim.** Each excess edge in $E \setminus E(T)$ is mapped to by at most four edges $e \in E(T)$.

To see this first note that any excess edge $f$ is mapped to by at most two edges of type 1. In fact, if an excess edge $f$ is contained in two triangles $\Delta_1$ and $\Delta_2$ that contain only one excess edge, then there is a cycle of size 4 in $T$, which cannot be true since $T$ is a tree.

Therefore, it suffices to show that any excess edge $f = (v, w)$ is mapped to by at most two edges $e \in E(T)$ of type 2. Suppose that there exist two edges $e_1 = (v_1, v)$ and $e_2 = (v_2, v)$ in $E(T)$ such that $\phi(e_1) = \phi(e_2) = f$. Then, by the definition of $\phi$, $d(\rho, v_1) = d(\rho, v_2) = d(\rho, v) - 1$, which again violates the fact that $T$ is a tree. Hence, for any endpoint of an excess edge $f$, at most one edge $e \in E(T)$ of type 2 incident on it which is mapped to $f$. Thus, any $f = (v, w)$ is mapped to by at most two edges $e \in E(T)$ of type 2, concluding the proof of the claim.

Therefore, since $\phi : E(T) \to E \setminus E(T)$ is an at most 4 to 1 mapping, we have

$$E(T) \leq 4(|E| - |E(T)|),$$

which finishes the proof.

**Lemma 3.3.** Let $G = (V, E)$ be a connected graph. Then,

$$\Delta \leq \frac{\sqrt{3}}{3}(|E| - |V| + 1)^{3/2} + 3(|E| - |V| + 1).$$

**Proof.** As in the previous lemma, let $T$ be a spanning tree of $G$ and $\rho$ be an arbitrary root. Since $T$ is a tree, each triangle $\Delta$ in $G$ is of one of the following three types.
Type 1. $\Delta$ has one excess edge.
Type 2. $\Delta$ has two excess edges.
Type 3. $\Delta$ has three excess edges.

We define a map $\phi$ from the set of triangles of type 1 or type 2 to $E \setminus E(T)$, the set of excess edges. For any triangle $\Delta$ of type 1, define $\phi(\Delta)$ to be the unique edge in $E \setminus E(T)$ contained in $\Delta$. For any triangle $\Delta$ of type 2, let $e = (v, w)$ be the unique edge of $\Delta$ in $E(T)$. Without loss of generality, we assume $d(\rho, v) > d(\rho, w)$. Then, define $\phi(\Delta)$ to be the edge in $E \setminus E(T)$ which has $v$ as an endpoint.

Since $T$ is a tree, each edge in $E \setminus E(T)$ is mapped to by at most one triangle of type 1. In addition, by the same reasoning as in the proof of Lemma 3.2, each edge in $E \setminus E(T)$ is mapped to by at most two triangles of type 2. Hence,

$$|\text{type 1 triangles}| + |\text{type 2 triangles}| \leq 3(|E| - |E(T)|) = 3(|E| - |V| + 1). \tag{23}$$

In order to control the number of triangles of type 3, we use Lemma 3.1 upper bounding the number of triangles in a graph in terms of the number of edges,

$$|\text{type 3 triangles}| \leq \frac{\sqrt{2}}{3}(|E| - |V| + 1)^{3/2}. \tag{24}$$

Therefore, combining (23) and (24), we conclude the proof.

The next result states that if a connected graph has not too many tree excess edges, then it admits a dense subgraph, in the sense that the number of edges is of the order of the square of the vertices, which accounts for most of its triangles.

**Lemma 3.4.** Let $\xi > 0$ be a sufficiently small constant. Let $G = (V, E)$ be a connected graph containing at least $\ell$ triangles such that

$$|E| - |V| \leq \frac{1}{2} \xi^{-1/2} \ell^{2/3} - 1.$$  

Then, for sufficiently large $\ell$, there exists a subgraph $G'$ in $G$ such that

$$|V(G')| \leq \xi^{-3/2} \ell^{1/3}, \tag{25}$$

$$|E(G')| \geq \left(\frac{3}{\sqrt{2}}\right)^{3/2} (1 - 2\xi^{1/2})^{2/3} \ell^{2/3}, \tag{26}$$

$$|\Delta(G')| \geq (1 - 2\xi^{1/2}) \ell. \tag{27}$$

We need the following easy lemma.

**Lemma 3.5.** Let $G = (V, E)$ be a connected graph. Then there is a subgraph $G'$ such that

$$\Delta(G') \geq \Delta(G) - 3(|E| - |V| + 1), \quad |E(G')| = |E| - |V| + 1. \tag{28}$$

**Proof.** Consider a spanning tree $T$ of $G$. Let $G'$ be the subgraph induced by the excess edges $E \setminus E(T)$. Then, in terms of the terminology in the proof of Lemma 3.3, $\Delta(G')$ is nothing other than the number of type 3 triangles. Therefore, by (23), we have (28).

Equipped with this lemma, we proceed to proving Lemma 3.4.

**Proof of Lemma 3.4.** By Lemma 3.5, there exists a subgraph $G_1$ such that

$$\Delta(G_1) \geq \ell - \frac{3}{2} \xi^{-1/2} \ell^{2/3}, \quad |E(G_1)| \leq \frac{1}{2} \xi^{-1/2} \ell^{2/3}.$$
For \( v \in V(G_1) \), let \( d_v \) be the degree of \( v \) in \( G_1 \). Define
\[
A := \{ v \in V(G_1) : d_v \leq \xi \ell^{1/3} \}.
\]
Then, the number of triangles in \( G_1 \) having at least one vertex in \( A \) is bounded by
\[
\sum_{v \in A} d_v^2 \leq \xi \ell^{1/3} \sum_{v \in A} d_v \leq 2 \xi \ell^{1/3} |E(G_1)| \leq \xi \ell^{1/2} \ell.
\] (29)
Denote by \( G' \) the subgraph of \( G_1 \) induced by the vertex set \( V(G_1) \setminus A \). Then, for large enough \( \ell \),
\[
\Delta(G') \geq \ell - \frac{3}{2} \xi^{-1/2} \ell^{2/3} - \xi^{1/2} \ell \geq (1 - 2 \xi^{1/2}) \ell.
\]
Thus, we have (27), and by the bound in Lemma 3.1, we subsequently obtain (26). In addition, using the fact
\[
\xi^{-1/2} \ell^{2/3} \geq 2E(G_1) \geq \sum_{v \in G'} d_v \geq |V(G')| \xi \ell^{1/3},
\]
we have (25).

The final result of this section provides a structural description of when (22) is almost sharp. As has already been mentioned in Section 1.3, after the completion of the paper, Noga Alon pointed out to us that a similar “stability” result for the Kruskal-Katona theorem had been proved earlier in [24]. However, since our proof appears to be different and could be of independent interest, we decided to include the proof.

**Proposition 3.6.** Let \( \varepsilon > 0 \) be a sufficiently small constant. Let \( G = (V, E) \) be a graph with at least \( \frac{1}{6}(1-\varepsilon)k \) many triangles and at most \( (1+\varepsilon)\frac{1}{6}k^{2/3} \) many edges. Then, there is a subset \( V' \subseteq V \) such that \( |V'| \leq (1 + 6\varepsilon^{1/4})k^{1/3} \) and \( \Delta(V') \geq \frac{1}{6}(1 - 6\varepsilon^{1/4})k \).

In order to prove this, we analyze the principal eigenvector of the adjacency matrix of \( G \). Note that the top eigenvector of \( G \) is ‘almost’ uniformly distributed. We show that the top eigenvector of \( G \) is ‘almost’ uniformly distributed on some subset \( V' \subseteq V \) with \( |V'| \approx k^{1/3} \).

**Proof.** Let \( \lambda_1 \geq \cdots \geq \lambda_n \) and \( v_1, \cdots, v_n \) be eigenvalues and eigenvectors of the adjacency matrix of \( G \) which is denoted by \( A \). Then,
\[
\text{tr}(A^2) = \sum \lambda_i^2 = 2|E| \leq (1 + \varepsilon)k^{2/3},
\]
\[
\text{tr}(A^3) = \sum \lambda_i^3 = 6\Delta \geq (1 - \varepsilon)k.
\]
Since \( \sum \lambda_i^3 \leq \lambda_1 \sum \lambda_i^2 \) (recall that \( \lambda_1 \) is positive by the Perron-Frobenius theorem), we have
\[
\frac{1 - \varepsilon}{1 + \varepsilon} k^{1/3} \leq \lambda_1 \leq (1 + \varepsilon)^{1/2} k^{1/3},
\] (30)
which implies
\[
\lambda_2^2 + \cdots + \lambda_n^2 \leq (1 + \varepsilon - \left( \frac{1 - \varepsilon}{1 + \varepsilon} \right)^2) k^{2/3}.
\] (31)
By (30) and (31), \( \lambda_2, \cdots, \lambda_n \) are negligible compared to \( \lambda_1 \) for small enough \( \varepsilon > 0 \). Motivated by this, we decompose
\[
A = \lambda_1 v_1 v_1^T + \sum_{i=2}^n \lambda_i v_i v_i^T =: A_1 + A_2.
\]
Define
\[ \tilde{\varepsilon} := 1 + \varepsilon - \left( \frac{1 - \varepsilon}{1 + \varepsilon} \right)^2. \] (32)

Then, for small enough \( \varepsilon > 0 \),
\[ \varepsilon < \tilde{\varepsilon} < 8\varepsilon, \] (33)
and by (31),
\[ \|A_2\|_F^2 = \sum_{i=2}^{n} \lambda_i^2 \leq \tilde{\varepsilon} k^{2/3}. \] (34)

Letting \( v_1 = ((v_1)_i)_{i=1,...,n} \), we next partition the vertex set \( V = \{1, \cdots, n\} \) in the following way.

\[ S_1 := \left\{ i : |(v_1)_i| > \frac{1 + \tilde{\varepsilon}^{1/4}}{k^{1/6}} \right\}, \]
\[ S_2 := \left\{ i : \frac{\tilde{\varepsilon}^{1/8}}{k^{1/6}} < |(v_1)_i| < \frac{1 - \tilde{\varepsilon}^{1/4}}{k^{1/6}} \right\}, \]
\[ S_3 := \left\{ i : |(v_1)_i| \leq \frac{\tilde{\varepsilon}^{1/8}}{k^{1/6}} \right\}, \]
\[ S_4 := \left\{ i : \frac{1 - \tilde{\varepsilon}^{1/4}}{k^{1/6}} \leq |(v_1)_i| \leq \frac{1 + \tilde{\varepsilon}^{1/4}}{k^{1/6}} \right\}. \]

We will ultimately take \( S_4 \) to be \( V' \) appearing in the statement of the proposition. To be able to show this, we need to establish the contributions coming from the remainder of the graph to be negligible.

Towards this we first show that
\[ |S_1| < 2\tilde{\varepsilon}^{1/4}k^{1/3}. \] (35)

Using (30) and the definition of \( S_1 \), it follows that each entry of the sub-matrix obtained by the restriction of \( A_1 = \lambda_1 v_1 v_1^T \) to \( S_1 \times S_1 \) has absolute value greater than \( (1 + \tilde{\varepsilon}^{1/4})^2 \frac{1 - \varepsilon}{1 + \varepsilon} - 1 > (1 + 2\tilde{\varepsilon}^{1/4}) \frac{1 - \varepsilon}{1 + \varepsilon} - 1 = 2\tilde{\varepsilon}^{1/4} - \frac{2\varepsilon}{1 + \varepsilon} \). Since each entry of \( A \) is either 0 or 1 (as it is the adjacency matrix of a graph), for small \( \varepsilon > 0 \), absolute value of each entry in the sub-matrix of \( A_2 = A - A_1 \) restricted to \( S_1 \times S_1 \) is greater than
\[ (1 + \tilde{\varepsilon}^{1/4})^2 \frac{1 - \varepsilon}{1 + \varepsilon} - 1 > (1 + 2\tilde{\varepsilon}^{1/4}) \frac{1 - \varepsilon}{1 + \varepsilon} - 1 = 2\tilde{\varepsilon}^{1/4} - \frac{2\varepsilon}{1 + \varepsilon} (1 + 2\tilde{\varepsilon}^{1/4}) > 2\tilde{\varepsilon}^{1/4} - 4\varepsilon > \frac{1}{2} \tilde{\varepsilon}^{1/4}. \]

This implies that the Frobenius norm of \( A_2 \) restricted to \( S_1 \times S_1 \) is greater than \( \frac{1}{2} \tilde{\varepsilon}^{1/4} \sqrt{2} |S_1| \). Hence, by (34), we obtain (35).

Next, we show that
\[ |S_2| < 2\tilde{\varepsilon}^{1/4}k^{1/3}. \] (36)

Using (30) and the definition of \( S_2 \), it follows that each entry of the sub-matrix \( A_1 = \lambda_1 v_1 v_1^T \) restricted to \( S_2 \times S_2 \) has an absolute value between \( \tilde{\varepsilon}^{1/4} \frac{1 - \varepsilon}{1 + \varepsilon} \) and \( (1 - \tilde{\varepsilon}^{1/4})^2 (1 + \varepsilon)^{1/2} \). For small enough \( \varepsilon > 0 \), \( \tilde{\varepsilon}^{1/4} \frac{1 - \varepsilon}{1 + \varepsilon} > \frac{1}{2} \tilde{\varepsilon}^{1/4} \) and
\[ 1 - (1 - \tilde{\varepsilon}^{1/4})^2 (1 + \varepsilon)^{1/2} > 1 - (1 - \tilde{\varepsilon}^{1/4}) \left( 1 + \frac{1}{2} \varepsilon \right) > \tilde{\varepsilon}^{1/4} - \frac{1}{2} \varepsilon > \frac{1}{2} \tilde{\varepsilon}^{1/4}. \]
Since each entry of $A$ is either 0 or 1, each entry of $A_2$ restricted to $S_2 \times S_2$ is greater than $\frac{1}{2} \varepsilon^{1/4}$. This implies that the Frobenius norm of $A_2$ restricted to $S_2 \times S_2$ is greater than $\frac{1}{2} \varepsilon^{1/4} \sqrt{|S_2|}$, and in conjunction with (34) implies (36).

In addition, since $\|v\|_2 = 1$, by Markov's inequality

$$|S_4| \leq \frac{1}{(1 - \varepsilon^{1/4})^2} k^{1/3}. \quad (37)$$

Recalling that $\Delta(W)$ denotes the number of triangles consisting of vertices in $W$ and that our goal is to prove that most of the triangles in $G$ come from $S_4$, we next show that for small enough $\varepsilon$,

$$\Delta(G) - \Delta(S_4) \leq 3\varepsilon^{1/4} k. \quad (38)$$

The remainder of the proof is devoted to proving the above. There are quite a few technical steps. First, using the bounds from (35) and (36), we bound the number of triangles having at least one vertex in $S_1$ or $S_2$. Then, using the fact that $|\langle v_1 \rangle|$ is small for $i \in S_3$, we argue that there cannot be many triangles having at least one vertex in $S_3$.

We start by defining some new notation. For any $W_1, W_2 \subseteq V$, let $\Delta_1(W_1, W_2)$ be the set of triangles having one vertex in $W_1$ and two vertices in $W_2$, $\Delta_2(W_1, W_2)$ be the set of triangles having two vertices in $W_1$ and one vertex in $W_2$, and set $\Delta(W_1, W_2) = \Delta_1(W_1, W_2) + \Delta_2(W_1, W_2)$. In addition, for $W_1, W_2, W_3 \subseteq V$, let $\Delta(W_1, W_2, W_3)$ be the set of triangles with one vertex each in $W_1, W_2, W_3$ respectively.

Since $\{S_1, S_2, S_3, S_4\}$ is a partition of the vertex set,

$$\Delta(G) - \Delta(S_4) = \Delta(S_1) + \Delta(S_2) + \Delta(S_3) + \Delta(S_1, S_2) + \Delta(S_1, S_3) + \Delta(S_1, S_4) + \Delta(S_2, S_3) + \Delta(S_2, S_4) + \Delta(S_3, S_4) + \Delta(S_1, S_2, S_3) + \Delta(S_1, S_2, S_4) + \Delta(S_1, S_3, S_4) + \Delta(S_2, S_3, S_4). \quad (39)$$

We now seek to control each term.

By (35), using $|E| < (1 + \varepsilon)^{\frac{1}{2}} k^{2/3}$,

$$\Delta(S_1) + \Delta(S_1, S_2) + \Delta(S_1, S_3) + \Delta(S_1, S_4) + \Delta(S_1, S_2, S_3) + \Delta(S_1, S_2, S_4) + \Delta(S_1, S_3, S_4) \leq |S_1| \cdot |E| \leq \varepsilon^{1/4}(1 + \varepsilon)k. \quad (40)$$

The above uses the trivial bound that the number of triangles with one endpoint in $S_1$ is at most $|S_1| \cdot |E|$.

Similarly, by (36),

$$\Delta(S_2) + \Delta(S_2, S_1) + \Delta(S_2, S_3) + \Delta(S_2, S_4) + \Delta(S_1, S_2, S_4) + \Delta(S_2, S_3, S_4) \leq |S_2| \cdot |E| \leq \varepsilon^{1/4}(1 + \varepsilon)k. \quad (41)$$

Consider the subgraph $G_1$ induced by edges with both endpoints in $S_3$, or one in $S_3$ and $S_2 \cup S_4$ each. We show that

$$|E(G_1)| < \varepsilon^{2/3} k. \quad (42)$$
Note that by definition of $S_2, S_3, S_4$, for any vertices $i, j$ connected by an edge in $G_1$, $$(v_i)_i(v_j)_j \leq \frac{\tilde{\varepsilon}^{1/2}}{k^{1/6}} \cdot \frac{1 + \varepsilon^{1/4}}{k^{1/6}}.$$ By (30), the absolute value of each entry of $A_1$ restricted to $G_1$ is (upper) bounded by $$\frac{\tilde{\varepsilon}^{1/2}}{k^{1/6}} \cdot \frac{1 + \varepsilon^{1/4}}{k^{1/6}} \cdot (1 + \varepsilon)^{1/2} k^{1/3} = (1 + \varepsilon)^{1/2} \tilde{\varepsilon}^{1/8} (1 + \varepsilon^{1/4}).$$ Suppose that (42) fails, i.e., that there are at least $\tilde{\varepsilon} k^{2/3}$ many edges in $G_1$. Then, in the matrix $A_2$, there are at least $2\tilde{\varepsilon} k^{2/3}$ many entries greater than $1 - (1 + \varepsilon)^{1/2} \tilde{\varepsilon}^{1/8} (1 + \varepsilon^{1/4})$. This implies that if (42) fails, then the Frobenius norm of $A_2$ is greater than $$\sqrt{2}\tilde{\varepsilon}^{1/2} k^{1/3} (1 - (1 + \varepsilon)^{1/2} \tilde{\varepsilon}^{1/8} (1 + \varepsilon^{1/4})).$$ This contradicts (34) for small $\varepsilon > 0$. Hence, we have (42) which in particular implies that the number of edges in the subgraph induced by $S_3$, denoted by $|E(S_3)|$, satisfies $$|E(S_3)| \leq |E(G_1)| \leq \tilde{\varepsilon} k^{2/3}. \tag{43}$$ Hence, by Lemma 3.1, $$\Delta(S_3) \leq \frac{\sqrt{2} \tilde{\varepsilon}^{1/2}}{3} k^{3/2}. \tag{44}$$ Finally, in order to bound (39), the last quantity we need to bound is $\Delta(S_3, S_4)$. We show that $$\Delta(S_3, S_4) \leq \left( \frac{\tilde{\varepsilon}}{(1 - \tilde{\varepsilon}^{1/4})^2} + (1 + \varepsilon) \tilde{\varepsilon}^{1/2} \right) k. \tag{45}$$ First, by (37) and (43), $$\Delta_2(S_3, S_4) \leq |E(S_3)| \cdot |S_4| \leq \frac{\tilde{\varepsilon}}{(1 - \tilde{\varepsilon}^{1/4})^2} k. \tag{46}$$ Next, in order to bound $|\Delta_1(S_3, S_4)|$, let $d_1, \ldots, d_m$ be the number of neighbors in $S_3$ of each vertex in $S_3$, i.e., the degree of each vertex of $S_3$ in $S_4$, and let $$W := \{ i \in S_3 : d_i < \varepsilon^{1/2} k^{1/3} \}.$$ By (42), $$|W^c \cap S_4| < \varepsilon^{1/2} k^{1/3}. \tag{47}$$ Since the number of edges in the subgraph induced by $S_4$ is upper bounded by the total number of edges $\frac{1}{2}(1 + \varepsilon) k^{2/3}$, the number of triangles having one vertex in $W^c \cap S_3$ and two vertices in $S_4$ is bounded by $\frac{1}{2}(1 + \varepsilon) \tilde{\varepsilon}^{1/2} k$. In addition, using (42), the number of triangles having one vertex in $W$ and two vertices in $S_4$ is bounded by $$\sum_{i \in W} d_i^2 \leq \left( \max_{i \in W} d_i \right) (d_1 + \cdots + d_m) \leq \varepsilon^{1/2} k^{1/3} \cdot \varepsilon^{1/2} k^{2/3} = \varepsilon^{3/2} k.$$

Thus, we have $$\Delta_1(S_3, S_4) \leq \frac{1}{2} (1 + \varepsilon) \varepsilon^{1/2} k + \varepsilon^{3/2} k \leq (1 + \varepsilon) \varepsilon^{1/2} k. \tag{47}$$
Combining (46) and (47), we have (45). Therefore, applying (40), (41) and (44) to (39), we obtain (38). Hence,

\[ \Delta(S_4) \geq \Delta(G) - 3\varepsilon^{1/4}k \geq \left( \frac{1}{6}(1 - \varepsilon) - 3\varepsilon^{1/4} \right)k \geq \frac{1}{6}(1 - 6\varepsilon^{1/4})k. \]

In addition, by (37),

\[ |S_4| \leq \frac{1}{(1 - \varepsilon^{1/4})^2}k^{1/3} \leq (1 + 3\varepsilon^{1/4})k^{1/3} \leq (1 + 6\varepsilon^{1/4})k^{1/3}. \]

This implies that the subset \( S_4 \) satisfies the desired property. \( \square \)

4. Rarity of triangle induced graphs

We start by defining the function alluded to in Section 1.3, and establish several of its properties. Using the latter we then obtain the estimate in (11). We start by defining a function \( f : [0, \infty) \to (0, \infty) \) governed by the inequality in Lemma 3.3, by

\[ f(x) = \frac{\sqrt{2}}{3}(x + 1)^{3/2} + 3(x + 1). \] (48)

\( h \) will be essentially the inverse of \( f \), which we now formally define. Many of the forthcoming arguments establishing its properties are rather technical in nature, and can be safely skipped on first read.

Let \( h : (10, \infty) \to (1, \infty) \) be the inverse function of \( f \), restricted to \((10, \infty)\). The number 10 is chosen so that \( \inf_{y>10} h(y) > 1 \), which follows from the fact \( f(1) < 10 \). Since \( f \) is convex, \( h \) is concave on \((10, \infty)\). While our applications will only make use of the values of \( h(y) \) for integers \( y \), it would be convenient for expository reasons to regard \( h \) as a function of real numbers which we define by setting:

\[ h(y) = 1 \quad \text{for} \quad 2 \leq y \leq 10 \quad \text{and} \quad h(y) = 0 \quad \text{for} \quad y < 2. \]

We next make some observations about the function \( h \), that will be of use in the sequel. Notice that \( h \) is monotone, and since \( h \) is concave on \((10, \infty)\), the mean value theorem implies that for any \( x, y \geq 10 \),

\[ h(x) + h(y) \geq h(10) + h(x + y - 10). \] (49)

Since \( f(x) \geq \sqrt{2} \frac{x^{3/2}}{3} \) for large \( x \), and recalling \( h(y) \leq 1 \) for \( y \leq 10 \), setting \( a := (\frac{3}{\sqrt{2}})^{2/3} \), for all \( y \geq 1 \),

\[ h(y) \leq ay^{2/3}. \] (50)

Also, since \( f(x) \geq \sqrt{2} \frac{(x + 10)^{3/2}}{3} \) for large \( x \), for large enough \( y \),

\[ h(y) + 10 \leq ay^{2/3}. \] (51)

In addition, for any \( \kappa > 0 \), for large enough \( y > 0 \),

\[ h(y) > (a - \kappa)y^{2/3}. \] (52)

Note that for \( y > 10 \),

\[ h'(y) = f'(h(y))^{-1} = \left( \frac{\sqrt{2}}{2}(h(y) + 1)^{1/2} + 3 \right)^{-1}. \] (53)
Thus, by (50), (52) and the fact that \((\frac{\sqrt{2}}{2}a^{1/2})^{-1} = \frac{2}{3}a\) by our choice of \(a\), for any \(\kappa > 0\), for large enough \(y\),

\[
(a - \kappa)\frac{2}{3}y^{-1/3} \leq h'(y) \leq (a + \kappa)\frac{2}{3}y^{-1/3}.
\] (54)

Finally, let us examine the third derivative of \(h\). By taking the derivative of \(f(h(y)) = y\) three times, we get

\[
f'''(h(y))h'(y)^3 + 3f''(h(y))h'(y)h''(y) + f'(h(y))h'''(y) = 0.
\]

It is straightforward to verify that \(f', f'' > 0, f''' < 0\). Also, since \(h\) is increasing and concave on \((10, \infty)\), we have \(h'' < 0, h' > 0\) on this interval. Thus, for \(y > 10\),

\[
h''(y) > 0. \tag{55}
\]

We next record the following estimate.

**Lemma 4.1.** For any non-negative integers \(i\) and \(m\),

\[
i + h(m - 10i) \geq h(m - 10^5) \tag{56}
\]

and

\[
h(m - 10^5) \geq h(m) - h(10^6). \tag{57}
\]

The numbers \(10^5\) and \(10^6\) have no significant importance and they are chosen to be simply sufficiently large constants.

**Proof.** We first prove (56). When \(m \leq 10^5\), the RHS of (56) becomes 0, and when \(10i \leq 10^5\),

\[
h(m - 10i) \geq h(m - 10^5)\]

by the monotonicity of \(h\), so in those cases the result follows trivially. Thus it remains to prove the result for \(m > 10^5\) and \(10i > 10^5\).

We next divide the analysis into two cases: \(m \geq 10i + 10^4\) and \(m < 10i + 10^4\). In the case \(m \geq 10i + 10^4\), by the mean value theorem,

\[
h(m - 10^5) - h(m - 10i) \leq (10i - 10^5)h'(m - 10i) \leq 10ih'(10^4) \leq i
\]

(recall that \(h'\) is decreasing on \((10, \infty)\) since \(h\) is concave on that interval). The last inequality follows since \(f(600) < 10^4\) and hence \(h(10^4) > 600\) which along with (53) implies \(h'(10^4) < 0.1\). On the other hand, if \(m \leq 10i + 10^4\), then by (50), when \(m \geq 10^5\),

\[
i \geq \frac{1}{10}(m - 10^4) \geq am^{2/3} \geq h(m - 10^5).
\]

To prove (57), it suffices to consider the case \(m \geq 10^6\) since otherwise the LHS is non-negative while the RHS is non-positive. By the mean value theorem,

\[
h(m) - h(m - 10^5) \leq 10^5h'(m - 10^5) \leq 10^5h'(10^6 - 10^5) \leq h(10^6).
\]

The last inequality follows from (53) by observing that \(h(10^6), h(10^6 - 10^5) > 14000\).

\[\square\]

### 4.1. Estimates using \(h\)

Now given the function \(h\), we have the following estimate on the number of tree excess edges in a graph in terms of the number of triangles.

**Lemma 4.2.** For any connected graph \(G = (V, E)\), if \(\Delta \geq 1\),

\[
|E| - |V| \geq h(\Delta). \tag{58}
\]

**Proof.** The proof is a consequence of Lemma 3.3. In fact, recalling \(h : (10, \infty) \to (1, \infty)\) is the inverse of \(f\), (58) is true for \(\Delta > 10\). Also, for \(2 \leq \Delta \leq 10\), \(|E| - |V| \geq 1 = h(\Delta)\). This is because if
\(|E| - |V| = 0\), then the number of triangles in \(G\) is at most 1 (we obtain at most one triangle once one additional edge is added to a spanning tree, this is where the connectivity is crucially used). Finally, for \(\Delta = 1\), \(|E| - |V| \geq 0 = h(\Delta)\), since \(G\) cannot be a tree. \(\square\)

**Remark 4.3.** Observe that (58) implies that for any \(TIG = (V, E)\) spanned by \(\ell \geq 1\) triangles,

\[
|E| - |V| \geq h(\ell).
\]  

This is because \(\Delta \geq \ell\) and \(h\) is monotone.

We now come to the following key estimate. For a positive integer \(\ell\), let \(F_\ell\) be the event that there exists a triangle-induced subgraph (to be denoted \(TISG\)) spanned by \(\ell\) triangles. In the next lemma, we bound the probability of \(F_\ell\).

**Lemma 4.4.** Let \(\mu > 0\) be a constant. For any constant \(\eta > 0\), for sufficiently large \(n\) and any \(1 \leq \ell \leq \frac{n^{1/6}}{10}\),

\[
\mathbb{P}(F_\ell) \leq \left(\frac{1}{n}\right)^{(1-\eta)h(\ell)}.
\]

**Proof.** For positive integers \(\ell, v, e\), let \(F_{\ell,v,e}\) be the event that there exists a \(TISG\) spanned by \(\ell\) triangles with \(v\) vertices and \(e\) edges. The number of subgraphs with \(v\) vertices and \(e\) edges is bounded by \(\binom{n}{v} \left(\frac{\ell}{e}\right)^{\ell v^2 e}\) \(\leq n^v v^{2e}\). Since \(e \leq 5(e - v) + 5\) by Lemma 3.2,

\[
\mathbb{P}(F_{\ell,v,e}) \leq \frac{d^e}{n^e} n^v v^{2e} = \frac{d^e}{n^{e-v}} v^{2e} \leq \left(\frac{d^5 v^{10}}{n}\right)^{e-v} d^5 v^{10}.
\]  

By (59), we have

\[
\mathbb{P}(F_\ell) \leq \mathbb{P}(\bigcup_{e-v \geq h(\ell)} F_{\ell,v,e}).
\]

It turns out that a naive union bound applied to the RHS using (60) does not yield a sharp bound, unless \(\ell\) is small. For large \(\ell\), we use Lemma 3.4 instead. The estimates for large and small \(\ell\), where large and small are defined in terms of a parameter \(\xi\) to be defined shortly, are obtained separately in the following two steps respectively.

**Step 1.** Fixing a large constant \(b\) to be chosen later, we have

\[
\mathbb{P}(F_\ell) \leq \mathbb{P}(\bigcup_{e-v \leq 6b^2/3-1} F_{\ell,v,e}) + \mathbb{P}(\bigcup_{e-v \geq 6b^2/3-1} F_{\ell,v,e}).
\]  

In the second term, a naive union bound effectively works since there are many tree-excess edges. In order to control the first term, we first apply Lemma 3.4 which states that the event \(\cup_{e-v \leq 6b^2/3-1} F_{\ell,v,e}\) implies the existence of a dense subgraph accounting for most of the triangles. More precisely, for a sufficiently small constant \(\xi > 0\), set \(b = \frac{1}{2} \xi^{-1/2}\). Then, by Lemma 3.4, under the event \(\cup_{e-v \leq 6b^2/3-1} F_{\ell,v,e}\), there exists a subgraph \(G'\) such that

\[
|V(G')| \leq \xi^{-3/2} \ell^{1/3}, \quad |E(G')| \geq a(1 - 2\xi^{1/2} (2/3) \ell^{2/3}).
\]

Hence, by a union bound and the fact that the number of graphs with \(v\) vertices is bounded by \(2^v\), for any large enough \(\xi > 0\), for large enough \(\ell\) and \(n \geq 2\xi^{-3}\),

\[
\mathbb{P}(\bigcup_{e-v \leq 6b^2/3-1} F_{\ell,v,e}) \leq n^{\xi^{-3/2}} \ell^{1/3} 2^{\xi^{-3} \ell^{2/3}} \left(\frac{d}{n}\right) a(1 - 2\xi^{1/2} (2/3) \ell^{2/3}) \leq n^{\xi^{-3/2}} \ell^{1/3} 2^{\xi^{-3} \ell^{2/3}} \left(\frac{d}{n}\right) (1 - 2\xi^{1/2}) (2/3) h(\ell).
\]  

(50)
via graph decomposition into $1.1$-triangles in each $\ell$

Note that the number of summands is less than $(3\xi)^{\ell h(\ell)}$ bounds. For small $\xi > 0$, since $n \to \infty$, this quantity is bounded by $(\frac{1}{n})^{a/2}$, which holds for small $\xi > 0$ since $n \geq 2^{\xi^{-5}}$ (any exponent greater than 3.5 works, the exponent 5 is an arbitrary choice).

To bound the other term in (61), namely $P(\cup_{e \in v \geq b^{2/3}-1} F_{\ell, v, e})$ it suffices to bound

$$\sum_{e \in v + b^{2/3}} P(F_{\ell, v, e}).$$

Note that the number of summands is less than $(3\ell)^2$, since the number of edges and vertices are bounded by $3\ell$. Thus, using (60), for any $\kappa > 0$, for large enough $\ell$,

$$\sum_{e \geq v + b^{2/3}} P(F_{\ell, v, e}) \leq (3\ell)^2 \left( \frac{d^5(3\ell)^{10}}{n} \right) b^{2/3} - 1 d^5(3\ell)^{10} \leq C(10 + \kappa) b^{2/3} + 12 \left( \frac{1}{n} \right) b^{2/3} - 1. \quad (63)$$

Since $b \to \infty$ as $\xi \to 0$ and $\ell \leq n^{1/\gamma - \mu}$, this quantity is bounded by $(\frac{1}{n})^{a/2}$ for small enough $\xi, \kappa > 0$. Therefore, applying (62) and (63) to (61), for small enough $\xi > 0$, there exists $\ell_0(\xi) > 0$ such that for $\ell \geq \ell_0(\xi)$ and large enough $n$,

$$P(F_{\ell}) \leq \left( \frac{1}{n} \right)^{(1-4\xi^{1/2})h(\ell)}.$$

**Step 2.** For small $\ell$, a simple union bound using (60) suffices. For $1 \leq \ell \leq \ell_0(\xi)$ and sufficiently large $n$ (depending on $\xi$), there exists a constant $C = C(\xi) > 0$ such that

$$P(F_{\ell}) \leq \sum_{e \geq v + h(\ell)} P(F_{\ell, v, e}) \leq (3\ell)^2 \left( \frac{d^5(3\ell)^{10}}{n} \right) h(\ell) d^5(3\ell)^{10} \leq C \left( \frac{1}{n} \right)^{h(\ell)} \leq \left( \frac{1}{n} \right)^{(1-\xi^{1/2})h(\ell)}.$$

Finally we have all the ingredients in place to prove Theorem 1.1.

5. **Proof of Theorem 1.1 via graph decomposition into TISGs.**

Recall that due to the results of Section 2, it only remains to prove the corresponding upper bounds.

For $1 \leq \ell_1 \leq \cdots \leq \ell_m$, let $E_{\ell_1, \cdots, \ell_m}$ be the event that there exist vertex-disjoint TISGs spanned by $\ell_i$ triangles. We start by observing that the occurrence of the event $\{N \geq k\}$ implies the occurrence of the event $E_{\ell_1, \cdots, \ell_m}$ for some $m \geq 1, 1 \leq \ell_1 \leq \cdots \leq \ell_m$ and $\ell_1 + \cdots + \ell_m = k$. In fact, if there are $N$ triangles, then there are disjoint TISGs $H_1, \cdots, H_r$ in $G$ such that $H_i$ contains exactly $\ell'_i$ triangles with $1 \leq \ell'_1 \leq \cdots \leq \ell'_r$ and $\ell'_1 + \cdots + \ell'_r = N$. Since $N \geq k$, one can choose $\ell_i$ connected triangles in each $H_i$, such that $0 \leq \ell_i \leq \ell'_i$, $1 \leq \ell_1 \leq \cdots \leq \ell_m$, $m \leq r$, and $\ell_1 + \cdots + \ell_m = k$. 


Denoting by $H'_1, \cdots, H'_m$ TISGs spanned by each such collection of $\ell_1, \cdots, \ell_m$ triangles\(^3\) respectively, by considering the (vertex-disjoint) union of these TISGs, one can deduce that $\{N \geq k\}$ implies the occurrence of $E_{\ell_1, \cdots, \ell_m}$.

We now move towards upper bounding the probability $\mathbb{P}(E_{\ell_1, \cdots, \ell_m})$. We assume that $1 = \ell_1 = \cdots = \ell_j < \ell_{j+1} \leq \cdots \leq \ell_m$. By BK inequality,

$$
\mathbb{P}(E_{\ell_1, \cdots, \ell_m}) \leq \mathbb{P}(E_{1, \cdots, 1}) \mathbb{P}(F_{\ell_{j+1}}) \cdots \mathbb{P}(F_{\ell_m})
$$

(there are $j$ 1s in $E_{1, \cdots, 1}$). We now obtain a rather precise estimate of $\mathbb{P}(E_{1, \cdots, 1})$. The probability that there exist $j$ vertex-disjoint triangles is bounded by

$$
\left( \frac{d}{n} \right)^3 j \frac{1}{j!} \left( \begin{array}{c} n \ni 3 \end{array} \right) \left( \begin{array}{c} n-3 \ni 3 \end{array} \right) \cdots \left( \begin{array}{c} n-3j+3 \ni 3 \end{array} \right) \leq \frac{1}{j!} \left( \frac{d^3}{6} \right)^j.
$$

Note that it is crucial that we do not use the upper bound $\mathbb{P}(E_i)^j$ here, which would only give an upper bound $\alpha^j$ ($c > 0$ is a constant) on the probability $\mathbb{P}(E_{1, \cdots, 1})$. This along with Lemma 4.4 (which is applicable since $\ell_i \leq k \leq n^{1\alpha-\mu}$ for $i = j+1, \cdots, m$), yield, for any $\eta > 0$, for large enough $n$,

$$
\mathbb{P}(E_{1, \cdots, 1}) \mathbb{P}(F_{\ell_{j+1}}) \cdots \mathbb{P}(F_{\ell_m}) \leq \frac{1}{j!} \left( \frac{d^3}{6} \right)^j \left( \frac{1}{n} \right)^{1-\eta} (h(\ell_{j+1}) + \cdots + h(\ell_m)).
$$

Thus,

$$
\mathbb{P}(N \geq k) \leq \sum_{j=0}^{k} \sum_{\ell_{j+1} + \cdots + \ell_m = k-j} \frac{1}{j!} \left( \frac{d^3}{6} \right)^j \left( \frac{1}{n} \right)^{1-\eta} (h(\ell_{j+1}) + \cdots + h(\ell_m)).
$$

Recall from (49), that for any $x, y \geq 10$,

$$
h(x) + h(y) \geq h(10) + h(x+y-10) = 1 + h(x+y-10).
$$

We then apply this inequality repeatedly to $\ell_{j+1}, \cdots, \ell_m$ until there is at most one element greater than 10. Let us denote by $i$ the number of elements eventually less than or equal to 10. Since this procedure preserves the total sum to be equal to $k-j$, by the monotonicity of $h$, we obtain

$$
h(\ell_{j+1}) + \cdots + h(\ell_m) \geq i + h(k-j-10i).
$$

Thus, applying Lemma 4.1 and the fact that the number of partitions of $n$ is bounded by $e^{\alpha\sqrt{n}}$ for some constant $\alpha > 0$ ([4]), we get

$$
\sum_{\ell_{j+1} + \cdots + \ell_m = k-j} \frac{1}{j!} \left( \frac{d^3}{6} \right)^j \left( \frac{1}{n} \right)^{1-\eta} (h(\ell_{j+1}) + \cdots + h(\ell_m)) \leq \frac{1}{j!} \left( \frac{d^3}{6} \right)^j e^{\alpha\sqrt{k-j}} \left( \frac{1}{n} \right)^{1-\eta} h(k-j-10^5).
$$

Hence, the RHS of (66) is bounded by

$$
\sum_{j=0}^{k} \frac{1}{j!} \left( \frac{d^3}{6} \right)^j e^{\alpha\sqrt{k-j}} \left( \frac{1}{n} \right)^{1-\eta} h(k-j-10^5).
$$

We now bound this in the cases $k^{1/3} \log k < (a-\delta) \log n$ and $k^{1/3} \log k > (a+\delta) \log n$ where the $j \approx k$ and $j \approx 0$ are the dominating terms respectively.

**Case 1.** $k^{1/3} \log k < (a-\delta) \log n$.

---

\(^3\)We say that $H$ is the TISG spanned by $\ell$ triangles $T_1, \cdots, T_\ell$, if $H$ is connected and is the union of $T_1, \cdots, T_\ell$. 

We show that each term in (68) is bounded by the term with \( j = k \) up to a small multiplicative factor. Precisely, we claim that there exists \( a_j \) satisfying
\[
\frac{1}{j!} \left( \frac{d^3}{6} \right)^j e^{\alpha \sqrt{k-j}} \left( \frac{1}{n} \right)^{(1-\eta)h(k-j-10^5)} \leq e^{a_j} \frac{1}{k!} \left( \frac{d^3}{6} \right)^k ,
\] (69)
and further, for large enough \( n \),
\[
\sum_{j=0}^{k} e^{a_j} \leq C e^{C \log k} ,
\] (70)
where \( C = C(\delta, \eta) > 0 \) is a constant. Clearly this finishes the argument, as, by (70), the summation (68) is bounded by
\[
Ce^{C \log k} \frac{1}{k!} \left( \frac{d^3}{6} \right)^k \leq e^{-k \log k + ck}
\]
for some constant \( c > 0 \) depending only on \( d \) (not depending on \( \delta \)) and large enough \( k \).

Now to prove (69), by taking a ratio of the LHS and RHS of the same, recalling \( d_0 = \frac{d^3}{6} \) and using \( \frac{k!}{j!} \leq k^{k-j} \), we note that the quantity \( a_j \) can be chosen as
\[
a_j := (k-j) \log(k/d_0) + \alpha \sqrt{k-j} - (1-\eta)h(k-j-10^5) \log n.
\] (71)
Now, since by (50) and (52), \( h(y) \) is asymptotically \( ay^{2/3} \), there exists \( M = M(\eta, \delta) > 0 \) such that for \( k-j \geq M \),
\[
h(k-j-10^5) \geq (1-\eta)h(k-j) \geq (1-\eta)(a-0.5\delta)(k-j)^{2/3},
\]
which implies
\[
a_j \leq (k-j) \log(k/d_0) + \alpha \sqrt{k-j} - (1-\eta)^2(a-0.5\delta)(k-j)^{2/3} \log n
\]
\[
\leq (k-j)^{2/3} \left( (k-j)^{1/3} \log(k/d_0) + \alpha - (1-\eta)^2(a-0.5\delta) \log n \right).
\]
Since \( k^{1/3} \log k \leq (a-\delta) \log n \), for sufficiently small \( \eta > 0 \), for \( k-j \geq M \),
\[
a_j \leq -(k-j)^{2/3} \cdot 0.1 \delta \log n < 0.
\] (72)
Whereas, for \( k-j \leq M \), by the expression of \( a_j \) in (71),
\[
a_j \leq M \log(k/d_0) + \alpha \sqrt{M}.
\] (73)
Hence, by (72) and (73),
\[
\sum_{j=0}^{k} e^{a_j} = \sum_{j=0}^{k-M} e^{a_j} + \sum_{j=k-M+1}^{k} e^{a_j} \leq k + Me^{M \log(k/d_0) + \alpha \sqrt{M}} \leq Ce^{C \log k},
\]
which implies (70).

**Case 2.** \( k^{1/3} \log k > (a+\delta) \log n \).

By (57), the summation (68) is bounded by
\[
\sum_{j=0}^{k} \frac{1}{j!} \left( \frac{d^3}{6} \right)^j e^{\alpha \sqrt{k-j}} \left( \frac{1}{n} \right)^{(1-\eta)h(k-j-10^5)} \leq e^{a_j} \frac{1}{k!} \left( \frac{d^3}{6} \right)^k ,
\] (74)
Recalling $d_1 = \frac{e^{\alpha^3}}{\alpha^3}$ and by Stirling’s formula (18), this is bounded by
\[
\sum_{j=0}^{k} \left(\frac{d_1}{j}\right)^j e^{\alpha \sqrt{k}} \left(\frac{1}{n}\right)^{(1-\eta)(h(k-j)-h(10^6))}
\] (75)
(we set $(\frac{d_1}{\alpha})^0 := 1$). We claim that for small enough constant $\rho > 0$, for sufficiently large $n$, each term above can be bounded by
\[
Ce^{\alpha \sqrt{k}} \left(\frac{1}{n}\right)^{(1-\eta)(h(k-\rho k)-h(10^6))}.
\] (76)

We first conclude the proof assuming this claim. Take $\rho > 0$ small enough so that for large $k$,
\[
(1 - \eta)(h(k-\rho k) - h(10^6)) > (1 - 2\eta)h(k).
\]
This is possible since $h(y) \approx ay^{2/3}$ (see (50) and (52)). Hence, (75) is bounded by
\[
Ck \cdot e^{\alpha \sqrt{k}} \left(\frac{1}{n}\right)^{(1-\eta)(h(k-\rho k)-h(10^6))} \leq Ce^{\alpha \sqrt{k}} \left(\frac{1}{n}\right)^{(1-2\eta)h(k)} \leq Ce^{h(k)} \left(\frac{1}{n}\right)^{(1-2\eta)h(k)} \leq \left(\frac{1}{n}\right)^{(1-3\eta)h(k)},
\]
which concludes the proof. Therefore, it suffices to verify the claim, i.e., each term in (75) is bounded by (76). This is equivalent to
\[
(1 - \eta)(h(k-\rho k) - h(k-j)) \log n - j \log(j/d_1) \leq \log C.
\] (77)
This immediately holds for $j < \rho k$, since in this case there exists $C_1 = C_1(d_1) > 0$ such that
\[
\text{LHS in (77)} \leq -j \log(j/d_1) \leq C_1.
\]
Thus, it only remains to verify the same for $j \geq \rho k$. Towards this end, we regard the LHS in (77) as a function of $j$. Let
\[
g(x) := -(1 - \eta)h(k-x) \log n + (1 - \eta)h(k-\rho k) \log n - x \log(x/d_1).
\]
Then,
\[
g'(x) = (1 - \eta)h'(k-x) \log n - \log(x/d_1) - 1.
\] (78)
Recall that by the upper bound for $h'$ in (54), for large enough $k$, $h'(k-\rho k) < (a + \delta)\frac{1}{3} \sqrt{\frac{k}{k-\rho k}}$. Hence, for small $\rho > 0$ and large enough $k$,
\[
g'(\rho k) = (1 - \eta)h'(k-\rho k) \log n - \log(\rho k/d_1) - 1
\leq (1 - \eta)(a + \delta)\frac{1}{3} \sqrt{\frac{k}{k-\rho k}} \log n - \log(\rho k/d_1) - 1
\leq \frac{2}{3} \left(\frac{k}{k-\rho k}\right)^{1/3} \log k - \log(\rho k/d_1) - 1
= \left(\frac{2}{3} \left(\frac{1}{k^2}\right)^{1/3} - 1\right) \log k - \log(\rho k/d_1) - 1 < 0.
\]
Note that we used the condition $(a + \delta) \log n < k^{1/3} \log k$ in the second inequality above. Using this fact, we next show that
\[
\max_{x \in [\rho k, k-10]} g(x) = \max(g(\rho k), g(k-10)).
\] (79)
By (55), $h'(k-x)$ is convex on $(\rho k, k-10)$, which also implies the convexity of $g'(x)$ (see the expression for $g'$ in (78)). Since $g'(\rho k) < 0$, there are only two possible scenarios: either $g' < 0$ on
(ρk, k − 10), or g′ < 0 on (ρk, x₀) and g′ > 0 on (x₀, k − 10) for some x₀ ∈ (ρk, k − 10). This implies (79).

Thus, in order to show (77), it suffices to consider the two cases, j ≥ k − 10 and j = ρk. In the case j = k − z with 0 ≤ z ≤ 10, the LHS of (77) can be bounded as

\[ (1 - \eta)(h(k - \rho k) - h(k - j)) \log n - j \log(j/d_1) \leq (1 - \eta)h(k - \rho k) \log n - (k - 10) \log((k - 10)/d_1) \leq h(k) \log n - k \log k + Ck \leq ak^{2/3} \log n - k \log k + Ck \leq 0. \]

Here, we used the condition k^{1/3} \log k > (a + \delta) \log n to deduce the last inequality. Whereas, in the case j = ρk, (77) immediately holds, completing the verification of the claim.

We now proceed towards proving Theorem 1.2.

6. Proof of Theorem 1.2

We start by recalling the events from Theorem 1.2:

\[ D_\varepsilon := \{ \text{There exist at least } (1 - \varepsilon)k \text{ vertex-disjoint triangles} \} \]

and

\[ C_\varepsilon := \{ \text{There exists } V' \subseteq V_n \text{ such that } |V'| \leq (1 + \varepsilon)6^{1/3}k^{1/3}, \Delta(V') \geq (1 - \varepsilon)k \}. \]

We now define an auxiliary event \( T_\varepsilon \), as follows:

\[ T_\varepsilon := \{ \text{There exists a triangle-induced subgraph spanned by at least } (1 - 2\varepsilon)k \text{ many triangles} \}. \]

Note that \( \Delta(G) \leq \frac{1}{6}|G|^3 \) for any graph G, and it is not too difficult to show that an “almost” equality holds if and only if G is “close” to a clique for appropriate formulation of the terms in quotes. Hence, \( C_\varepsilon \) implies that \( V' \) looks like a clique containing almost k many triangles.

The proof of Theorem 1.2 has broadly two parts. The first part is proving that conditional on the event \( \{ N \geq k \} \), the event \( D_\varepsilon \cup T_\varepsilon \) holds with high probability (see Proposition 6.1). The next part proves that conditionally on \( T_\varepsilon \cap \{ N \geq k \} \), the event \( C_{12\varepsilon/4} \) holds with high probability (see Proposition 6.2). By these two parts, with high probability conditionally on \( \{ N \geq k \} \), the event \( D_\varepsilon \cup C_{12\varepsilon/4} \) holds. By the monotonicity of the events \( D_\varepsilon \) and \( C_\varepsilon \), in \( \varepsilon \), one can conclude that conditionally on \( \{ N \geq k \} \), \( D_\varepsilon \cup C_\varepsilon \) holds with high probability.

Statements (5) and (6) follow from Theorem 1.1, along with the fact that for small enough \( \varepsilon > 0 \), it is unlikely that \( D_\varepsilon \cap C_\varepsilon \) happens given \( \{ N \geq k \} \). Validity of (5) and (6) for any \( \varepsilon > 0 \) follows again by the monotonicity of the events \( D_\varepsilon \) and \( C_\varepsilon \) in \( \varepsilon \).

**Proposition 6.1.** The following holds for sufficiently small \( \varepsilon > 0 \). For any k such that \( k \to \infty \) and \( k \leq n^{1/3 - \mu} \),

\[ \lim_{n \to \infty} P(D_\varepsilon \cup T_\varepsilon | N \geq k) = 1. \]

**Proposition 6.2.** The following holds for sufficiently small \( \varepsilon > 0 \). For any k such that \( k \to \infty \) and \( k \leq n^{1/3 - \mu} \),

\[ \lim_{n \to \infty} P(T_\varepsilon \cap C_{12\varepsilon/4} | N \geq k) = 0. \]

We now finish the proof of Theorem 1.2 first.
Proof of Theorem 1.2. By the monotonicity of the events $D_\varepsilon$ and $C_\varepsilon$ in $\varepsilon$, it suffices to prove them for small enough $\varepsilon$, an assumption that is particularly convenient for some of our arguments. By the above two propositions, for any small enough $\varepsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}(D_\varepsilon \cup C_{12\varepsilon/4}|N \geq k) = 1.$$  

Observing that $D_\varepsilon$ and $C_\varepsilon$ are decreasing as $\varepsilon \to 0$, by the arbitrariness of $\varepsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}(D_\varepsilon \cup C_\varepsilon|N \geq k) = 1. \quad (80)$$

To obtain (5) and (6), we first verify that for sufficiently small $\varepsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}(D_\varepsilon \cap C_\varepsilon|N \geq k) = 0. \quad (81)$$

By (22), $C_\varepsilon$ implies the existence of a subset $V'$ such that $|V'| \leq (1 + \varepsilon)6^{1/3}k^{1/3} =: v$ and $|E(V')| \geq a(1 - \varepsilon)^{2/3}k^{2/3} =: e$ (where, recall that the value of $a$ is specified to be $(\frac{3}{\sqrt{2}})^{2/3}$). By a first moment bound, there is $\gamma = \gamma(\varepsilon) > 0$ with $\lim_{\varepsilon \to 0} \gamma = 0$ such that for large enough $k$, the probability that such a subgraph exists is bounded by

$$\left(\frac{n}{v}\right)^e \left(\frac{\log n}{e}\right) \leq n^2e^{2e} \left(\frac{d}{n}\right)^e \leq n^{-(a - \gamma)k^{2/3}}. \quad (82)$$

Recall that $D_\varepsilon$ implies the existence of at least $(1 - \varepsilon)k$ disjoint triangles, and under the event $C_\varepsilon$, there are at most $|V'| = O(k^{1/3})$ vertex-disjoint triangles sharing some vertex in $V'$. Hence, for large enough $k$, $D_\varepsilon \cap C_\varepsilon$ implies the disjoint occurrence of the subgraph $V'$ and at least $(1 - 2\varepsilon)k$ disjoint triangles. Thus, by Remark 2.2 and BK inequality,

$$\mathbb{P}(D_\varepsilon \cap C_\varepsilon) \leq n^{-(a - \gamma)k^{2/3}} \cdot e^{-(1 - 2\varepsilon)k \log k + 2c'k} =: S.$$ 

Relying on Proposition 2.4, for small enough $\varepsilon > 0$, it follows that the quantity $S$ is negligible compared to $\mathbb{P}(N \geq k)$. To see this, observe that $S = o(1)e^{-k \log k - c k}$ is equivalent to

$$-(a - \gamma)k^{2/3} \log n + 2\varepsilon k \log k + 2c'k \to -\infty,$$

which holds for small enough $\varepsilon > 0$ when $k^{1/3} \log k \leq a \log n$. Further, $S = o(1)n^{-(1 + \varepsilon)ak^{2/3}}$ is equivalent to

$$(\gamma + a\varepsilon)k^{2/3} \log n - (1 - 2\varepsilon)k \log k + 2c'k \to -\infty,$$

which holds for small enough $\varepsilon > 0$ when $k^{1/3} \log k \geq a \log n$. Therefore, we have (81).

To conclude the proof, note that when $k^{1/3} \log k \leq (a - \delta) \log n$, by Lemma 2.1 and (82),

$$\lim_{n \to \infty} \mathbb{P}(C_\varepsilon|N \geq k) = 0,$$

whereas, in the case $k^{1/3} \log k \geq (a + \delta) \log n$, by Lemma 2.3 and Remark 2.2,

$$\lim_{n \to \infty} \mathbb{P}(D_\varepsilon|N \geq k) = 0.$$ 

This, along with (80) and (81) finishes the proof of (5) and (6) for sufficiently small $\varepsilon > 0$. Since events $D_\varepsilon$ and $C_\varepsilon$ are increasing in $\varepsilon$, this concludes the proof of (5) and (6) for any $\varepsilon > 0$. $\square$

It remains to prove the propositions which we turn to next.

Proof of Proposition 6.1. We show that

$$\mathbb{P}(D_\varepsilon^c \cap T_\varepsilon^c \cap \{N \geq k\}) = o(1)\mathbb{P}(N \geq k).$$
The event $D^c_i \cap T^c_i \cap \{N \geq k\}$ implies the occurrence of the event $E_{\ell_1, \ldots, \ell_m}$ for some $1 \leq \ell_1 \leq \cdots \leq \ell_m \leq (1-2\varepsilon)k$ with $\ell_1 + \cdots + \ell_m = k$ such that the number of 1s among the $\ell_j$s, denoted by $j$, is less than $(1-\varepsilon)k$.

Thus, for any $\eta > 0$, for sufficiently large $n$,

$$
\mathbb{P}(D^c_i \cap T^c_i \cap \{N \geq k\}) \leq \sum_{j=0}^{(1-\varepsilon)k} \sum_{\ell_{j+1} + \cdots + \ell_m = k-j \atop 2 \leq \ell_{j+1} \leq \cdots \leq \ell_m \leq (1-2\varepsilon)k} \mathbb{P}(E_{\ell_1, \ldots, \ell_m})
$$

By (65), the above sum can be bounded by

$$
\sum_{j=0}^{(1-\varepsilon)k} \frac{1}{j!} \left( \frac{d^3}{6} \right) j \left( \frac{1}{n} \right)^{(1-\eta)(h(\ell_{j+1}) + \cdots + h(\ell_m))} = o(1) \mathbb{P}(N \geq k).
$$

By (67), the above sum can be bounded by

$$
\sum_{j=\varepsilon k}^{(1-\varepsilon)k} \frac{1}{j!} \left( \frac{d^3}{6} \right) j \left( \frac{1}{n} \right)^{(1-\eta)h(k-j-10^5)}
$$

Recalling $d_1 = \frac{e^{d_1}}{6}$ and by Stirling’s formula (18), each term above is bounded by

$$
d_1^j e^{\alpha \sqrt{k-j}} e^{-(j \log j + (1-\eta)h(k-j-10^5)) \log n}.
$$

Suppose that $j =zk$ with $\varepsilon \leq z \leq 1 - \varepsilon$. Recalling that $h(k-j) \approx a(k-j)^{2/3}$ (see (50) and (52)), for any $\iota > 0$, for large enough $k$,

$$
h(k-j-10^5) \geq (1-0.1\iota)h(k-j) \geq (a-\iota)(k-j)^{2/3} = (a-\iota)(1-z)^{2/3} k^{2/3}.
$$

This implies that for some constant $c_1 > 0$,

$$
j \log j + (1-\eta)h(k-j-10^5) \log n \geq zk \log z + zk \log k + (1-\eta)(a-\iota)(1-z)^{2/3} k^{2/3} \log n
$$

$$
\geq -c_1 k + \left( z + (1-\eta)(a-\iota) \frac{\log n}{k^{1/3} \log k} \right)(1-z)^{2/3} k \log k.
$$

In order to lower bound the second term, let us examine the concave function $z \mapsto z + \chi(1-z)^{2/3}$ on $[0, 1]$, parametrized by $\chi > 0$. We will show that if $\varepsilon > 0$ is small enough, then for any $\varepsilon \leq z \leq 1 - \varepsilon$ and $\chi \in (0, 1)$,

$$
z + \chi(1-z)^{2/3} \geq (1+0.01\varepsilon)^{2/3} \chi.
$$

Let us first conclude the proof of Step 1 assuming (87). Using this, there exists $\gamma = \gamma(\varepsilon) > 0$ such that the following holds for small enough $\iota, \eta, \varepsilon > 0$ and any $\varepsilon \leq z \leq 1 - \varepsilon$. 

\begin{itemize}
\item when $k^{1/3} \log k \leq a \log n$:
\[ z + (1 - \eta) (a - t) \frac{\log n}{k^{1/3} \log k} (1 - z)^{2/3} \geq 1 + \gamma. \]
To see this, note that for any $\kappa > 0$, $(1 - \eta) (a - t) \frac{\log n}{k^{1/3} \log k} \geq 1 - \kappa$ for small enough $\eta, t > 0$. Thus,
\[ z + (1 - \eta) (a - t) \frac{\log n}{k^{1/3} \log k} (1 - z)^{2/3} \geq z + (1 - \kappa) (1 - z)^{2/3} \]
\[ \geq (1 + 0.01 \varepsilon)^{2/3} (1 - \kappa) \geq 1 + 0.001 \varepsilon \]
for small enough $\kappa > 0$, where the second last inequality follows by setting $\chi = 1 - \kappa$ in (87).
\end{itemize}

Therefore, for large enough $k$, using (86), the quantity (85) in this case is bounded by
\begin{equation}
kd_1^k e^\alpha \sqrt{K} e^{c_1 k} e^{-(1+\gamma)k \log k} \leq e^{-(1+\gamma/2)k \log k}. \tag{88}
\end{equation}

\begin{itemize}
\item when $k^{1/3} \log k \geq a \log n$: setting $\chi = (1 - \eta) (a - t) \frac{\log n}{k^{1/3} \log k} \in (0, 1)$ in (87), taking $\gamma > 0$ to be $1 + 2\gamma = (1 + 0.01 \varepsilon)^{2/3}$,
\[ z + (1 - \eta) (a - t) \frac{\log n}{k^{1/3} \log k} (1 - z)^{2/3} \geq (1 + 2\gamma) (1 - \eta) (a - t) \frac{\log n}{k^{1/3} \log k} \]
\[ \geq (1 + \gamma) a \frac{\log n}{k^{1/3} \log k}. \]
\end{itemize}
Hence, for large enough $k$, in this case the quantity (85) is bounded by
\begin{equation}
kd_1^k e^\alpha \sqrt{K} e^{c_1 k} e^{-a(1+\gamma)k^{2/3} \log n} \leq n^{-a(1+\gamma/2)k^{2/3}}. \tag{89}
\end{equation}

By Proposition 2.4, quantities (88) and (89) are negligible compared to $\mathbb{P}(N \geq k)$ for large enough $k$.

Thus to conclude Step 1 it remains to prove (87). By concavity of $z \mapsto z + \chi (1 - z)^{2/3}$, it suffices to consider only two cases $z = \varepsilon$ and $z = 1 - \varepsilon$. In the case $z = \varepsilon$, using $\chi < 1$ and mean value theorem, for small enough $\varepsilon > 0$,
\[ \chi((1 + 0.01 \varepsilon)^{2/3} - (1 - \varepsilon)^{2/3}) \leq 1.01 \varepsilon \cdot \frac{2}{3} \cdot 1.1 \leq \varepsilon. \]
which after rearranging yields (87). In the case $z = 1 - \varepsilon$, since $\chi \in (0, 1)$ it suffices to prove $(1 + 0.01 \varepsilon)^{2/3} - \varepsilon^{2/3} \leq 1 - \varepsilon$. By mean value theorem, for small enough $\varepsilon > 0$,
\[ (1 + 0.01 \varepsilon)^{2/3} - 1 \leq 0.01 \varepsilon \cdot \frac{2}{3} \cdot 1.1 \leq \left( \frac{1}{\varepsilon^{2/3}} - 1 \right) \varepsilon = \varepsilon^{2/3} - \varepsilon. \]

\textbf{Step 2.} We next show that
\begin{equation}
\sum_{j=0}^{\ell k} \sum_{2 \leq j_1 + \cdots + j_m = k - j} \frac{1}{j!} \left( \frac{d^3}{6} \right) j \left( \frac{1}{n} \right) (1 - \eta)^{h(\ell_1 + \cdots + h(\ell_m))} = o(1) \mathbb{P}(N \geq k). \tag{90}
\end{equation}

We claim that for sufficiently small $\varepsilon > 0$, for large enough $k$ and any $\ell_{j+1} + \cdots + \ell_m = k - j, \ell_{j+1} \leq \cdots \leq \ell_m < (1 - 2\varepsilon)k$ with $j \leq \varepsilon k$,
\begin{equation}
L(\ell_{j+1} + \cdots + \ell_m) \geq \ell^2 k^{2/3}. \tag{91}
\end{equation}
We first conclude the proof of (90) assuming this claim. Using the above along with Stirling’s formula (18), for sufficiently small $\eta > 0$, the quantity (90) is bounded by
\[
\sum_{j=0}^{\varepsilon k} \frac{d^j}{j!} e^{\alpha \sqrt{k}} \left( \frac{1}{n} \right)^{(1-\eta)(h(k)+3\varepsilon^2k^{2/3})} \leq C e^{\alpha \sqrt{k}} \left( \frac{1}{n} \right)^{h(k)+2\varepsilon^2k^{2/3}} \leq \left( \frac{1}{n} \right)^{h(k)+\varepsilon^2k^{2/3}},
\]
(92)
where the first inequality uses $\frac{d^j}{j!} \leq C$ for any $j \geq 0$ (we set $0^0 := 1$). Thus, along with the lower bound $\mathbb{P}(N \geq k)$ (Proposition 2.4) concludes the proof of (90).

Now, let us verify the claim (91). For any $x, y \in [10, (1-2\varepsilon)k]$, by (49), it holds that
\[
 h(x) + h(y) \geq h(10) + h(x + y - 10), \quad \text{if} \quad x + y - 10 \leq (1 - 2\varepsilon)k
\]
and by a similar reasoning we also have that
\[
 h(x) + h(y) \geq h(x + y - (1 - 2\varepsilon)k) + h((1 - 2\varepsilon)k), \quad \text{if} \quad x + y - 10 \geq (1 - 2\varepsilon)k.
\]
We apply this repeatedly to \(\ell_{j+1}, \ldots, \ell_m\) until there is at most one element strictly between 10 and \((1-2\varepsilon)k\). At the end of this procedure, let $i$ be the number of elements less than or equal to 10. Then, we obtain the following two cases, depending on whether \((1-2\varepsilon)k\) appears as a term or not (since this procedure preserves the summation, the term \((1-2\varepsilon)k\) appears at most once for small $\varepsilon > 0$):

Case 1. \(h((1-2\varepsilon)k)\) does not appear:
\[
h(\ell_{j+1}) + \cdots + h(\ell_m) \geq i + h(b_1)
\]
(93)
with \((1-2\varepsilon)k \geq b_1 \geq k - j - 10i\). The lower bound on $b_1$ is a consequence of the facts that the above procedure preserves the summation to be \(k - j\) and that $h$ is monotone.

Case 2. \(h((1-2\varepsilon)k)\) appears:
\[
h(\ell_{j+1}) + \cdots + h(\ell_m) \geq i + h(b_2) + h((1-2\varepsilon)k)
\]
(94)
with \(b_2 \geq k - j - (1-2\varepsilon)k - 10i = 2\varepsilon k - j - 10i\). The reason for the lower bound on $b_2$ is similar as above, along with a fact that the term \((1-2\varepsilon)k\) is obtained at the end of procedure.

We first consider the former case. Since \((1-2\varepsilon)k \geq b_1 \geq k - j - 10i\) and \(j \leq \varepsilon k\), we have \(10i \geq \varepsilon k\). Thus, for large enough $k$,
\[
i + h(b_1) \geq i + \frac{1}{10} \varepsilon k \geq h(k) + 3\varepsilon^2 k^{2/3} \quad \text{(50)}
\]
Note that the last inequality holds since LHS and RHS are of order $k$ and $k^{2/3}$ respectively.

We next consider the latter case (94). Since \(2\varepsilon k - j \geq 0\) (recall that $j \leq \varepsilon k$), by Lemma 4.1,
\[
i + h(b_2) + h((1-2\varepsilon)k) \geq i + h(2\varepsilon k - j - 10i) + h((1-2\varepsilon)k) \\
\geq h(2\varepsilon k - j) - h(10^6) + h((1-2\varepsilon)k).
\]
Thus, in order to verify (91), it suffices to show that for sufficiently small $\varepsilon > 0$, for large enough $k$,
\[
h(2\varepsilon k - j) + h((1-2\varepsilon)k) \geq h(k) + 3\varepsilon^2 k^{2/3} + h(10^6).
\]
(95)
In fact, by mean value theorem, for large enough $k$,
\[
h(k) - h((1-2\varepsilon)k) \leq 2\varepsilon k h'((1-2\varepsilon)k) \leq 2\varepsilon k \cdot 1.01 \cdot \frac{2}{3} \cdot (k - 2\varepsilon k)^{-1/3}.
\]
(54)
Since \( j \leq \varepsilon k \), by (52), for large enough \( k \),
\[
h(2\varepsilon k - j) \geq h(\varepsilon k) \geq \frac{1}{2}a(\varepsilon k)^{2/3} + h(10^6).
\]
By the above two displays, to prove (95), it suffices to check
\[
\frac{1}{2}a(\varepsilon k)^{2/3} \geq 2\varepsilon k \cdot 1.01 \cdot \frac{2}{3}(k - 2\varepsilon k)^{-1/3} + 3\varepsilon^2 k^{2/3}.
\]
Since the coefficients of \( k^{2/3} \) in LHS and RHS are of order \( \varepsilon^{2/3} \) and \( \varepsilon \) as \( \varepsilon \to 0 \) respectively, this inequality holds for sufficiently small \( \varepsilon > 0 \).
\[
\square
\]

We now move on to the proof of Proposition 6.2.

**Proof of Proposition 6.2.** Under the event \( \mathcal{T}_\varepsilon \), let \( G' \) be a TISG spanned by \( \hat{k} := [(1-2\varepsilon)k] \) triangles. We will prove that with high probability conditionally on \( \{N \geq k\} \), \( G' \) contains a subgraph which looks like a clique and contains at least \((1-12\varepsilon^1/4)k\) triangles. The argument has three parts (carried out in the sequel in Steps 1, 2, and 3 respectively). First, we show that with high probability conditioned on \( \{N \geq k\} \), the number of tree-excess edges in \( G' \) is at most of order \( k^{2/3} \) (Step 1). Hence by Lemma 3.4, \( G' \) contains a ‘dense’ subgraph \( G'' \) of size \( O(k^{1/3}) \) which contains close to \( k \) triangles. Next, we deduce that with high probability conditioned on \( \{N \geq k\} \), any subgraph of size \( O(k^{1/3}) \) has ‘almost’ less than \( h(k) \) edges (Step 2). This implies that \( G'' \) almost attains the equality in (22) and thus as an application of Proposition 3.6, \( G'' \) contains a subgraph which is close to a clique and contains almost \( k \) triangles. This is done in Step 3.

**Step 1.** We prove that there exists a large constant \( b > 0 \) such that following holds: With high probability conditionally on \( \{N \geq k\} \), any TISG \( H \) spanned by \( \hat{k} \) triangles satisfies
\[
|E(H)| - |V(H)| \leq b\hat{k}^{2/3} - 1. \tag{96}
\]
As in the proof of Lemma 4.4, let \( F_{k,v,e}^i \) be an event that there exists a TISG spanned by \( \hat{k} \) triangles with \( v \) vertices and \( e \) edges. Then,
\[
\mathbb{P}(\exists \text{TISG } H \text{ spanned by } \hat{k} \text{ triangles with } |E(H)| - |V(H)| \geq b\hat{k}^{2/3} - 1) \leq \sum_{\varepsilon \geq \varepsilon + b\hat{k}^{2/3} - 1} \mathbb{P}(F_{k,v,e}^i) \leq \sum_{\varepsilon \geq \varepsilon + b\hat{k}^{2/3} - 1} \frac{\left(d^5(3\hat{k})^{10}\hat{k}^{2/3} - 1\right)}{n} \cdot \left(d^5(3\hat{k})^{10}\hat{k}^{2/3} - 1\right)
\leq (3\hat{k})^{12} \cdot \left(C\hat{k}\right)^{10\hat{k}^{2/3} - 1} \cdot \left(C\hat{k}\right)^{10\hat{k}^{2/3} - 1} \leq C\hat{k}^{12} \cdot \left(C\hat{k}\right)^{10\hat{k}^{2/3} - 1} \leq C\hat{k}^{12} \cdot \left(C\hat{k}\right)^{10\hat{k}^{2/3} - 1} \leq C\hat{k}^{12} \cdot \left(C\hat{k}\right)^{10\hat{k}^{2/3} - 1}.
\]
We show that this is bounded by \( (\frac{1}{n})^{h(k) + \varepsilon bk^{2/3}} \) for small enough \( \xi, \varepsilon > 0 \). Once we show this, by the lower bound on \( \mathbb{P}(N \geq k) \) in Lemma 2.3, (96) follows.

Rearranging, we need to establish
\[
C^{10\hat{k}^{2/3} + 1} \cdot \hat{k}^{12 + 10\hat{k}^{2/3}} \leq n^{b\hat{k}^{2/3} - 1 - h(k) - \varepsilon bk^{2/3}}.
\tag{97}
\]
By the condition \( k \leq n^{\frac{1}{\mu^2}} \), for large enough \( n \), the LHS is bounded by
\[
\frac{1}{n^{b\mu k^{2/3}} \cdot n^{(1-10\mu)bk^{2/3} + 2}} = n^{(1-9\mu)bk^{2/3} + 2}.
\]
Whereas, for small enough \( \varepsilon > 0 \) and large enough \( b > 0 \), the exponent of \( n \) in the RHS of (97) is
\[
bk^{2/3} - 1 - h(k) - \varepsilon bk^{2/3} \geq ((1 - 2\varepsilon)^{2/3} - \varepsilon)bk^{2/3} - ak^{2/3} - 1 \geq ((1 - \mu)b - a)k^{2/3} > (1 - 9\mu)bk^{2/3} + 2,
\]
where the last inequality uses the fact that \( (1 - \mu)b - a > (1 - 9\mu)b \) for large enough \( b \). We thus obtain (97).

**Step 2.** For any constants \( c_1, c_2 > 0 \), with high probability conditionally on \( \{N \geq k\} \), any subgraph \( H \) (need not be a TISG or a connected graph) with \( |V(H)| < c_1k^{1/3} \) satisfies
\[
|E(H)| < h(k) + c_2k^{2/3} \tag{98}
\]
for large enough \( n \geq n_0(c_1, c_2) \) and \( k \geq k_0(c_1, c_2) \) (a useful consideration is the case when \( c_1 \) is large and \( c_2 \) is small). In fact, for large enough \( n \) and \( k \), the probability that there exists a subgraph \( H \) with \( |V(H)| < c_1k^{1/3} \) and \( |E(H)| \geq h(k) + c_2k^{2/3} \) is bounded by
\[
n^{c_1k^{1/3}/2}2^{-c_2k^{2/3}}\left(\frac{d}{n}\right)^{h(k)+c_2k^{2/3}} \leq \left(\frac{1}{n}\right)^{h(k)+\frac{c_2}{2}k^{2/3}}.
\]
By the lower bound on \( \mathbb{P}(N \geq k) \) in Lemma 2.3, (98) follows.

**Step 3.** Recall that \( G' \) is a TISG spanned by \( \tilde{k} \) triangles. By (96), for a large enough constant \( b > 0 \), with high probability conditionally on \( \{N \geq k\} \),
\[
|E(G')| - |V(G')| \leq bk^{2/3} - 1.
\]
Thus, setting \( \xi > 0 \) to be such that \( b = k^{1/2}\xi^{-1/2} \), \( G' \) satisfies the hypotheses of Lemma 3.4, on the application of which, we conclude that for sufficiently large \( k \), \( G' \) contains a dense subgraph \( G'' \) such that
\[
|V(G'')| \leq \xi^{-3/2}\tilde{k}^{1/3} \leq \xi^{-3/2}(1 - \varepsilon)^{1/3}k^{1/3} \tag{99}
\]
and (by taking large enough \( b \), i.e. small enough \( \xi > 0 \))
\[
\Delta(G'') \geq (1 - 2\xi^{1/2})\tilde{k} \geq (1 - 3\varepsilon)k. \tag{100}
\]
Note that Lemma 3.4 also implies a lower bound for \( |E(G'')| \) from (26). However for our purposes, we will instead be interested in a sharp upper bound for \( |E(G'')| \), which along with (100) will be used to finish the argument. Towards this, by (98) and (99), with high probability (conditionally),
\[
|E(G'')| \leq h(k) + a\varepsilon k^{2/3} \leq a(1 + \varepsilon)k^{2/3}. \tag{101}
\]
Thus (100) and (101) imply that \( G'' \) almost attains an equality in Lemma 3.1. This allows us to invoke Proposition 3.6. Applying the latter with \( 6k \) and \( 3\varepsilon \) in place of \( k \) and \( \varepsilon \) respectively, allows us to conclude that \( G'' \) contains a subgraph of size at most \( (1 + 12\varepsilon^{1/4})6^{1/3}k^{1/3} \) with at least \( (1 - 12\varepsilon^{1/4})k \) triangles. This concludes the proof. □

**References**

[1] Noga Alon. On the number of subgraphs of prescribed type of graphs with a given number of edges. *Israel Journal of Mathematics*, 38(1):116–130, 1981.

[2] Luisa Andreis, Wolfgang König, Heide Langhammer, and Robert I. A. Patterson. A large-deviations principle for all the components in a sparse inhomogeneous random graph, 2021.
[3] Luisa Andreis, Wolfgang König, and Robert I. A. Patterson. A large-deviations principle for all the cluster sizes of a sparse erdős–reýny graph. *Random Structures & Algorithms*, 59(4):522–553, 2021.

[4] George E. Andrews. *The Theory of Partitions*. Cambridge University Press, 1984.

[5] Fanny Augeri. Nonlinear large deviation bounds with applications to traces of wigner matrices and cycles counts in Erdős-Rényi graphs. *Annals of Probability, to appear*, 2020.

[6] Bhaswar B Bhattacharya, Persi Diaconis, and Sumit Mukherjee. Universal limit theorems in graph coloring problems with connections to extremal combinatorics. *The Annals of Applied Probability*, 27(1):337–394, 2017.

[7] Bhaswar B Bhattacharya, Shirshendu Ganguly, Eyal Lubetzky, and Yufei Zhao. Upper tails and independence polynomials in random graphs. *Advances in Mathematics*, 319:313–347, 2017.

[8] Béla Bollobás. Threshold functions for small subgraphs. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 90, pages 197–206. Cambridge University Press, 1981.

[9] Béla Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.

[10] Suman Chakraborty, Remco van der Hofstad, and Frank den Hollander. Sparse random graphs with many triangles. *arXiv preprint arXiv:2112.06526*, 2021.

[11] Sourav Chatterjee. The missing log in large deviations for triangle counts. *Random Structures Algorithms*, 40(4):437–451, 2012.

[12] Sourav Chatterjee. An introduction to large deviations for random graphs. *Bull. Amer. Math. Soc. (N.S.)*, 53(4):617–642, 2016.

[13] Sourav Chatterjee and Amir Dembo. Nonlinear large deviations. *Adv. Math.*, 299:396–450, 2016.

[14] Sourav Chatterjee and Partha S. Dey. Applications of Stein’s method for concentration inequalities. *Ann. Probab.*, 38(6):2443–2485, 2010.

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

[16] Nicholas Cook and Amir Dembo. Large deviations of subgraph counts for sparse erdős–reýny graphs. *Advances in Mathematics*, 373:107289, 2020.

[17] B. DeMarco and J. Kahn. Tight upper tail bounds for cliques. *Random Structures Algorithms*, 41(4):469–487, 2012.

[18] B. DeMarco and J. Kahn. Upper tails for triangles. *Random Structures Algorithms*, 40(4):452–459, 2012.

[19] Matan Harel, Frank Mousset, and Wojciech Samotij. Upper tails via high moments and entropic stability. *arXiv:1904.08212*, 2019.

[20] Svante Janson, Tomasz Łuczak, and Andrzej Ruciński. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

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

[22] Svante Janson and Andrzej Ruciński. The infamous upper tail. *Random Structures Algorithms*, 20(3):317–342, 2002.

[23] Svante Janson and Andrzej Ruciński. The deletion method for upper tail estimates. *Combinatorica*, 24(4):615–640, 2004.

[24] Peter Keevash. Shadows and intersections: Stability and new proofs. *Advances in Mathematics*, 218(5):1685–1703, Aug 2008.

[25] J. H. Kim and V. H. Vu. Divide and conquer martingales and the number of triangles in a random graph. *Random Structures Algorithms*, 24(2):166–174, 2004.

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

[27] Eyal Lubetzky and Yufei Zhao. On the variational problem for upper tails in sparse random graphs. *Random Structures Algorithms*, 50(3):420–436, 2017.

[28] David Reimer. Proof of the van den Berg–Kesten conjecture. *Combinatorics, Probability and Computing*, 9(1):27–32, 2000.

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

DEPARTMENT OF STATISTICS, EVANS HALL, UNIVERSITY OF CALIFORNIA, BERKELEY, CA 94720, USA

Email address: sganguly@berkeley.edu
