On the Upper Tail of Counts of Strictly Balanced Subgraphs

Matas Šileikis

Department of Mathematics and Computer Science,
Adam Mickiewicz University, Poznań, Poland

`matas.sileikis@gmail.com`

Abstract

Let X_G be the number of copies of G in the Erdős-Rényi binomial random graph $\mathbb{G}(n, p)$. Janson, Oleszkiewicz and Ruciński proved that for every $t > 1$

$$\exp\{-O_t(M_G^* \ln(1/p))\} \leqslant \mathbb{P}\{X_G \geqslant t\,\mathbb{E}X_G\} \leqslant \exp\{-\Omega_t(M_G^*)\},$$

where M_G^* is a certain function of n and p. For $G = K_3$ the logarithmic gap between the bounds was closed by Chatterjee and, independently, DeMarco and Kahn. We provide matching bounds for strictly balanced G, when $\mathbb{E}X_G \leqslant \ln n$. Also, we give matching bounds for C_4, K_4, and stars $K_{1,k}$ in a broader range of $\mathbb{E}X_G$. In particular, this improves some results of Janson and Ruciński for which the so called deletion method was used.

1 Introduction

For a fixed graph G, let X_G be the number of copies of G in the random graph $\mathbb{G}(n, p)$ (see, e.g., [5] for definition) and let $\mu_G = \mathbb{E}X_G$. We consider the asymptotics of

$$-\ln \mathbb{P}\{X_G \geqslant t\mu_G\}$$

for fixed $t > 1$, as $n \to \infty$. In the sequel, we always assume that G has at least one edge.

We use the asymptotic notation $O, \Omega, \Theta, o, \asymp, \gg$ as it is defined in [5] with implicit constants possibly depending on G. Subscripts added to these symbols indicate additional parameters on which the implicit constants depend. We treat p as a function of n, and n is assumed to be sufficiently large. The numbers of vertices and edges of a graph H are denoted, respectively, by v_H and e_H.

Let $\Psi_H = n^{v_H} p^{e_H} \asymp \mu_H$, and $\Phi_G = \min_{H \subseteq G} \Psi_H$, where the minimum is taken over subgraphs H with $e_H > 0$. Let $m_G = \max_{H \subseteq G} e_H/v_H$ be the *maximum density* of G.

In the present paper we assume, for simplicity, that $p \geqslant n^{-1/m_G}$. This is equivalent to $\Phi_G \geqslant 1$. Recall that n^{-1/m_G} is the threshold for the event $\{X_G > 0\}$. See [6, Remark 8.1] for an explanation of why we are not interested in p below the threshold.

Also, to avoid some trivialities, we assume that $t\mu_G$ is at most the number of copies of G in K_n, since otherwise $\mathbb{P}\{X_G \geqslant t\mu_G\} = 0$.

Janson, Oleszkiewicz and Ruciński [6] proved that for every $t > 1$

$$p^{O_t(M_G^*)} \leqslant \mathbb{P}\{X_G \geqslant t\mu_G\} \leqslant \exp\{-\Omega_t(M_G^*)\}. \tag{1}$$

In (1), M_G^* is a function of n and p satisfying

$$M_G^* \asymp \begin{cases} \min_{H \subseteq G} \Psi_H^{1/\alpha_H^*}, & \text{if } p \leqslant n^{-1/\Delta_G}, \\ n^2 p^{\Delta_G}, & \text{if } p \geqslant n^{-1/\Delta_G}, \end{cases} \tag{2}$$

where α_H^* stands for the *fractional independence number*, defined as the maximum of $\sum_v \alpha_v$ over all assignments of nonnegative weights α_v to the vertices satisfying $\alpha_u + \alpha_v \leqslant 1$ for every edge uv of H (see [6] for some properties of α_H^*), while Δ_G is the maximum degree. For example, $\alpha_{K_3}^* = 3/2$, and $M_{K_3}^* \asymp n^2 p^2$ for all p.

One should bear in mind that the right-hand side of (2) is continuous as a function of p, and takes a simple form $n^a p^b$ with some constants $a, b > 0$ in each of a finite number of intervals (see [6, Section 7] for a thorough analysis).

The logarithms of the upper and lower bounds in (1) differ by a multiplicative factor $\ln(1/p)$. Recently Chatterjee [1] and, independently, DeMarco and Kahn [3] closed this logarithmic gap for $G = K_3$. Chatterjee proved that if $p > Cn^{-1}\ln n$, where C is a constant depending on t, then

$$\mathbb{P}\{X_{K_3} \geqslant t\mu_{K_3}\} = \exp\left\{-\Theta_t\left(n^2 p^2 \ln(1/p)\right)\right\}, \tag{3}$$

which means that in this range of p the lower bound in (1) is sharp. Shortly after the preprint of Chatterjee [1] was posted, a preprint by DeMarco and Kahn [3] appeared on arXiv with an alternative proof of (3) for $p > n^{-1}\ln n$. They also proved that if $n^{-1} \leqslant p \leqslant n^{-1}\ln n$, then

$$\mathbb{P}\{X_{K_3} \geqslant t\mu_{K_3}\} = \exp\left\{-\Theta_t\left(n^3 p^3\right)\right\}. \tag{4}$$

Thus, in a small range above the threshold neither of the bounds in (1) is sharp. Asymptotics (3) and (4) can be combined into a single result:

$$\mathbb{P}\{X_{K_3} \geqslant t\mu_{K_3}\} = \exp\left\{-\Theta_t\left(\min\left\{\Psi_{K_3}, M_{K_3}^* \ln(1/p)\right\}\right)\right\}. \tag{5}$$

A graph G is called *balanced* if $\max_{H \subseteq G} e_H/v_H$ is attained by $H = G$, and, in particular, *strictly balanced*, if it is attained by $H = G$ only.

The first result of this paper is the following theorem, which gives asymptotics analogous to (4) for strictly balanced G, when Ψ_G is logarithmically small.

Theorem 1. *Let G be a strictly balanced graph. Then, for every $t > 1$ and $\Psi_G \leqslant \ln n$,*

$$\mathbb{P}\left\{X_G \geqslant t\mu_G\right\} \leqslant \exp\{-\Omega_t(\mu_G)\}, \tag{6}$$

and for $\Psi_G \leqslant \ln^a n$, where $a = \alpha_G^*/(\alpha_G^* - 1)$,

$$\mathbb{P}\left\{X_G \geqslant t\mu_G\right\} \geqslant \exp\{-O_t(\mu_G)\}. \tag{7}$$

Let us clarify the above restrictions on Ψ_G. The upper (respectively, lower) bound holds for $\Psi_G \leqslant B \ln n$ (respectively, $\Psi_G \leqslant B \ln^a n$) for any fixed $B > 0$, if we let the implicit constants depend on B. Therefore, for simplicity, we take $B = 1$. For the upper bound (6) the restriction $\Psi_G \leqslant \ln n$ seems to be a technical limitation. It is the cost we pay for the simplicity of the proof. We hope that with more effort it can be relaxed to $\Psi_G \leqslant \ln^a n$ to match the range of the validity of the lower bound. A suggestion how one could try to obtain such a relaxation is outlined in Remark 3.

Remark 1. The lower bound (7) is obtained by extending the approach of DeMarco and Kahn [3]. The bound holds much more generally, but in Theorem 1 we state it only for the range where it is better than the lower bound in (1). In other words, inequality $\Psi_G \leqslant \ln^a n$ is roughly equivalent to $\mu_G \leqslant M_G^* \ln(1/p)$ (see Remark 2).

Actually, in the proof of (7) we bound the tail $\mathbb{P}\left\{X_G \geqslant t\mu_G\right\}$ from below by the probability that the number of copies is exactly $\lceil t\mu_G \rceil$, and, moreover, these copies are vertex-disjoint.

During his plenary lecture at the conference "Random Structures and Algorithms, Atlanta, May 2011" J. Kahn presented a result (to appear in [2]) that for any clique $G = K_r$ asymptotics analogous to (5) hold, namely,

$$\mathbb{P}\left\{X_{K_r} \geqslant t\mu_{K_r}\right\} = \exp\left\{-\Theta_t\left(\min\{\mu_{K_r}, M_{K_r}^* \ln(1/p)\}\right)\right\}. \tag{8}$$

In the same lecture the following conjecture of DeMarco and Kahn was stated:

Conjecture 1. *If G is a graph and $t > 1$, then for every p*

$$\mathbb{P}\left\{X_G \geqslant t\mu_G\right\} = \exp\{-\Theta_t(\min\{\Phi_G, M_G^* \ln(1/p)\})\}.$$

Moreover, in the aforementioned talk it was mentioned that for any graph G and any subgraph $H \subseteq G$ it is not hard to prove that

$$\mathbb{P}\left\{X_G \geqslant 2\mu_G\right\} \geqslant \exp\left\{-O(\mu_H)\right\}, \tag{9}$$

which is more general than (7). However, we do not know whether a proof of (9) in full generality is going to be published soon. Therefore, acknowledging the origin of the idea, we provide a proof of (7).

In this paper we also prove the asymptotics for $G = K_4, C_4$, and $G = K_{1,k}$, $k \geqslant 2$, thus supporting Conjecture 1. We still assume some conditions for p, but these are less restrictive than the condition $\Psi_G \leqslant \ln n$ assumed in Theorem 1: the order of Ψ_G may now be a power of n. Note that all these graphs are strictly balanced.

Theorem 2. *Let $\gamma > 0$ and $t > 1$.*

(i) *If $p \leqslant n^{-4/5-\gamma}$, then*

$$\mathbb{P}\left\{X_{C_4} \geqslant t\mu_{C_4}\right\} = \exp\left\{-\Theta_{t,\gamma}\left(\min\left\{\Psi_{C_4}, M_{C_4}^* \ln n\right\}\right)\right\}.$$

(ii) *If $p \leqslant n^{-1/2-\gamma}$, then*

$$\mathbb{P}\left\{X_{K_4} \geqslant t\mu_{K_4}\right\} = \exp\left\{-\Theta_{t,\gamma}\left(\min\left\{\Psi_{K_4}, M_{K_4}^* \ln n\right\}\right)\right\}.$$

(iii) *For every $k \geqslant 2$, if $K = K_{1,k}$ is the k-armed star and $p \leqslant n^{-\theta-\gamma}$, where $\theta = 1 + k^{-1} - k^{-2}$, then*

$$\mathbb{P}\left\{X_K \geqslant t\mu_K\right\} = \exp\left\{-\Theta_{t,\gamma}\left(\min\left\{\Psi_K, M_K^* \ln n\right\}\right)\right\}.$$

Notice that the threshold for the existence of the star $K_{1,k}$ is $p = n^{-1-k^{-1}}$.

Well before the asymptotics for triangles were found, Janson and Ruciński [8] improved the upper bound in (1) in some range of p for graphs $G = K_4$ and $G = C_4$ by inserting a factor $\ln^{1/2} n$ in the exponent. Assume that $\mu_{C_4} \geqslant C \ln n$ for some large $C > 0$ and $p \leqslant n^{-2/3-\gamma}$. Then

$$\mathbb{P}\left\{X_{C_4} \geqslant 2\mu_{C_4}\right\} \leqslant \exp\{-\Omega_\gamma(M_{C_4}^* \ln^{1/2} n)\}. \tag{10}$$

If $\mu_{K_4} \geqslant C \ln n$ and $p \leqslant n^{-1/2-\gamma}$ for some constant $\gamma > 0$, then

$$\mathbb{P}\left\{X_{K_4} \geqslant 2\mu_{K_4}\right\} \leqslant \exp\{-\Omega_\gamma(M_{K_4}^* \ln^{1/2} n)\}. \tag{11}$$

Although these inequalities are stated for $t = 2$, the proof in [8] can be easily extended to all $t > 1$.

Theorem 2.2 improves (10) for $p \leqslant n^{-4/5-\gamma}$ and Theorem 2.2 improves (11) for $p \leqslant n^{-1/2-\gamma}$. Also, note that Theorem 2 assumes no restriction on p from below, except for the universally assumed $p \geqslant n^{-1/m_G}$. Finally, observe that part (ii) of Theorem 2 is a special case of (8).

The inequalities (10) and (11) were obtained by the so called deletion method. We prove Theorem 2 using an older method of approximation by disjoint copies, originating from Spencer [9] (see [7, Section 2.3.4]). However, in the proof of Theorem 2 for C_4 and K_4, we retain the ad hoc part of the original proof of (10) and (11) by Janson and Ruciński [8], which involves a two-fold application of Chernoff's bound and heavily relies on the simplicity of C_4 and K_4.

2 The Proof of Theorem 1

The example of the triangle suggests that, roughly speaking, for small p the upper tail of X_G behaves as if the copies of G were disjoint. Recalling that $n^3 p^3 \asymp \mu_{K_3}$, compare (4)

with the following fact. Let D_G^e be the maximum number of edge-disjoint copies of G, and let $\varphi(\varepsilon) = (1 + \varepsilon) \ln(1 + \varepsilon) - \varepsilon$. Lemma 2.46 in [5] states that for every $\varepsilon > 0$

$$\mathbb{P}\left\{D_G^e \geqslant (1 + \varepsilon)\mu_G\right\} \leqslant \exp\{-\mu_G\varphi(\varepsilon)\} \tag{12}$$

$$\leqslant \exp\left\{-\frac{\varepsilon^2}{2(1 + \varepsilon/3)}\mu_G\right\}. \tag{13}$$

We proceed with some auxiliary facts. The advantage of using Ψ_G instead of μ_G is that it satisfies the log-modularity property: if G_1, G_2 are graphs, then

$$\Psi_{G_1 \cup G_2}\Psi_{G_1 \cap G_2} = \Psi_{G_1}\Psi_{G_2}. \tag{14}$$

Recall that for strictly balanced G we have $m_G = e_G/v_G$, so the threshold for existence of G is $p = n^{-v_G/e_G}$.

Proposition 3. *If a graph G is strictly balanced, then the following facts hold.*

(i) *G is connected.*

(ii) *If H is a proper subgraph of G and $p \geqslant n^{-v_G/e_G}$, then*

$$\Psi_H \geqslant n^c, \tag{15}$$

where $c = c(H) = v_H - e_H v_G/e_G > 0$.

(iii) *There is a constant $b \in (0, v_G/e_G)$ such that for $p \in [n^{-v_G/e_G}, n^{-b}]$ and $H \subseteq G$ with $v_H > 0$ we have*

$$\Psi_G \leqslant \Psi_H. \tag{16}$$

(iv) *Let $a > 0$ be a constant. If $p \geqslant n^{-v_G/e_G}$ and $\Psi_G \leqslant \ln^a n$, then*

$$M_G^* \asymp \Psi_G^{1/\alpha_G^*}.$$

Proof. (i) If G were not connected, then it would have a connected component as dense as G itself, contradicting the fact that G is strictly balanced.

(ii) To prove (15), note that the condition $p \geqslant n^{-v_G/e_G}$ implies

$$\Psi_H = n^{v_H}p^{e_H} \geqslant n^{v_H - e_H v_G/e_G}.$$

Since G is strictly balanced, we have $e_G/v_G > e_H/v_H$, which is equivalent to the inequality $v_H - e_H v_G/e_G > 0$.

(iii) Inequality (16) follows from (ii) with $b = (v_G - \min_{H \subsetneq G} c(H))/e_G$:

$$\Psi_G = n^{v_G}p^{e_G} \leqslant n^{v_G - be_G} \leqslant n^{c(H)} \leqslant \Psi_H, \qquad H \subsetneq G.$$

(iv) Clearly $\Psi_G \leqslant \ln^a n$ implies $p \ll n^{-1/\Delta_G}$. Therefore by (2)

$$M_G^* \asymp \min_{H \subseteq G} \Psi_G^{1/\alpha_G^*}.$$

Let H be a proper subgraph of G. We have $\Psi_H \geqslant n^c$ from (ii). It follows that

$$\Psi_H^{1/\alpha_H^*} \geqslant n^{c/\alpha_H^*} \gg (\ln n)^{a/\alpha_G^*} \geqslant \Psi_G^{1/\alpha_G^*}.$$

$\square$

Proof of (6). Let $\mathcal{F}_G = \{F = G_1 \cup G_2 : 0 < e_{G_1 \cap G_2} < e_G\}$ be the set of all unlabeled graphs obtained by taking a union of two distinct copies of G with at least one common edge. For example $\mathcal{F}_{K_3} = \{K_4^-\}$, where K_4^- is K_4 with one edge removed. Recall that D_G^e is the maximum number of edge-disjoint copies of G in $\mathbb{G}(n,p)$. Obviously, either $X_F \geqslant 1$ for some $F \in \mathcal{F}_G$ or $D_G^e = X_G$, therefore

$$\mathbb{P}\{X_G \geqslant t\mu_G\} \leqslant \mathbb{P}\{D_G^e \geqslant t\mu_G\} + \sum_{F \in \mathcal{F}_G} \mathbb{P}\{X_F \geqslant 1\}. \tag{17}$$

Clearly, it suffices to show that all probabilities on the right-hand side of (17) are bounded by $\exp\{-\Omega_t(\mu_G)\}$. Applying (13) with $\varepsilon = t - 1$, we get

$$\mathbb{P}\{D_G^e \geqslant t\mu_G\} \leqslant \exp\{-\Omega_t(\mu_G)\}.$$

We bound the remaining terms in (17) using Markov's inequality:

$$\mathbb{P}\{X_F \geqslant 1\} \leqslant \mu_F \leqslant \Psi_F.$$

If $F = G_1 \cup G_2 \in \mathcal{F}_G$ and $H = G_1 \cap G_2$, then, by the log-modularity (14), $\Psi_F = \Psi_G^2/\Psi_H$. By Proposition 3.(ii) we have that $\Psi_H \geqslant n^c$ for some constant $c > 0$. Moreover, by the assumption, $\Psi_G \leqslant \ln n$. Therefore

$$\ln \Psi_F \leqslant 2\ln\ln n - c\ln n = -\Omega(\ln n).$$

Finally, since $\mu_G \leqslant \Psi_G \leqslant \ln n$, we get

$$\mathbb{P}\{X_F \geqslant 1\} \leqslant \Psi_F \leqslant \exp\{-\Omega(\ln n)\} \leqslant \exp\{-\Omega(\mu_G)\}.$$

$\square$

Proof of (7). Let $N(F, H)$ be the number of copies of a graph H in another graph F. Let xG stand for a union of x vertex-disjoint copies of G. Let

$$D_G^v = \max\{x : \mathbb{G}(n,p) \text{ contains } xG\}$$

be the size of a largest collection of vertex disjoint copies. Writing $x = \lceil t\mu_G \rceil$, we have

$$\mathbb{P}\{X_G \geqslant t\mu_G\} \geqslant \mathbb{P}\{X_G = D_G^v = x\}. \tag{18}$$

Let $\mathcal{G}$ be the set of all copies of G in K_n. Consider the family

$$\mathbf{F} = \{\mathcal{S} \subset \mathcal{G} : |\mathcal{S}| = x, \text{ copies in } \mathcal{S} \text{ are vertex disjoint}\}.$$

For every $\mathcal{S} \in \mathbf{F}$, we define events

$$R_{\mathcal{S}} = \{\mathcal{S} \text{ is the set of all copies of } G \text{ in } \mathbb{G}(n,p)\},$$

$$Q_{\mathcal{S}} = \{\text{every copy } S \in \mathcal{S} \text{ appears in } \mathbb{G}(n,p)\}.$$

By Proposition 3.(i), G is connected, so the events $\{R_{\mathcal{S}}\}_{\mathcal{S} \in \mathbf{F}}$ are mutually exclusive. Therefore

$$\mathbb{P}\{X_G = D_G^v = x\} = \sum_{\mathcal{S} \in \mathbf{F}} \mathbb{P}\{R_{\mathcal{S}}\}. \tag{19}$$

Let $\mathrm{aut}(G)$ stand for the number of automorphisms of G. Then

$$|\mathbf{F}| = \frac{(n)_{v_G x}}{\mathrm{aut}(G)^x x!} \geqslant \frac{(n)_{v_G x}}{\mathrm{aut}(G)^x x^x}.$$

Using a standard inequality $(n)_m \geqslant (n/e)^m$, we get

$$|\mathbf{F}| \geqslant \frac{n^{v_G x}}{e^{v_G x} \mathrm{aut}(G)^x x^x} = \frac{n^{v_G x}}{x^x} \exp\{-\Theta(x)\}. \tag{20}$$

Note that

$$\mathbb{P}\{R_{\mathcal{S}}\} = \mathbb{P}\{R_{\mathcal{S}}|Q_{\mathcal{S}}\}\,\mathbb{P}\{Q_{\mathcal{S}}\} = \mathbb{P}\{R_{\mathcal{S}}|Q_{\mathcal{S}}\}\, p^{e_G x}. \tag{21}$$

Also, by symmetry the probability $q := \mathbb{P}\{R_{\mathcal{S}}|Q_{\mathcal{S}}\}$ is independent of $\mathcal{S}$. Therefore from (19), (20), and (21) we infer that

$$\mathbb{P}\{X_G = D_G^v = x\} \geqslant \frac{(n^{v_G} p^{e_G})^x q}{x^x} \exp\{-\Theta(x)\}. \tag{22}$$

By the assumption that $p \geqslant n^{-1/m_G}$, we have $\Psi_G = n^{v_G} p^{e_G} \asymp x/t$, so (22) implies

$$\mathbb{P}\{X_G = D_G^v = x\} \geqslant q \exp\{-\Theta_t(\Psi_G)\}. \tag{23}$$

Fix $\mathcal{S} \in \mathbf{F}$ and let $F = \cup_{S \in \mathcal{S}} S$. Let $\mathbb{G}_{\mathcal{S}}(n,p)$ be the random graph $\mathbb{G}(n,p)$ conditioned on $Q_{\mathcal{S}}$, i.e., the random graph obtained by adding to F each of the remaining $\binom{n}{2} - e_G x$ edges with probability p, independently of others. We have

$$q = \mathbb{P}\Big\{ \bigcap_{G' \in \mathcal{G}\backslash\mathcal{S}} \{G' \not\subseteq \mathbb{G}_{\mathcal{S}}(n,p)\} \Big\}. \tag{24}$$

Notice that each of the events $\{G' \not\subseteq \mathbb{G}_{\mathcal{S}}(n,p)\}$ is decreasing. Therefore the right-hand side of (24) can be bounded using the FKG inequality (see, e.g., [5, Theorem 2.12]; the set Γ to which this theorem refers is $E(K_n) \setminus E(F)$ in the present context). Thus we obtain

$$q \geqslant \prod_{G' \in \mathcal{G}\backslash\mathcal{S}} \mathbb{P}\{G' \not\subseteq \mathbb{G}_{\mathcal{S}}(n,p)\}. \tag{25}$$

Given $H \subsetneq G$, the number of copies $G' \in \mathcal{G} \setminus \mathcal{S}$ whose intersection with F is a copy of H is at most $N(F, H)(n - v_H)_{v_G - v_H} = N(xG, H)\frac{(n)_{v_G}}{(n)_{v_H}}$. The probability that such a copy G' does not exist in $\mathbb{G}_{\mathcal{S}}(n, p)$ is

$$1 - p^{e_G - e_H} \geqslant \exp\left\{-\frac{p^{e_G - e_H}}{1 - p^{e_G - e_H}}\right\} \geqslant \exp\left\{-c_p \frac{p^{e_G}}{p^{e_H}}\right\},$$

where $c_p = 1/(1 - p)$. Hence (25) implies

$$q \geqslant \prod_{H \subsetneq G} \prod_{G' \cap F \cong H} \exp\left\{-c_p \frac{p^{e_G}}{p^{e_H}}\right\}$$

$$\geqslant \prod_{H \subsetneq G} \exp\left\{-c_p N(xG, H) \cdot \frac{(n)_{v_G} p^{e_G}}{(n)_{v_H} p^{e_H}}\right\}. \tag{26}$$

The assumption $\Psi_G \leqslant \ln^a n$ is equivalent to

$$p \leqslant n^{-1/m_G} (\ln n)^{a/e_G}. \tag{27}$$

By (27), we have $p = o(1)$, therefore $c_p \to 1$. Also, $(n)_{v_G} p^{e_G} \asymp \Psi_G$ and $(n)_{v_H} p^{e_H} \asymp \Psi_H$. Hence (26) gives that

$$q \geqslant \exp\left\{-\Theta\left(\Psi_G \max_{H \subsetneq G} \frac{N(xG, H)}{\Psi_H}\right)\right\}. \tag{28}$$

If H is empty (i.e., has no vertices), then $N(xG, H) = 1 = \Psi_H$. Hence the maximum in (28) is at least 1. Therefore from (18), (23), and (28) we get

$$\mathbb{P}\{X_G \geqslant t\mu_G\} \geqslant \exp\left\{-O_t\left(\Psi_G \max_{H \subsetneq G} \frac{N(xG, H)}{\Psi_H}\right)\right\}. \tag{29}$$

To finish the proof, it is enough to show that for $H \subsetneq G$ we have $N(xG, H) = O_t(\Psi_H)$. Indeed, then $\Psi_G \asymp \mu_G$ and (29) imply (7). As we have seen, if H is the empty graph, then $N(xG, H) = \Psi_H$. Let us further assume that $v_H > 0$. By (27) and Proposition 3.(iii), the inequality $\Psi_G \leqslant \Psi_H$ holds for n large enough, i.e., $\Psi_G = O(\Psi_H)$. If H is connected, then any copy of H counted in $N(xG, H)$ must lie entirely in one of the x copies of G. Hence $N(xG, H) = xN(G, H) \asymp_t \Psi_G = O_t(\Psi_H)$. If H is not connected, then let $H_1, \ldots, H_c$ be the connected components of H. Since each component H_i has $xN(G, H_i)$ copies in xG,

$$N(xG, H) \leqslant x^c \prod_{i=1}^{c} N(G, H_i) \asymp_t \Psi_G^c.$$

On the other hand, by log-modularity, $\Psi_H = \prod_i \Psi_{H_i} = \Omega(\Psi_G^c)$ and therefore $N(xG, H) = O_t(\Psi_H)$. $\qquad\square$

3 The Proof of Theorem 2

Note that the conditions on p imply $\ln(1/p) \asymp \ln n$. Therefore, in view of Remark 1, the lower bounds are given by (1) and (7). So it remains to prove the upper bounds.

The proof is presented as follows. We start with an argument, which works for any graph G. Then we finish the proof for $G = C_4$. Since the proof for $G = K_4$ is very similar, we just mention the differences. Finally, we point out the changes that need to be made to the proof for C_4 in order to prove the result for stars $K_{1,k}$.

Let L be the random intersection graph, the vertices of which are the copies of G in $\mathbb{G}(n, p)$ and two vertices are connected by an edge if the corresponding copies have an edge in common. An easy graph-theoretic result, appearing implicitly in Spencer [9] (see, e.g., Janson [4] for the proof) states that for any graph L

$$v_L \leqslant \alpha_L + 2\beta_L \Delta_L, \tag{30}$$

where α_L is the independence number of L, β_L is the size of a largest induced matching in L, and, recall, Δ_L is the maximum degree of L.

In our setting, $v_L = X_G$ and $\alpha_L = D_G^e$, while $\beta_L \leqslant \sum_{F \in \mathcal{F}_G} D_F^e$, where $\mathcal{F}_G$ is the set of all graphs formed by a union of two distinct edge-intersecting copies of G, as defined in the proof of Theorem 1.

For fixed vertices u and v, let X_G^{uv} be the number of copies of G containing the edge uv. Then $\Delta_L \leqslant e_G \max_{uv} X_G^{uv}$, where the maximum is taken over all edges of K_n. Clearly, the distribution of X_G^{uv} does not depend on uv. Therefore, when the choice of uv does not matter, we write $X_G^{(2)}$ instead.

In view of the observations above, (30) implies

$$X_G \leqslant D_G^e + \sum_{F \in \mathcal{F}_G} D_F^e \cdot 2e_G \max_{uv} X_G^{uv}. \tag{31}$$

Let δ be such that $t = 1 + \delta(1 + 2e_G |\mathcal{F}_G|)$. Then, by (31), for every $d > 0$,

$$\mathbb{P}\{X_G \geqslant t\mu_G\} \leqslant \mathbb{P}\left\{D_G^e \geqslant (1+\delta)\mu_G\right\}$$
$$+ \sum_{F \in \mathcal{F}_G} \mathbb{P}\left\{D_F^e \geqslant \frac{\delta \mu_G}{d}\right\} + \binom{n}{2}\mathbb{P}\left\{X_G^{(2)} > d\right\}. \tag{32}$$

Proof of Theorem 2.2. From [6] we know that

$$M_{C_4}^* \asymp n^2 p^2 = \Psi_{C_4}^{1/2}. \tag{33}$$

Hence our aim is to prove that the right-hand side of (32) is bounded by

$$\exp\left\{-\Omega_{\delta,\gamma}\left(\min\left\{\Psi_{C_4}, \Psi_{C_4}^{1/2}\ln n\right\}\right)\right\}.$$

By inequality (13), the first term in (32) is at most

$$\exp\{-\Omega_\delta(\mu_{C_4})\} \leqslant \exp\{-\Omega_\delta(\min\{\Psi_{C_4}, \Psi_{C_4}^{1/2}\ln n\})\}.$$

To bound the second term in (32), we apply the following slightly weaker but more convenient form of (12). Noting that $\varphi(\varepsilon) \geqslant (1+\varepsilon) \ln \frac{1+\varepsilon}{e}$ and using a substitute $x = (1+\varepsilon)\mu_G$, we get an inequality

$$\mathbb{P}\{D_G^e \geqslant x\} \leqslant \exp\left\{-x \ln \frac{x}{e\mu_G}\right\}, \qquad e = 2.71\ldots. \tag{34}$$

Hence, for $F \in \mathcal{F}_{C_4}$, by (34),

$$\mathbb{P}\left\{D_F^e \geqslant \frac{\delta\mu_{C_4}}{d}\right\} \leqslant \exp\left\{-\frac{\delta\mu_{C_4}}{d} \ln \frac{\delta\mu_{C_4}}{ed\mu_F}\right\}.$$

Put

$$d = \max\{\ln n, \Psi_{C_4}^{1/2}\} = \max\left\{\ln n, n^2 p^2\right\}.$$

Then

$$\mathbb{P}\left\{D_F^e \geqslant \frac{\delta\mu_{C_4}}{d}\right\} \leqslant \exp\left\{-\Omega_\delta\left(\min\left\{\frac{\Psi_{C_4}}{\ln n}, \Psi_{C_4}^{1/2}\right\} \ln \frac{\Psi_{C_4}}{d\Psi_F}\right)\right\}. \tag{35}$$

It remains to check that the logarithmic factor in (35) is of order $\ln n$. Note that by log-modularity (14), $\Psi_{C_4}/\Psi_F = \Psi_H/\Psi_{C_4}$, where H is the intersection of the two copies of C_4 that make up F. Hence, it suffices to show that $\Psi_H/(d\Psi_{C_4}) = \Omega(n^c)$ for some $c > 0$, probably depending on γ. Consider two cases.

Case (i): $\Psi_{C_4} \leqslant \ln^2 n$. Then $d = \ln n$. By Proposition 3.(ii), we have that Ψ_H is at least some positive power of n. Therefore

$$\frac{\Psi_H}{d\Psi_{C_4}} \geqslant \frac{\Psi_H}{\ln^3 n} = \Omega(n^c), \quad c > 0.$$

Case (ii): $\Psi_{C_4} > \ln^2 n$. Then $d = \Psi_{C_4}^{1/2}$, and so

$$\frac{\Psi_H}{d\Psi_{C_4}} = \frac{\Psi_H}{\Psi_{C_4}^{3/2}} = n^{v_H - 6} p^{e_H - 6}. \tag{36}$$

Restriction $p \leqslant n^{-4/5-\gamma}$ implies that the right-hand side of (36) is at least $n^{b(H)}$ with

$$b(H) = v_H - 6 + (4/5 + \gamma)(6 - e_H). \tag{37}$$

Note that H is a proper subgraph with at least one edge. Observe that $b(H) \geqslant b(H')$, where H' is a proper subgraph of G on v_H vertices and with maximum number of edges. Checking that $H' = K_2, P_3, P_4$ give corresponding values $b(H') = 5\gamma, 1/5 + 4\gamma, 2/5 + 3\gamma$, we obtain that $\Psi_H/(d\Psi_{C_4}) \geqslant n^c$, where $c = \min\{5\gamma, 2/5 + 3\gamma\} > 0$, thus concluding Case (ii).

Concerning the third term in (32), we use the proof from [8, Example 6.3] which, by a double application of the Chernoff bound, shows that for every $m > 0$,

$$\mathbb{P}\left\{X_{C_4}^{(2)} > d\right\} \leqslant \exp\left\{-m \ln\left(\frac{m}{2enp}\right)\right\} + \exp\left\{-\frac{d}{2} \ln\left(\frac{d}{em^2p}\right)\right\}. \tag{38}$$

Recall that $d = \max\{\ln n, \Psi_{C_4}^{1/2}\}$. Set $m = d \ln n$. Hence

$$d = \max\{\ln n, n^2 p^2\}, \qquad \text{and} \qquad m = \max\{\ln^2 n, n^2 p^2 \ln n\}.$$

Note that in (38) the logarithmic factor in the first term is $\Omega(1)$, while the one in the second term is $\Omega_\gamma(\ln n)$. This is indeed so, since

$$\frac{m}{np} = \max\left\{\frac{\ln^2 n}{np}, np \ln n\right\} \geqslant \ln^{3/2} n, \qquad \text{and}$$

$$\frac{d}{m^2 p} = \frac{1}{dp \ln^2 n} \geqslant \min\left\{\frac{1}{p \ln^3 n}, \frac{1}{n^2 p^3 \ln^2 n}\right\} \geqslant \min\left\{\frac{n^{4/5+\gamma}}{\ln^3 n}, \frac{n^{2/5+3\gamma}}{\ln^2 n}\right\}.$$

Thus, (38) implies that

$$\mathbb{P}\left\{X_{C_4}^{(2)} \geqslant d\right\} \leqslant \exp\left\{-\Omega_\gamma\left(\max\left\{\ln^2 n, \Psi_{C_4}^{1/2} \ln n\right\}\right)\right\}.$$

We can ignore the factor $\binom{n}{2}$ in (32), since it contributes to the exponent only an additive term $O(\ln n)$. This completes the proof, since

$$\max\{\ln^2 n, \Psi_{C_4}^{1/2} \ln n\} \geqslant \min\{\Psi_{C_4}, \Psi_{C_4}^{1/2} \ln n\}.$$

$\square$

Proof of Theorem 2.2. The proof is very similar to that of Theorem 2.2. One needs to replace $n^2 p^2$ by $n^2 p^3$ in (33), and $p^{e_H - 6}$ by $p^{e_H - 9}$ in (36). Also we use the restriction $p \leqslant n^{-1/2-\gamma}$, so (37) becomes $b(H) = v_H - 6 + (1/2 + \gamma)(9 - e_H)$. In the present proof we need to check $b(H') > 0$ for $H' = K_2, K_3, K_4^-$. The proof from [8, Example 6.2] gives that

$$\mathbb{P}\left\{X_{K_4}^{(2)} > d\right\} \leqslant \exp\left\{-m \ln\left(\frac{m}{enp^2}\right)\right\} + \exp\left\{-d \ln\left(\frac{2d}{em^2 p}\right)\right\}. \tag{39}$$

Just like in the proof for C_4, we set $d = \max\left\{\ln n, \Psi_{K_4}^{1/2}\right\}$ and $m = d \ln n$, so

$$d = \max\left\{\ln n, n^2 p^3\right\} \qquad \text{and} \qquad m = \max\left\{\ln^2 n, n^2 p^3 \ln n\right\}.$$

Further, we note that the logarithmic factors in the exponents of (39) are $\Omega(1)$ and $\Omega_\gamma(\ln n)$, respectively, because

$$\frac{m}{np^2} = \max\left\{\frac{\ln^2 n}{np^2}, np \ln n\right\} \geqslant \left(\frac{\ln^3 n}{p}\right)^{1/2} \geqslant \ln^{3/2} n,$$

and

$$\frac{d}{m^2 p} = \frac{1}{dp \ln^2 n} \geqslant \min\left\{\frac{1}{p \ln^3 n}, \frac{1}{n^2 p^4 \ln^2 n}\right\} \geqslant \min\left\{\frac{n^{1/2+\gamma}}{\ln^3 n}, \frac{n^{4\gamma}}{\ln^2 n}\right\}.$$

Finally, we finish the proof precisely as in the case $G = C_4$. $\square$

Proof of Theorem 2.2. From [6, Corollary 1.8] we know that in the given range of p we have $M_K^* \asymp \Psi_K^{1/k}$.

The argument is, again, similar to that of Theorem 2.2. At the appropriate step we set $d = \max\left\{\ln n, \Psi_K^{(k-1)/k}\right\}$. Then we apply (34) to obtain a bound analogous to (35). To show that $\Psi_H/(d\Psi_K) = \Omega(n^c)$, we notice that the intersection H of two distinct copies of K is a star with at most $k-1$ arms and use the condition $p \leqslant n^{-\theta-\gamma}$.

Further we observe, that the number of k-stars containing a particular edge uv is $\binom{\deg(u)}{k-1} + \binom{\deg(v)}{k-1}$. Therefore, writing $\mathrm{Bi}(n,p)$ for a binomial random variable, we get that

$$\mathbb{P}\left\{X_K^{(2)} \geqslant d\right\} \leqslant 2\mathbb{P}\left\{\binom{\mathrm{Bi}(n-2,p)}{k-1} \geqslant d/2\right\}.$$

Then the Chernoff bound yields

$$\mathbb{P}\left\{X_K^{(2)} > d\right\} \leqslant \exp\left\{-\Omega\left(d^{1/(k-1)}\ln\frac{d^{1/(k-1)}}{\mathrm{e}np}\right)\right\}. \tag{40}$$

It is easy to check that the logarithmic factor in the exponent in (40) is $\Omega(\ln n)$, whence the order of the exponent is

$$\max\left\{(\ln n)^{1+\frac{1}{k-1}}, \Psi_K^{1/k}\ln n\right\} \geqslant \min\left\{\Psi_K, \Psi_K^{1/k}\ln n\right\} \asymp \min\left\{\Psi_K, M_K^*\ln n\right\}.$$

$\square$

4 Further Remarks

Remark 2. Let G be strictly balanced. From the proof of (7) one can see that a lower bound $\exp\{-\Theta_t(\mu_G)\}$ holds, as long as $\Psi_G = \min_{H \subseteq G} \Psi_H$ and $v_G t\mu_G \leqslant n$. However, this bound is better than the lower bound in (1) only when $\Psi_G \leqslant \ln^a n$. More precisely, if $\Psi_G \leqslant \ln^a n$, then

$$\Psi_G = O_t(M_G^*\ln(1/p)), \tag{41}$$

while if $\Psi_G \gg \ln^a n$, then

$$\Psi_G \gg M_G^*\ln(1/p). \tag{42}$$

Let us prove (41) and (42). The restriction $v_G t\mu_G \leqslant n$ implies

$$p = O_t(n^{-(v_G-1)/e_G}). \tag{43}$$

The assumption $p \geqslant n^{-1/m_G}$ and (43) clearly imply $\ln(1/p) \asymp_t \ln n$. So it is enough to prove (41) and (42) with $\ln n$ instead of $\ln(1/p)$.

The assumption $\Psi_G \leqslant \ln^a n$ is equivalent to $\Psi_G \leqslant \Psi_G^{1/\alpha_G^*}\ln n$. On the other hand, by Proposition 3.(iv), $\Psi_G^{1/\alpha_G^*} \asymp M_G^*$, so (41) holds.

Since $1/\Delta_G \leqslant v_G/(2e_G) \leqslant (v_G - 1)/e_G$, inequality (43) gives that $p = O_t(n^{-1/\Delta_G})$. Hence from (2) we have that

$$M_G^* \asymp_t \min_{H \subseteq G} \Psi_H^{1/\alpha_H^*}.$$

Thus, the right-hand side of (42) is $O_t\left(\Psi_G^{1/\alpha_G^*} \ln n\right)$. Since $\Psi_G \gg \ln^a n$ is equivalent to $\Psi_G \gg \Psi_G^{1/\alpha_G^*} \ln n$, we conclude that (42) holds.

Remark 3. Let L be the intersection graph of copies of G, as defined in the proof of Theorem 2. In the proof of the upper bound in Theorem 1, we implicitly use the trivial fact that if L has no edges, then the largest independent set in L is the whole vertex set. This implies the inequality (17). However, one can get a stronger conclusion by using the fact that if edges are few, then L has a large independent set. Indeed, by Turán's Theorem applied to the complement of L we get that

$$D_G^e \geqslant \frac{X_G^2}{X_G + 2 \sum_{F \in \mathcal{F}_G} X_F}.$$

This gives an inequality

$$\mathbb{P}\{X_G \geqslant t\mu_G\} \leqslant \mathbb{P}\{D_G^e \geqslant (1 + \varepsilon)\mu_G\} + \sum_{F \in \mathcal{F}_G} \mathbb{P}\{X_F \geqslant \varepsilon\mu_G\},$$

with properly chosen $\varepsilon = \varepsilon(t) > 0$. Therefore it is of interest to find out in which range of p the inequality

$$\mathbb{P}\{X_F \geqslant \varepsilon\mu_G\} \leqslant \exp\{-\Omega_\varepsilon(\mu_G)\} \tag{44}$$

holds, since this might extend the upper bound in Theorem 1 to larger p (hopefully, up to $n^{-1/m_G}(\ln n)^{a/e_G}$). As we have seen in the proof of Theorem 1, if $\Psi_G \leqslant \ln n$, then (44) is given by Markov's inequality. It is plausible that a stronger inequality might hold. Note that since Ψ_G is less than a power of $\ln n$, we have $\mu_F \ll 1$. In particular, p is below the threshold of the existence of F.

References

[1] Sourav Chatterjee. The missing log in large deviations for triangle counts. *Random Struct. Algorithms*, 2011. doi: 10.1002/rsa.20381.

[2] Bobby DeMarco and Jeff Kahn. Upper tails for cliques. Preprint available from `http://arxiv.org/pdf/1111.6687v1`, 2011.

[3] Bobby DeMarco and Jeff Kahn. Upper tails for triangles. *Random Struct. Algorithms*, 2011. doi: 10.1002/rsa.20382.

[4] Svante Janson. Poisson approximation for large deviations. *Random Struct. Algorithms*, 1(2):221–229, 1990.

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

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

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

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

[9] Joel Spencer. Counting extensions. *J. Combin. Theory Ser. A*, 55(2):247–255, 1990.

[10] Van H. Vu. On the concentration of multi-variate polynomials with small expectation. *Random Struct. Algorithms*, 16(4):344–363, 2000.