Journal of Statistical Physics (2026) 193:81
https://doi.org/10.1007/s10955-026-03647-6

Large Deviations for Subgraphs in Inhomogeneous Random Graphs

Riccardo Michielan[1] · Clara Stegehuis[2] · Bert Zwart[3]

Received: 8 March 2026 / Accepted: 8 June 2026

Abstract
Inhomogeneous random graphs are fundamental models for real-world networks, where prescribed degrees are imposed as soft constraints. A common assumption in such models is that the degree distribution follows a power-law, capturing the heavy-tailed nature observed in many contexts. While various graph functionals have been studied in this setting, inhomogeneity makes their analysis significantly more challenging. Here, we investigate the large deviations of subgraph counts in inhomogeneous random graphs. Rare events concerning these functionals translate into quantifying the probability that extremely large hubs appear in the graph. This can be achieved by defining a specific optimization problem that captures the most likely way to generate numerous additional subgraphs. When the expected number of subgraphs is sublinear in the graph size, polynomially large deviations are possible, and in this case, we can derive sharp results on clique counts.

Keywords Large deviations · Random graphs · Subgraphs · Cliques · Scale-free

1 Introduction

Many real-world networks exhibit heavy-tailed degree distributions that can be modeled by power-laws with parameter α, often with infinite second moment [1]. To model such networks, random graph models with heavy-tailed degree distributions have become central benchmarks. These models are designed to replicate the degree of heterogeneity observed in empirical networks, but typically do not enforce other structural properties. As a result, understanding the behavior of various network statistics in such models has become a key area of

Communicated by Satya Majumdar.

✉ Riccardo Michielan
riccardo.michielan@gssi.it

Clara Stegehuis
c.stegehuis@utwente.nl

Bert Zwart
bert.zwart@cwi.nl

1 Gran Sasso Science Institute, L'Aquila, Italy

2 University of Twente, Enschede, Netherlands

3 Centrum voor Wiskunde en Informatica, Amsterdam, Netherlands

research [2–7]. In this paper, we focus on the occurrence of cliques and, more generally, arbitrary fixed subgraphs. The count of such subgraphs provides insight into the network's local density and so-called network motifs, which have been widely studied in various applications of network science [8–10]. While many real-world networks contain many dense subgraphs, classical sparse random graph models tend to be locally tree-like and thus contain few such structures in the large-network limit. However, in power-law random graphs, the presence of high-degree vertices (hubs) can significantly increase the likelihood of observing dense subgraphs, even when the average degree remains finite [11, 12]. This raises two natural questions: how unlikely is it for a sparse power-law random graph to contain an unusually large number of cliques or other fixed subgraphs, and through which mechanism is such a rare event realized? In particular, is the deviation spread across many vertices, or does it concentrate on a small number of exceptional vertices?

The typical abundance and maximal size of cliques in scale-free networks have been studied from a statistical-physics perspective in [13, 14]. These works focus on uncorrelated scale-free ensembles and analyze, among other things, how cutoffs and overlaps between cliques affect clique counts and clique-number estimates. Related work studied loops and Hamilton cycles in scale-free networks, including the role of degree correlations and growing network models [15–17]. Our focus is different: we fix the size k and study upper-tail large deviations of the clique count in the ensemble. In this fixed-k setting, the relevant overlap issue is the dependence between different copies of the same k-clique, and this dependence is controlled by analyzing how many cliques a single edge can be part of.

Large deviation principles for subgraph counts have been extensively studied in Erdős–Rényi graphs. For example, the tail behavior of triangle counts has been characterized in both dense and sparse regimes [18–23], revealing that rare events with many triangles are typically driven by the emergence of localized dense regions. Other work studies the large deviation behavior of random walks and other stochastic processes on graphs [24, 25]. These insights have been extended to general subgraphs in the dense setting, but the sparse case–especially under heavy-tailed degree distributions–remains far less understood. From a statistical-physics perspective, this is a question about the structure of a rare fluctuation. A key distinction is whether the fluctuation is delocalized, involving many degrees of freedom, or condensed, meaning that it is carried by a small number of exceptional ones. In the present inhomogeneous setting, the relevant degrees of freedom are the vertex weights, and our main result shows that upper-tail deviations in clique counts are realized via a condensation mechanism.

Existing large deviation results for inhomogeneous random graphs often rely on the assumption of dense graphs [26] or finite second moments [22, 27, 28], which allow for tractable analysis via factorized connection probabilities. In contrast, when the degree distribution has infinite variance, the saturation of rank-one kernels introduces nonlinear dependencies on vertex weights and structural degree-degree correlations [29, 30]. Our results apply specifically to this sparse saturated rank-one setting with infinite-variance weights, and should not be interpreted as applying to power-law random graphs in general or to uncorrelated null models. The saturation is mathematically important, since it limits the effect of increasing very large weights and is one of the mechanisms behind the condensation phenomena analyzed below. Some progress has been made in regimes where the average degree diverges [31], but these results do not apply to the sparse, power-law regime. To date, large deviation results in this setting are limited to specific functionals such as PageRank [32], edge counts [33, 34], giant component [35], and triangle counts [36].

In this paper, we derive asymptotic estimates for the upper tail probabilities of clique counts and a class of general subgraph counts in sparse, inhomogeneous random graphs with power-law degree distributions that have finite mean and infinite second moment. For k-cliques, we identify the precise decay rate of the upper tail and show that the deviation is realized through a condensation mechanism: the dominant excess number of cliques is generated by $k-2$ exceptional hubs whose degrees are much larger than the typical maximum degree. Thus, the cost of the rare event matches the cost of creating these few atypical vertices, excluding the need for a collective rearrangement of the whole graph. For general subgraphs, the optimal rare-event structure may be different, and we characterize the degree configurations that drive the deviation. This highlights the interplay between subgraph topology and the heavy-tailed nature of the degree distribution. Finally, our result also reveals a transition in the deviation mechanism: for $\alpha > 2 - 2/k$, the upper tail is supported by a finite condensed set of hubs, whereas below this threshold a finite number of hubs no longer suffices, suggesting a qualitatively different regime involving a growing number of atypical vertices.

Notation. We denote $[k] = \{1, 2, \dots, k\}$. We say that a sequence of events $(\mathcal{E}_n)_{n\geq 1}$ happens with high probability (w.h.p.) if $\lim_{n\to\infty} \mathbb{P}(\mathcal{E}_n) = 1$ and we use $\xrightarrow{\mathbb{P}}$ for convergence in probability. We write $f(n) = o(g(n))$ if $\lim_{n\to\infty} f(n)/g(n) = 0$, and $f(n) = O(g(n))$ if $|f(n)|/g(n)$ is uniformly bounded. We write $f(n) = \Theta(g(n))$ if $f(n) = O(g(n))$ as well as $g(n) = O(f(n))$. We say that $X_n = O_{\mathbb{P}}(g(n))$ for a sequence of random variables $(X_n)_{n\geq 1}$ if $|X_n|/g(n)$ is a tight sequence of random variables, and $X_n = o_{\mathbb{P}}(g(n))$ if $X_n/g(n) \xrightarrow{\mathbb{P}} 0$.

Organization of the paper. We first describe the version of the inhomogeneous random graph that we study in Section 2, and proceed to state our main results on clique counts and more general subgraph counts. We then provide a discussion in Section 3, and the proofs of our results for cliques are in Sections 5, and for general subgraphs in Section 6.

2 Main Results

2.1 Model

We consider the rank-1 inhomogeneous random graph (IRG) [37] with heavy-tailed weights. The IRG is constructed by assigning a positive weight W_i to each vertex $i \in [n]$. These weights are assumed to be independent and identically distributed (i.i.d.) samples from a random variable W with values on the interval $[1, \infty)$ and with probability density function $f_W(x) = \alpha x^{-\alpha-1}(1 + o(1))$, where $\alpha \in (1, 2)$. In particular, W is heavy-tailed, and it satisfies

$$\bar{F}(x) := \mathbb{P}(W_i > x) = x^{-\alpha}(1 + o(1)), \quad x \geq 1. \tag{1}$$

Edges between vertices i and j are formed independently with probability

$$p_{ij} = \Theta\left(\min\left(\frac{W_i W_j}{\mu n}, 1\right)\right), \tag{2}$$

where $\mu = \mathbb{E}[W_i]$ is the expected vertex weight. This connection rule ensures that the expected degree of vertex i scales proportionally to its weight, i.e., $\mathbb{E}[\deg(i)] \approx W_i$ [37, 38], up to constant factors. This class includes the usual rank-one kernels with saturation, such as the Chung–Lu kernel $p_{ij} = \min(W_i W_j/(\mu n), 1)$ [37] and the generalized random graph kernel $p_{ij} = W_i W_j/(\mu n + W_i W_j)$ [39]. In the infinite-variance regime considered here, the saturation in these kernels when $W_i W_j \gg n$ induces structural degree-degree correlations. Thus, the model is not an uncorrelated Chung–Lu-type null model in the usual sense; the induced correlations are part of the mechanism studied in this paper.

Given a subset of k vertices with weights $(h_1, \ldots, h_k)$, the probability that they form a clique is asymptotically proportional to

$$f_n(h_1, \ldots, h_k) := \prod_{1 \le i < j \le k} \min\left(\frac{h_i h_j}{\mu n}, 1\right). \tag{3}$$

2.2 Large Deviations for Clique Counts

The number of k-cliques in IRG, denoted by $\mathcal{K}_n^{(k)}$, is a random variable influenced by two sources of randomness: the vertex weight assignment and the random edge formation. Understanding the typical and atypical behavior of $\mathcal{K}_n^{(k)}$ is crucial for characterizing the higher-order connectivity structure of the graph.

We begin by recalling a result from [40, 41] that estimates the expected number of k-cliques, denoted by $m_n^{(k)} := \mathbb{E}[\mathcal{K}_n^{(k)}]$.

Lemma 2.1 *Let* $\alpha \in (1, 2)$. *Then*

$$m_n^{(k)} = H\left(n\bar{F}(\sqrt{n})\right)^k (1 + o(1)),$$

for some constant $H > 0$. *In particular,* $m_n^{(k)}$ *is regularly varying with index* $k(2-\alpha)/2$.

This result shows that the expected number of cliques grows polynomially with n. Furthermore, it can be shown that when $\alpha \in (1, 2)$, most cliques are formed around a small number of vertices with weight approximately $\sqrt{n}$.

The central goal of this section is to establish a large deviation principle for $\mathcal{K}_n^{(k)}$. Specifically, for any $a > 0$, we aim to quantify the probability of observing significantly more cliques than expected,

$$\mathbb{P}(\mathcal{K}_n^{(k)} > (1+a)m_n^{(k)}).$$

The main result of this section states that, in the regime $\alpha > 2 - 2/k$, the upper tail of the k-clique count is governed by the probability of creating $k-2$ exceptionally large hubs. In particular,

$$\mathbb{P}(\mathcal{K}_n^{(k)} > (1+a)m_n^{(k)}) = \Theta(\mathbb{P}\,(\text{there are } k-2 \text{ vertices with weight} \ge c_a(n)))$$

where $c_a(n)$ is the necessary weight scale to create additionally many $am_n^{(k)}$ cliques. Precisely, one can understand what is the smallest size required from these $k-2$ hubs to generate an expected excess of $am_n^{(k)}$ cliques, defining

$$c_a(n) := \inf\left\{c : \binom{n}{2}\int_0^\infty \int_0^\infty f_n(x, y, c, ..., c)\, \mathrm{d}F(x)\, \mathrm{d}F(y) > am_n^{(k)}\right\}. \tag{4}$$

We now present in detail the result, matching asymptotic bounds for the order of magnitude of the large deviation probability for clique counts.

Theorem 2.2 *Let* $\alpha > 2 - 2/k$ *and* $a > 0$ *and let* $c_a(n)$ *be as in (4). Then, as* $n \to \infty$,

$$\mathbb{P}(\mathcal{K}_n^{(k)} > (1+a)m_n^{(k)}) = \Theta\left(n\mathbb{P}(W > c_a(n))\right)^{k-2}.$$

Moreover, the following lemma characterizes the asymptotic scaling of $c_a(n)$:

Lemma 2.3 *Let* $c_a(n)$ *be as in (4). Then,*

$$c_a(n) = K_2 a^{\frac{1}{2(\alpha-1)}} n^{\alpha^*}(1 + o(1)),$$

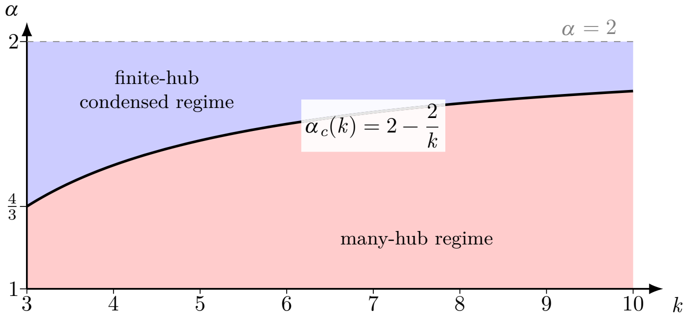

Fig. 1 Phase diagram for upper-tail deviations of k-clique counts. The critical line $\alpha = 2 - 2/k$ separates a finite-hub condensed regime, rigorously described by Theorem 2.2, from a conjectural many-hub regime below the line. Above the line, the excess of k-cliques is produced by $k-2$ exceptional hubs, and the deviation probability is polynomial. At the line, $c_a(n)$ reaches order n, so that finitely many hubs no longer suffice because of the saturation of the connection kernel

where $\alpha^* = 1 - \frac{2-k(2-\alpha)}{4(\alpha-1)}$, *for some positive constant* K_2.

Theorem 2.2 only holds in the regime $\alpha > 2 - 2/k$, and this induces a transition in the upper-tail mechanism, as depicted in Figure 1. In the large α regime, the deviation is realized by a finite condensed set of $k-2$ hubs. As α approaches $2 - 2/k$, the hub scale $c_a(n)$ approaches order n, so that a single hub cannot increase the clique count by further increasing its weight, because the connection probability (2) remains roughly constant when the weight increases further. This indicates that, below the threshold, a finite number of hubs can no longer produce the required excess of k-cliques, and that the deviation mechanism must instead involve an increasing number of atypical hubs. We therefore expect a different regime below $\alpha = 2 - 2/k$, with exponentially small deviation probabilities rather than the polynomial scale of Theorem 2.2. For the triangl case, $k = 3$, this picture has been proved rigorously in [36]: the threshold is $\alpha_c = 4/3$, above which the deviation is caused by one hub and has polynomial probability, below which it is caused by polynomially many hubs of order n and has semi-exponential probability, while at the boundary finitely many linear-size hubs may contribute. This provides evidence for the analogous many-hub scenario for larger cliques, although a complete treatment for general k remains open.

When choosing $a = a(n)$ growing with n, we can apply Theorem 2.2 to compute the probability of larger deviations, such as polynomial deviations. In particular, consider deviations of order $n^\gamma \gg m_n^{(k)}$, for $\gamma > 0$ sufficiently large. Lemma 2.1 yields $n^\gamma = am_n^{(k)}(1 + o(1))$, with $a = n^{\gamma - k(2-\alpha)/2}(1 + o(1))$. Then, Lemma 2.3 gives

$$c_a(n) = \Theta\left(a^{\frac{1}{2(\alpha-1)}} n^{1-\frac{2-k(2-\alpha)}{4(\alpha-1)}}\right) = \Theta\left(n^{1+\frac{\gamma-1}{2(\alpha-1)}}\right),$$

from which we can obtain the probability of deviations using Theorem 2.2. Interestingly, the scale at which $c_a(n)$ grows in n is independent of k, meaning that the required size of hubs necessary to achieve prescribed polynomial clique deviation is independent of the clique size.

2.3 Deviations for Subgraph Counts

Next, we investigate deviations of other subgraph counts than cliques. Given a small graph H on k vertices, we denote by $N(H)$ the number of times that H occurs as a subgraph in the

IRG. Furthermore, we denote the conditional expectation of $N(H)$ given the weights of all n vertices $W_1, W_2, ..., W_n$,

$$C^{(n)}(H) = \mathbb{E}\left[N(H) \mid \{W_i\}_{i\in[n]}\right]. \tag{5}$$

Our goal is to calculate

$$\mathbb{P}\left(C^{(n)}(H) > n^{\gamma}\right),$$

for some $\gamma > 0$.

We will focus on a specific class of subgraphs:

Assumption 1 Consider the optimization problem

$$B(H) := \max_{\boldsymbol{\beta}\in[0,1]^k} \sum_{i\in[k]}(1-\beta_i) + \sum_{(i,j)\in H} \min(\beta_i + \beta_j - 1, 0). \tag{6}$$

The optimal solution of (6) for graph H satisfies $\boldsymbol{\beta}^* \in [0, 1/\alpha)^k$.

Informally, the solution of (6) describes the optimal vertex weight profile in the typical regime. In fact, the objective function $B(H)$ encodes the polynomial scaling (with respect to n) of the expected number of copies of H formed on vertices with weights proportional to $\beta_1, ..., \beta_k$; thus, $B(H)$ identifies the expected subgraph count in the IRG, and the optimal $\boldsymbol{\beta}$ indicates which weight configuration yields maximum contribution.

Assumption 1 ensures that the random variable $N(H)$ concentrates around its mean. If this condition does not hold, then, with high probability, $N(H)$ has a smaller order of magnitude than its expectation, and therefore the probability of deviation from its mean is automatically considerably large. In particular, Assumption 1 ensures that

$$\frac{N(H)}{n^{k(2-\alpha)/2}} \xrightarrow{\mathbb{P}} K_H \tag{7}$$

for some $0 < K_H < \infty$ [40, Theorem 2.2]. Furthermore, by [40, Lemma 4.2], this assumption implies that the optimal solution is unique. The set of graphs satisfying Assumption 1 is quite large and contains, for example, all cliques and Hamiltonian graphs with an odd number of nodes. However, it is easy to check that not all graphs satisfy it. A simple example is the star graph, where a direct computation shows that the solution to (6) is achieved at value $\beta = 1$ for the star center. When k gets large, however, a large fraction of connected graphs is Hamiltonian, so that 1 is satisfied by a large proportion of graphs.

The upper-tail event $C^{(n)}(H) > n^{\gamma}$ can be analyzed through atypical weight profiles. The variables $\beta_1, ..., \beta_k$ encode the scales of the vertex weights involved in a copy of H, and the large-deviation problem amounts to identifying which such profile can produce the prescribed excess of subgraphs at the smallest probabilistic cost. In this sense, the optimization problem below can be viewed as a variational principle for the rare event. We define

$$R(H) := \max_{\boldsymbol{\beta}\in[0,1]^k} \sum_{i\in[k]} \min\left(1-\alpha\beta_i, 0\right) \tag{8}$$

$$\text{s.t.} \sum_{i\in V_H} \max\left(1-\alpha\beta_i, 0\right) + \sum_{\{i,j\}\in E_H} \min\left(\beta_i+\beta_j-1, 0\right) \geq \gamma. \tag{9}$$

The constraint in (9) selects the weight profiles that are capable of producing at least n^{γ} copies of H, and may therefore be interpreted as a microcanonical constraint on the subgraph excess. The objective in (8) measures the polynomial cost of creating the corresponding atypical weights. Since the objective is non-positive, maximizing $R(H)$ is equivalent to selecting the least costly feasible profile. In this sense, $R(H)$ plays the role of a variational free energy associated with the upper-tail event.

The following theorem shows that this variational problem determines the exponent of the upper-tail event $C^{(n)}(H) > n^\gamma$, and therefore identifies the dominant large-deviation mechanism.

Theorem 2.4 *Let $\gamma > k(2-\alpha)/2$, and let H satisfy Assumption 1. When (9) is feasible, then*

$$R(H) = \lim_{n\to\infty} \frac{\log(\mathbb{P}\left(C^{(n)}(H) > n^\gamma\right))}{\log(n)}. \tag{10}$$

Theorem 2.4 is a statement about the conditional expectation rather than about the actual number of copies $N(H)$. It therefore identifies the probability cost of producing a weight configuration for which the expected number of copies of H, conditional on the weights, exceeds n^γ. We expect the same exponent $R(H)$ to describe the upper tail of $N(H)$ itself. To prove this, however, one has to show that, once such a favorable weight configuration is present, the random edges produce, with high probability, of order n^γ copies of H. This requires controlling dependencies between overlapping copies of H, and no general argument is currently available for arbitrary subgraphs. We return to this point in Section 3.

Under Assumption 1, if $\gamma \le B(H)$ the solution of (8)+(9) is trivially $R(H) = 0$. In other words, n^γ is below the typical value, and there is no need to introduce extra-large hubs to observe n^γ copies of H in the IRG. The problem (8) + (9) is piece-wise linear and can be rewritten as a Mixed-Integer linear program (MILP). Before writing the general MILP, let us illustrate the encoding on the triangle K_3. In this case the nonlinear constraint contains the terms $\theta_i = \max(1-\alpha\beta_i, 0)$, for $i = 1, 2, 3$, and $\zeta_{ij} = \min(\beta_i+\beta_j-1, 0)$, for $1 \le i < j \le 3$, so that it reads

$$\theta_1 + \theta_2 + \theta_3 + \zeta_{12} + \zeta_{13} + \zeta_{23} > \gamma.$$

The objective contains $\delta_i = \min(1 - \alpha\beta_i, 0)$. The variables δ_i and ζ_{ij} are easy to encode linearly: since the problem is a maximization problem, the constraints $\delta_i \le 1 - \alpha\beta_i$, $\delta_i \le 0$ force $\delta_i = \min(1-\alpha\beta_i, 0)$ to be at the optimum. Similarly, $\zeta_{ij} \le \beta_i + \beta_j - 1$; $\zeta_{ij} \le 0$ forces $\zeta_{ij} = \min(\beta_i + \beta_j - 1, 0)$. The only term requiring binary variables is $\theta_i = \max(1-\alpha\beta_i, 0)$. For this, we introduce $b_i \in \{0, 1\}$ and impose $\theta_i \le 1 - \alpha\beta_i + 1 - b_i$ and $\theta_i \le b_i$. If $b_i = 0$, then $\theta_i \le 0$; if $b_i = 1$, then $\theta_i \le 1-\alpha\beta_i$. As this is a maximization problem, the optimization program will select b_i such that the maximum is achieved. Thus the binary variable selects the correct branch of the maximum. The same idea applies to the following general MILP formulation of (8)+(9):

$$\max \sum_{i\in V_H} \delta_i \tag{11}$$

$$\text{s.t.} \sum_{i\in V_H} \theta_i + \sum_{i,j\in E_H} \zeta_{ij} > \gamma \tag{12}$$

$$\delta_i \le -\beta_i\alpha + 1, \quad \delta_i \le 0 \tag{13}$$

$$\zeta_{ij} \le \beta_i + \beta_j - 1, \quad \zeta_{ij} \le 0 \tag{14}$$

$$\theta_i \le 1 - \beta_i\alpha + 1 - b_i \tag{15}$$

$$\theta_i \le b_i \tag{16}$$

$$b_i \in \{0, 1\} \tag{17}$$

Indeed, δ_i encodes the term $\min(1-\alpha\beta_i, 0)$, while ζ_{ij} encodes $\min(\beta_i+\beta_j-1, 0)$. The variables θ_i encode $\max(1-\alpha\beta_i, 0)$ through the binary variables b_i: when $b_i = 0$, the constraints force $\theta_i \leq 0$, while when $b_i = 1$, they force $\theta_i \leq 1 - \alpha\beta_i$. Since the constraint is easier to satisfy when the θ_i's are as large as possible, the optimal choice is $\theta_i = \max(1 - \alpha\beta_i, 0)$. Therefore the MILP is equivalent to the original piecewise-linear optimization problem.

Intuitively, the optimal value of β_i gives the scaling of the minimal weight necessary to have n^γ copies of H. That is, in the rare-event configuration producing at least n^γ copies of H, a vertex playing the role of vertex i in H typically has weight of order n^{β_i}.

The pattern in Figure 2 can be interpreted in terms of the *hub leverage* of the graph. The objective in (8) is the cost of lifting vertices above the typical scale, while the constraint in (9) measures how many copies can be generated by the resulting weight profile. If a small set of vertices touches many edges of H, or belongs to a dense core, then making these vertices hubs simultaneously increases several edge probabilities. Such graphs have relatively small deviations, and their $R(H)$ values are closer to zero. Conversely, if the edge constraints are spread more evenly over the graph, then no small hub set is sufficient; the optimizer has to lift several vertices to atypical scales, which leads to a more negative value of $R(H)$. Thus, the clusters visible in the figure are not determined by the number of edges alone, but by how effectively the topology of H can convert a small number of atypical hubs into many additional copies.

3 Discussion

Unconditional subgraph deviations. Theorem 2.4 concerns the conditional expectation $C^{(n)}(H) = \mathbb{E}[N(H) \mid \{W_i\}_{i\in[n]}]$, rather than the unconditional count $N(H)$. We expect the same exponent $R(H)$ to hold for $N(H)$. The reason is that the main cost should be the creation of the atypical weight profile selected by the variational problem. Once this profile is present, the expected number of copies, conditionally on the weights, is of order n^γ. The remaining edge randomness should therefore not change the polynomial exponent, provided that these copies are sufficiently concentrated around their conditional mean.

This can be made precise as a conditional second-moment criterion. Let $\beta^\star$ be an optimizing profile, and let $N_{\beta^\star}(H)$ be the number of copies of H using vertices in the weight classes prescribed by $\beta^\star$. Suppose that, on the event G_n that this profile is present,

$$\mathbb{P}(G_n) = n^{R(H)+o(1)}, \qquad \mathbb{E}[N_{\beta^\star}(H) \mid \{W_i\}_{i\in[n]}] \geq n^{\gamma+\varepsilon}$$

for some $\varepsilon > 0$, and

$$\mathrm{Var}(N_{\beta^\star}(H) \mid \{W_i\}_{i\in[n]}) = o\big(\mathbb{E}[N_{\beta^\star}(H) \mid \{W_i\}_{i\in[n]}]^2\big)\,.$$

Then, conditional Chebyshev's inequality gives

$$\mathbb{P}(N(H) > n^\gamma) \geq n^{R(H)+o(1)}.$$

Thus, the conjectured exponent gives the unconditional lower bound whenever the copies generated by the optimal weight profile are conditionally concentrated. The matching upper bound is more delicate. One has to rule out alternative edge-driven mechanisms and control dependencies between overlapping copies of H. For cliques, this is done in Theorem 2.2, using the special overlap structure exploited in Lemma 5.1. For a general H, no analogous universal argument is available.

Other subgraphs. Theorem 2.4 only holds for a specific class of subgraphs. For large k, this class is quite large, as it contains all Hamiltonian subgraphs among the others. We believe that there exists a different class of subgraphs for which a different approach is necessary.

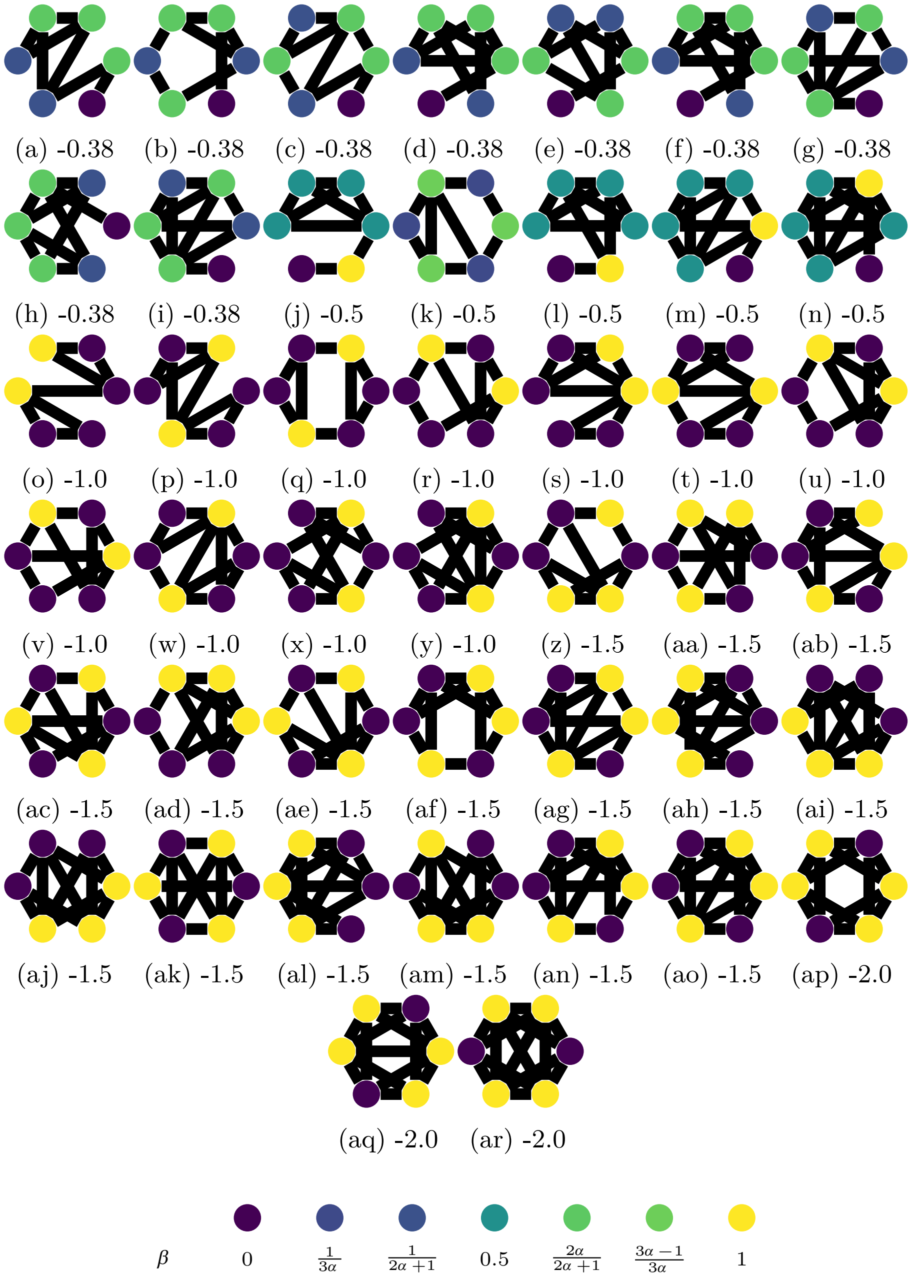

Fig. 2 Deviations for all connected subgraphs of size 6 satisfying Assumption 1, for $\alpha = 1.5$ and $\gamma = 2$. The color of each vertex indicates the optimal value of β_i in the variational problem (8)+(9), so that the corresponding vertex has weight scale n^{β_i} in the optimal rare-event configuration. The number below each graph is the value of $R(H)$; values closer to zero correspond to cheaper deviations. The clustering of the rates reflects how efficiently the topology of H can exploit a small set of atypically large vertices. Graphs in which a few vertices control many edge constraints have cheaper deviations, while graphs with more distributed connectivity require more vertices to be lifted to atypical weight scales and therefore have more negative rates

Indeed, many subgraph counts are known not to concentrate and to depend on the presence of hubs, so their large deviation properties are expected to differ significantly.

General U-statistics. Instead of focusing on subgraph counts, it can also be interesting to focus on the large deviations properties of general U-statistics. Given a graph G and an integer k, a U-statistic is a function of the kind

$$F_n(G) := \sum_{v_1,...,v_k \in V} h_n(v_1, ..., v_k)$$

for some local function $h_n : V^k \to \mathbb{R}$. U-statistics include subgraph counts or homomorphism densities, as well as other relevant global observables such as clustering coefficients, several centrality measures, and average shortest path distances. Any two vertices with the same weight are indistinguishable in IRG; thus, with a slight abuse of notation, we can define h_n as a function of the weights. It would be interesting to see the large deviation properties of these functions. We believe that our methods here could extend to specific types of U-statistics, for which

$$g(\beta_1, \dots, \beta_k) := \frac{\lim_{n\to\infty} \log(h_n(n^{\beta_1}, ..., n^{\beta_k}))}{\log(n)} \in (0, \infty), \tag{18}$$

for all $\boldsymbol{\beta} \in [0, 1]^k$, and

$$\lim_{n\to\infty} \frac{\log(h_n(n^{\beta_1}, \dots n^{\beta_k}))}{\log(h_n(b_1 n^{\beta_1}, \dots b_k n^{\beta_k}))} = 1, \tag{19}$$

for all fixed $b_1, \dots, b_k > 0$. Finally, we assume that

$$B_F := \max_{\boldsymbol{\beta}} \sum_{i=1}^{k} (1 - \alpha\beta_i) + g(\beta_1, ..., \beta_k) \tag{20}$$

has feasible solution $\boldsymbol{\beta}^* \in [0, 1/\alpha]^k$, which is a necessary condition for F_n to concentrate around its mean. In that case, (8)+(9) becomes

$$\begin{aligned} R_F := \max_{\boldsymbol{\beta}\in[0,1]^k} & \sum_{i\in V_H} \min(1 - \alpha\beta_i, 0) \\ \text{s.t.} & \sum_{i\in V_H} \max(1 - \alpha\beta_i, 0) + g(\beta_1, \dots, \beta_k) > \gamma. \end{aligned} \tag{21}$$

Whether this optimization problem is efficiently solvable depends on the shape of the function $g(\beta_1, \dots, \beta_k)$. For recent work on the case where h_n does not depend on n, we refer to [42].

Exponential deviations. When (8)+(9) is infeasible, then it is not possible to obtain the desired deviation using only a finite number of hubs, and a growing number of hubs is necessary to obtain the desired deviation. The optimization problem below should be understood as a heuristic extension of the finite-hub variational principle to this many-hub regime. A complete rigorous analysis is currently available for triangles, where the transition at $\alpha = 4/3$ has been characterized in [36]: above the threshold the deviation is caused by one hub and has polynomial probability, below it the deviation is caused by polynomially many linear-size hubs and has semi-exponential probability, while at the boundary finitely many linear-size hubs may contribute. For larger cliques, the analogous problem requires controlling configurations with polynomially many saturated hubs and their overlaps, and we leave this for future work.

The probability of having at least un^{θ} vertices of weight at least n^{δ} is

$$\mathbb{P}\left(un^{\theta} \text{ vertices of weight } > n^{\delta}\right) = \exp\left(-u(\theta - 1 + \alpha\delta)n^{\theta}\log(n)\right)(1 + o(1)), \quad (22)$$

when $\theta > 1 - \alpha\delta$. This equation shows that increasing θ influences the probability to a much larger extent than δ. We therefore first assume that $\delta = 1$. We again write an optimization problem similar to (9), where we maximize the probability that the hubs are present, subject to the constraint that they form the desired number of subgraphs. Now, conditionally on n^{β} vertices of weight proportionally to n to be present, the number of vertices of weight at least n^{β_i} scales as $n^{\max(1-\beta_i\alpha,\theta)}$. Indeed, as at least n^{θ} vertices of weight proportionally to n are present, this number is at least n^{θ}, and is otherwise dictated by the power-law distribution. The probability that a subgraph forms is similar to that in (9). This yields the optimization problem

$$\begin{aligned} &\min_{\beta\in[0,1]^k,\theta\geq 0} \quad \theta \\ &\text{s.t.} \sum_{i\in V_H} \max\left(1-\beta_i\alpha, \theta\right) + \sum_{i,j\in E_H} \min\left(\beta_i + \beta_j - 1, 0\right) > \gamma \end{aligned} \quad (23)$$

As (9), this optimization problem is infeasible when the desired deviation is not possible (for example, more than n^3 triangles). Again, this can be rewritten as a Mixed Integer Linear Program. When polynomial deviations suffice, setting $\theta = 0$ gives a feasible solution, as the constraint then reduces to the same one as in (9).

4 LD for Conditional Number of Cliques

An important ingredient in proving Theorem 2.2 is showing that the conditional expectation of $\mathcal{K}_n^{(k)}$ given the weights of all n vertices $W_1, W_2, ..., W_n$,

$$C_n = \mathbb{E}\left[\mathcal{K}_n^{(k)} \mid \{W_i\}_{i\in[n]}\right], \quad (24)$$

concentrates around its mean. This, combined with the following proposition on the large deviations properties of C_n, will prove Theorem 2.2.

Proposition 4.1 *Assume $\alpha > 2 - 2/k$. Then,*

$$\mathbb{P}\left(C_n > (1+a)m_n^{(k)}\right) = \Theta\left(n\mathbb{P}\left(W > c_a(n)\right)\right)^{k-2}. \quad (25)$$

The rest of this section will be devoted to proving Proposition 4.1.

We now analyze the number of k-cliques conditionally on the n weights. Let C_n be expected number of k-cliques, given weights $\{W_i\}_{i=1,\dots,n}$. We aim to get upper and lower bounds for $\mathbb{P}\left(C_n > (1+a)m_n^{(k)}\right)$. We will see that the most likely way to generate additional $a \cdot m_n^{(k)}$ cliques is to have a certain number of hubs in the graph. The conditional number of cliques is given by

$$C_n = \binom{n}{k}\int_0^{\infty}\cdots\int_0^{\infty} f_n(x_1, ..., x_k)\, \mathrm{d}F_n(x_1)\cdots \mathrm{d}F_n(x_k), \quad (26)$$

where F_n is the empirical weight distribution

$$F_n(x) = \frac{1}{n}\sum_{i}^{n} \mathbb{1}_{(W_i\leq x)}. \quad (27)$$

We make use of a convenient upper bound for the complementary empirical distribution function $\bar{F}_n$. Define $a(n) = \bar{F}^{-1}(1/n)n^{-\delta}$ and $b(n) = \bar{F}^{-1}(1/n)n^{\delta}$. In particular, $a(n)$ and $b(n)$ are regularly varying of index $1/\alpha - \delta$ and $1/\alpha + \delta$. Assume $\delta \in (0, 1/\alpha - 1)$, $c > 0$, $A > 0$ and let $E_n(A, c, \delta)$ be the event

$$\left\{ \sup_{x<a(n)} \frac{\bar{F}_n(x)}{\bar{F}(x)} \leq 1 + A, \sup_{x \in [a(n), b(n)]} \bar{F}_n(x) \leq (1 + A)\bar{F}(a(n)), \sup_{x>b(n)} \bar{F}_n \leq (1 + A)c/n \right\}.$$

By Proposition 2.2 in [36], $\mathbb{P}(E_n(A, c, \delta)) \geq 1 - o(n^{-\beta})$ for any fixed $A > 0$, as soon as $\lceil c \rceil > \beta/\alpha\delta$. In particular, on the event $E_n(A, c, \delta)$, the empirical distribution function can be bounded by

$$\bar{F}^*_n(x) = (1 + A) \left[\bar{F}(x)\mathbb{1}_{(x<a(n))} + \bar{F}(a(n))\mathbb{1}_{(a(n) \leq x<b(n))} + c/n\mathbb{1}_{(x \geq b(n))} \right]. \tag{28}$$

4.1 Contribution from Big Hubs

Assume the IRG contains h hubs. A hub is a vertex i with weight $W_i \geq b(n)$, which is regularly varying of index $\alpha_b \in (1/2, 1)$. We want to compute the contribution of these h hubs to the total number of cliques in the graph. Formally, let $\mathcal{B} = \{b_1(n), ..., b_h(n)\}$ be regularly varying of indices $\alpha_{b_1}, ..., \alpha_{b_h} \in (1/2, 1)$. Assume that the IRG contains h hubs with weights $\mathcal{B}$. Then, the expected number of cliques generated on these hubs is $\binom{n-h}{k-h} S^{(h)}_{\mathcal{B}}(n)$, with

$$S^{(h)}_{\mathcal{B}}(n) := \int_1^\infty \cdots \int_1^\infty f_n(b_1(n), ..., b_h(n), x_{h+1}, ..., x_k) \,\mathrm{d}F(x_{h+1}) \cdots \,\mathrm{d}F(x_k). \tag{29}$$

We now compute $S^{(h)}_{\mathcal{B}}(n)$ in three different cases for h.

Lemma 4.2 (Less than $k - 2$ hubs) *Let* $1 \leq h \leq k - 3$ *and* $\mathcal{B} = \{b_i(n)\}_{i=1,...,h}$. *Then*

$$S^{(h)}_{\mathcal{B}}(n) = O(n^{-\alpha(k-h)/2}).$$

Proof We have

$$\begin{aligned} S^{(h)}_{\mathcal{B}} &= \int_1^\infty \cdots \int_1^\infty f_n(b_1(n), ..., b_h(n), x_{h+1}, ..., x_k) \,\mathrm{d}F(x_{h+1}) \cdots \,\mathrm{d}F(x_k) \\ &\leq \int_1^\infty \cdots \int_1^\infty \prod_{h+1 \leq i<j \leq k} \left(1 \wedge \frac{x_i x_j}{\mu n}\right) \mathrm{d}F(x_{h+1}) \cdots \,\mathrm{d}F(x_k). \end{aligned} \tag{30}$$

Furthermore, the right-hand side in the bound of (30) is the probability that a clique of size $k - h$ is formed on $k - h$ uniformly chosen vertices. By Lemma 2.1, this quantity is $\Theta(n^{-\alpha(k-h)/2})$. □

Lemma 4.3 ($k - 1$ hubs) *Let* $\mathcal{B} = \{b_i(n)\}_{i=1,...,k-1}$. *Then*

$$S^{(k-1)}_{\mathcal{B}} = O\left(\frac{\max_i b_i(n)}{n}\right)^{\alpha}.$$

Proof Let $\bar{b} = \max_i b_i(n)$. Then

$$S^{(k-1)}_{\mathcal{B}} = \int_1^\infty f_n(b_1(n), ..., b_{k-1}(n), x) \,\mathrm{d}F(x)$$

$$= \int_1^\infty \prod_{i=1}^{k-1} \left(1 \wedge \frac{xb_i(n)}{\mu n}\right) \mathrm{d}F(x) \le \int_1^\infty \left(1 \wedge \frac{x\bar{b}}{\mu n}\right)^{k-1} \mathrm{d}F(x). \tag{31}$$

To compute the last integral, we split the integration domain

$$\begin{aligned} S_{\mathcal{B}}^{(k-1)} &\le \left(\frac{\bar{b}}{\mu n}\right)^{k-1} \int_1^{\mu n/\bar{b}} x^{k-1} \,\mathrm{d}F(x) + \int_{\mu n/\bar{b}}^\infty \mathrm{d}F(x) \\ &= \left(\frac{\bar{b}}{\mu n}\right)^{k-1} \frac{(\mu n/\bar{b})^{k-\alpha-1} - 1}{k-\alpha-1}(1+o(1)) + \bar{F}(\mu n/\bar{b}) \\ &= \Theta\left(\frac{\bar{b}}{n}\right)^{\alpha}. \end{aligned} \tag{32}$$

□

Lemma 4.4 ($k-2$ hubs) *Assume that $\mathcal{B} = \{b_i(n)\}_{i=1,\dots,k-2}$ are ordered increasingly and consider $S_{\mathcal{B}}^{(k-2)}$. When $b(n) = b_1(n) = \dots = b_{k-2}(n)$, we have*

$$S_{\mathcal{B}}^{(k-2)}(n) = \frac{K_1}{n}\left(\frac{b(n)}{n}\right)^{2(\alpha-1)} (1+o(1)) \tag{33}$$

for some $K_1 > 0$. In particular, when $b(n)$ is regularly varying of index α_b, $S_{\mathcal{B}}^{(k-2)}$ is regularly varying of index $2(\alpha-1)(\alpha_b-1)-1$.
Moreover, if $b_1(n) \ll b_{k-2}(n)$, then

$$S_{\mathcal{B}}^{(k-2)}(n) = o\left(\frac{1}{n}\left(\frac{b_{k-2}(n)}{n}\right)^{2(\alpha-1)}\right) \tag{34}$$

Before proving the statement, to illustrate the mechanism behind Lemma 4.4, consider the case $k = 3$. Then $k-2 = 1$, so the relevant rare structure consists of a single hub of weight $b = b(n)$, with $\sqrt{n} \ll b \ll n$. The expected number of triangles using this hub, up to the combinatorial factor, is

$$S_b^{(1)}(n) = \int_1^\infty \int_1^\infty \left(\frac{xy}{\mu n} \wedge 1\right)\left(\frac{xb}{\mu n} \wedge 1\right)\left(\frac{yb}{\mu n} \wedge 1\right) dF(x)dF(y).$$

The contribution of the region $xy > \mu n$ is at most $\mathbb{P}(XY > \mu n) = O(n^{-\alpha}\log n)$, which is negligible on the scale considered below. On the complementary region $xy < \mu n$, the first factor equals $xy/(\mu n)$, and the integral is asymptotically separable:

$$S_b^{(1)}(n) = \frac{1+o(1)}{\mu n}\left[\int_1^\infty x\left(\frac{xb}{\mu n} \wedge 1\right) dF(x)\right]^2.$$

Splitting the one-dimensional integral at $x = \mu n/b$, and using $dF(x) = \alpha x^{-\alpha-1}(1+o(1))dx$, gives

$$\begin{aligned} \int_1^\infty x\left(\frac{xb}{\mu n} \wedge 1\right) dF(x) &= \frac{b}{\mu n}\int_1^{\mu n/b} x^2 \, dF(x) + \int_{\mu n/b}^\infty x \, dF(x) \\ &= K\left(\frac{b}{n}\right)^{\alpha-1} (1+o(1)), \end{aligned}$$

for a constant $K > 0$. Hence

$$S_b^{(1)}(n) = \frac{K^2 + o(1)}{\mu n} \left(\frac{b}{n}\right)^{2(\alpha-1)}.$$

This is exactly the $k = 3$ instance of Lemma 4.4. The calculation also shows the geometry of the rare event: the hub alone is not enough to create triangles. It needs two additional vertices whose weights are large enough to connect to the hub, while the edge between those two vertices still contributes the factor $xy/(\mu n)$. The dominant contribution balances these two effects at the scale $x, y \approx n/b$.

Proof of Lemma 4.4 In this setting

$$S_{\mathcal{B}}^{(k-2)}(n) = \int_0^\infty \int_0^\infty f_n(b_1(n), ..., b_{k-2}(n), x, y)\, \mathrm{d}F(x)\, \mathrm{d}F(y). \tag{35}$$

Using the definition of f_n gives

$$S_{\mathcal{B}}^{(k-2)}(n) = \int_0^\infty \int_0^\infty \left(\frac{xy}{\mu n} \wedge 1\right) \prod_{i=1}^{k-2} \left(\frac{xb_i(n)}{\mu n} \wedge 1\right) \left(\frac{yb_i(n)}{\mu n} \wedge 1\right) \mathrm{d}F(x)\, \mathrm{d}F(y). \tag{36}$$

In the region $xy > \mu n$, the integrand is upper bounded by 1, so that the contribution of that region to $S_{\mathcal{B}}^{(k-2)}(n)$ can be upper bounded by $\mathbb{P}(xy > \mu n)$, which is $\Theta(n^{-\alpha}\log(n))$. In particular, from the Lemma assumption, $b_i(n)$ is slowly varying with index in $(1/2, 1)$ for any i, therefore $n^{-\alpha}\log(n)$ is negligible compared to the right hand sides of Equations (33), (34). Indeed, when $b_1(n) = ... = b_{k-2}(n) = b(n)$,

$$\frac{1}{n}\left(\frac{b(n)}{n}\right)^{2(\alpha-1)} > n^{-1}(n^{\delta-1/2})^{2(\alpha-1)} = \omega(n^{-\alpha+2\delta(\alpha-1)}) = \omega(n^{-\alpha}\log(n)),$$

for some $\delta > 0$, and similarly

$$\frac{1}{n}\left(\frac{b_{k-2}(n)}{n}\right)^{2(\alpha-1)} = \omega(n^{-\alpha}\log(n)).$$

Next, assume that $xy < \mu n$. First, observe that the integral in this region is asymptotically lower bounded by the integral restricted to $x < \sqrt{\mu n}$, $y < \sqrt{\mu n}$, and it is upper bounded asymptotically by the integral where the constraint $xy < \mu n$ is dropped. These two quantities are asymptotically identical. Therefore, we only need to solve the equivalent separable and symmetric integral:

$$S_{\mathcal{B}}^{(k-2)}(n) = \frac{1}{\mu n}\left[\int_1^\infty x \prod_{i=1}^{k-2}\left(\frac{xb_i(n)}{\mu n} \wedge 1\right) \mathrm{d}F(x)\right]^2 (1 + o(1)). \tag{37}$$

For each i, the minimum in the product is achieved at 1 when $x > \frac{\mu n}{b_i(n)}$. Hence, we can solve the integral by splitting the integration domain into sub-intervals $I_\ell := \left[\mu n/b_{\ell+1}(n), \mu n/b_\ell(n)\right]$ with $\ell = 0, ..., k-2$ (and conventionally $I_{k-2} = [0, \mu n/b_{k-2}(n)]$, $I_0 = [\mu n/b_1(n), \infty)$). Then,

$$S_{\mathcal{B}}^{(k-2)}(n) = \frac{1}{\mu n}\left(\int_{\mu n/b_1(n)}^\infty x\, \mathrm{d}F(x) + \sum_{\ell=1}^{k-2} \frac{\prod_{j=1}^{\ell} b_j(n)}{(\mu n)^\ell} \int_{I_\ell} x^{\ell+1}\, \mathrm{d}F(x)\right)^2 (1 + o(1))$$

$$= (1+o(1))\frac{1}{\mu n}\left(\frac{\alpha}{\alpha-1}\left(\frac{b_1(n)}{\mu n}\right)^{\alpha-1} + \right. \tag{38}$$

$$\left. \sum_{\ell=1}^{k-2} \frac{\alpha}{\ell-\alpha+1}\frac{\prod_{j=1}^{\ell} b_j(n)}{(\mu n)^{\alpha-1}}\left(b_\ell(n)^{-(\ell+1-\alpha)} - b_{\ell+1}(n)^{-(\ell+1-\alpha)}\right)\right)^2 \tag{39}$$

When $b_1(n) = b_2(n) = ... = b_{k-2}(n)$ most of the terms cancel out and

$$S_{\mathcal{B}}^{(k-2)}(n) = \frac{1}{\mu n}\left(\frac{\alpha(k-2)}{(\alpha-1)(k-\alpha-1)}\left(\frac{b_i(n)}{\mu n}\right)^{\alpha-1}\right)^2(1+o(1)) \tag{40}$$

which proves (33).

Next, assume that $b_1(n) \ll b_{k-2}(n)$. On the right-hand side of (38) the term

$$\left(\frac{b_1(n)}{\mu n}\right)^{\alpha-1} = o\left(\frac{b_{k-2}(n)}{n}\right)^{\alpha-1}$$

and for each term in the sum when $\ell > 1$ (using the bound $b_2(n) \le b_3(n) \le ... \le b_{k-2}(n)$)

$$\begin{aligned}\frac{\prod_{j=1}^{\ell} b_j(n)}{(\mu n)^{\alpha-1}}\left(b_\ell(n)^{-(\ell+1-\alpha)} - b_{\ell+1}(n)^{-(\ell+1-\alpha)}\right) &\le \frac{b_1(n)b_{k-2}(n)^{\ell-1}}{(\mu n)^{\alpha-1}} b_{k-2}(n)^{-(\ell+1-\alpha)} \\ &= \frac{b_1(n)}{b_{k-2}(n)}\left(\frac{b_{k-2}(n)}{\mu n}\right)^{\alpha-1} \\ &= o\left(\frac{b_{k-2}(n)}{n}\right)^{\alpha-1}\end{aligned} \tag{41}$$

Hence (34) follows. □

4.2 LD for Conditional Expected Cliques

Denote by $L_n(z)$ the number of vertices with a weight larger than z. We now show that whenever the vertex weights are all smaller than $\varepsilon c_a(n)$, $\mathcal{K}_k \le (1+a)m_n^{(k)}$ with high probability.

Proposition 4.5 (No hubs) *Assume that $\alpha > 2-2/k$ and let $c_a(n)$ be defined as in (4). Then, there exists $\varepsilon > 0$ such that*

$$\mathbb{P}\left(C_n > (1+a)m_n^{(k)}; L_n(\varepsilon c_a(n)) = 0\right) = o(n^{-\beta}) \tag{42}$$

for arbitrary β.

Proof We begin by decomposing

$$\begin{aligned}&\mathbb{P}\left(C_n > (1+a)m_n^{(k)}\,;\; L_n(\varepsilon c_a(n)) = 0\right) \\ &\quad \le \mathbb{P}\left(C_n > (1+a)m_n^{(k)}\,;\; L_n(\varepsilon c_a(n)) = 0\,;\; E_n(A,\delta,c)\right) + \mathbb{P}\left(E_n(A,\delta,c)^c\right).\end{aligned} \tag{43}$$

Since $\mathbb{P}(E_n(A,\delta,c)^c) = o(n^{-\beta})$ whenever $\lceil c\rceil > \beta/(\alpha\delta)$, it suffices to prove that

$$\mathbb{P}\left(C_n > (1+a)m_n^{(k)}\,;\; L_n(\varepsilon c_a(n)) = 0\,;\; E_n(A,\delta,c)\right) = 0 \tag{44}$$

for sufficiently small $A, \varepsilon > 0$.

On the event

$$\{L_n(\varepsilon c_a(n)) = 0\} \cap E_n(A, \delta, c),$$

the empirical tail satisfies

$$\bar{F}_n(x) \leq \bar{F}_n^*(x)\, \mathbb{1}_{\{x < \varepsilon c_a(n)\}},$$

where $\bar{F}_n^*$ is defined in (28). Since f_n is coordinate-wise non-decreasing, the conditional number of k-cliques is bounded by

$$C_n \leq \binom{n}{k} \int_1^\infty \cdots \int_1^\infty f_n(x_1, \ldots, x_k)\, dF_n^*(x_1) \cdots dF_n^*(x_k).$$

For each coordinate x_i, split the domain into:

small $(s) : x_i < a(n)$, medium $(m) : a(n) \leq x_i < b(n)$, large $(l) : b(n) \leq x_i < \varepsilon c_a(n)$.

Bounding medium coordinates by $b(n)$ and large coordinates by $\varepsilon c_a(n)$, we obtain

$$\frac{C_n}{(1+A)^k} \leq \sum_{\substack{s,m,l \geq 0 \\ s+m+l=k}} \binom{n}{k} \bar{F}(a(n))^m \left(\frac{c}{n}\right)^l$$

$$\times \int_0^\infty \cdots \int_0^\infty f_n \underbrace{x_1, \ldots, x_s}_{s \text{ small}}, \underbrace{b(n), \ldots, b(n)}_{m \text{ medium}}, \underbrace{\varepsilon c_a(n), \ldots, \varepsilon c_a(n)}_{l \text{ large}} dF(x_1) \cdots dF(x_s). \quad (45)$$

We now estimate each contribution according to s.
Case $s = k$ **(all small).** This term is at most $m_n^{(k)}$.
Case $2 < s < k$**.** Since $\bar{F}(a(n)) = O(n^{-1})$, we have

$$\bar{F}(a(n))^m (c/n)^l = O\,\bar{F}(a(n))^{m+l}.$$

The integral is bounded by $S_{\mathcal{B}}^{(k-s)}(n)$, so by Lemma 4.2,

$$\binom{n}{k} \bar{F}(a(n))^{m+l} S_{\mathcal{B}}^{(k-s)}(n) = O\left(n^k n^{-(m+l)(1-\delta\alpha)} n^{-s\alpha/2}\right) = O\left(n^{s(2-\alpha)/2+\delta\alpha(m+l)}\right).$$

Since $s < k$ and δ can be chosen arbitrarily small, these terms are $o(m_n^{(k)})$.
Case $s = 1$**.** Again $\bar{F}(a(n))^m (c/n)^l = O(\bar{F}(a(n))^{m+l})$. Lemma 4.3 yields

$$\binom{n}{k} \bar{F}(a(n))^{k-1} S_{\mathcal{B}}^{(k-1)}(n) = O\left(n^k n^{-(k-1)(1-\delta\alpha)} n^{-\alpha(1-\alpha^*)}\right) = O\left(n^{\alpha\delta(k-1)+1-\alpha(1-\alpha^*)}\right).$$

Since $\alpha\delta(k-1)$ can be made arbitrarily small and

$$1 - \alpha(1-\alpha^*) \leq \frac{k(\alpha-2)}{2}$$

for all $k \geq 3$ and $\alpha \in (2 - 2/k, 2)$, these terms are also $o(m_n^{(k)})$.
Case $s = 2$**.** If $m > 0$, Lemma 4.4 with $b_1(n) = b(n)$ and $b_{k-1}(n) = \varepsilon c_a(n)$ gives

$$\binom{n}{k} \bar{F}(a(n))^m (c/n)^l S_{\mathcal{B}}^{(k-2)}(n) = o\left(n \left(\frac{c_a(n)}{n}\right)^{2(\alpha-1)}\right) = o(m_n^{(k)}),$$

using Lemma 2.3. If $l = k-2$, Lemma 4.4 with all large coordinates equal to $\varepsilon c_a(n)$ yields

$$\binom{n}{k}(c/n)^l S_{\mathcal{B}}^{(k-2)}(n) = c^{k-2}\varepsilon^{2(\alpha-1)} a\, m_n^{(k)}(1+o(1)).$$

Summing all contributions in (45), we obtain, on $E_n(A,\delta,c) \cap \{L_n(\varepsilon c_a(n)) = 0\}$, that

$$C_n \le (1+A)^k 1 + c^{k-2}\varepsilon^{2(\alpha-1)} a m_n^{(k)}.$$

For sufficiently small $A, \varepsilon > 0$, the right-hand side is strictly less than $(1+a)m_n^{(k)}$, which contradicts the event in (44). Thus, the probability in (44) is zero, completing the proof. □

Next, we investigate the probability of exceeding $(1+a)m_n^{(k)}$ k-cliques when fewer than $k-2$ large hubs of size at least $c_a(n)$ are present.

Proposition 4.6 (Fewer than $k-2$ hubs) *Assume that $\alpha > 2 - 2/k$ and let $c_a(n)$ be defined as in (4). Then, for any $1 \le i < k-2$, there exists $\varepsilon > 0$ such that*

$$\mathbb{P}\left(C_n > (1+a)m_n^{(k)}; L_n(\varepsilon c_a(n)) = i\right) = O(n\mathbb{P}(W > c_a(n)))^{k-2}. \tag{46}$$

Proof Since $L_n(\varepsilon c_a(n))$ is binomial with parameters n and $p_n := \mathbb{P}(W > \varepsilon c_a(n))$ and $p_n = o(1/n)$, we have, for each fixed $i \in \mathbb{N}$,

$$\mathbb{P}(L_n(\varepsilon c_a(n)) = i) = (1+o(1))(n\mathbb{P}(W > \varepsilon c_a(n)))^i.$$

Hence it suffices to prove that, for every fixed i with $1 \le i < k-2$,

$$\mathbb{P}\left(C_n > (1+a)m_n^{(k)} \,\middle|\, L_n(\varepsilon c_a(n)) = i\right) = O((n\mathbb{P}(W > c_a(n)))^{k-i-2}). \tag{47}$$

Condition on the event $\{L_n(\varepsilon c_a(n)) = i\}$. For any k-clique that contains a certain number $j \in \{0, \ldots, i\}$ of these hubs, the remaining $k-j$ vertices must form a $(k-j)$-clique among the non-hub vertices. Let $C_n^{(k-j)}$ denote the total number of $(k-j)$-cliques in the graph. Then, on $\{L_n(\varepsilon c_a(n)) = i\}$, the total number of k-cliques satisfies

$$C_n \le \sum_{j=0}^{i} C_n^{(k-j)}.$$

Fix $\bar{a} := a/(i+1)$. If $C_n > (1+a)m_n^{(k)}$, then at least one of the $i+1$ terms in the above sum must exceed its mean by at least $\bar{a}m_n^{(k)}$. More precisely,

$$\begin{aligned} &\mathbb{P}\left(C_n > (1+a)m_n^{(k)} \,\middle|\, L_n(\varepsilon c_a(n)) = i\right) \\ &\le \mathbb{P}\left(C_n > (1+a)m_n^{(k)} \,\middle|\, L_n(\varepsilon c_a(n)) = 0\right) \\ &\quad + \sum_{j=1}^{i} \mathbb{P}\left(C_n^{(k-j)} > \bar{a}m_n^{(k)}\right). \end{aligned} \tag{48}$$

By Proposition 4.5, we can choose $\varepsilon > 0$ such that

$$\mathbb{P}\left(C_n > (1+\bar{a})m_n^{(k)} \,\middle|\, L_n(\varepsilon c_a(n)) = 0\right)$$

decays faster than any fixed polynomial in n. In particular, this term is $O((n\mathbb{P}(W > c_a(n)))^{k-i-2})$, since $n\mathbb{P}(W > c_a(n)) \to 0$ at a polynomial rate under our assumptions.

We now control the remaining terms in (48) by induction on the clique size. From [36], the statement of Proposition 2.2 is known for $k = 3$ (triangles), which serves as the base case. Fix $k \geq 4$ and assume that Proposition 2.2 holds for all clique sizes $3, \dots, k-1$. Then, for any $j \in \{1, \dots, k-2\}$ and any $\tilde{a} > 0$,

$$\mathbb{P}\left(C_n^{(k-j)} > (1+\tilde{a})m_n^{(k-j)}\right) = O((n\mathbb{P}\left(W > c_{\tilde{a}}^{(k-j)}(n)\right))^{k-j-2}), \tag{49}$$

where $m_n^{(k-j)}$ is the expected number of $(k-j)$-cliques, and $c_{\tilde{a}}^{(k-j)}(n)$ is the threshold (as in (4)) corresponding to having an expected excess of $\tilde{a}m_n^{(k-j)}$ $(k-j)$-cliques. Fix $j \in \{1, \dots, i\}$. We choose $\tilde{a} = \tilde{a}(j, n)$ such that

$$(1+\tilde{a})m_n^{(k-j)} = \bar{a}m_n^{(k)}. \tag{50}$$

By Lemma 2.1, we have $m_n^{(k-j)} = o(m_n^{(k)})$, so (50) implies

$$\tilde{a} = \frac{\bar{a}m_n^{(k)}}{m_n^{(k-j)}} - 1 = (1+o(1))\frac{\bar{a}m_n^{(k)}}{m_n^{(k-j)}} = \Theta(n^{j(1-\alpha/2)}).$$

Applying Lemma 2.3 to $(k-j)$-cliques, we obtain

$$c_{\tilde{a}}^{(k-j)}(n) = \Theta\left(\tilde{a}^{\frac{1}{2(\alpha-1)}} n^{1-\frac{2-(k-j)(2-\alpha)}{4(\alpha-1)}}\right) = \Theta\left(n^{1-\frac{2-k(2-\alpha)}{4(\alpha-1)}}\right) = \Theta(c_a(n)).$$

Hence $\mathbb{P}\left(W > c_{\tilde{a}}^{(k-j)}(n)\right)$ is of the same order as $\mathbb{P}(W > c_a(n))$, and by (49),

$$\begin{aligned}
\mathbb{P}\left(C_n^{(k-j)} > \bar{a}m_n^{(k)}\right) &= \mathbb{P}\left(C_n^{(k-j)} > (1+\tilde{a})m_n^{(k-j)}\right) \\
&= O((n\mathbb{P}\left(W > c_{\tilde{a}}^{(k-j)}(n)\right))^{k-j-2}) \\
&= O((n\mathbb{P}(W > c_a(n)))^{k-j-2}).
\end{aligned}$$

Since $j \leq i$ and $n\mathbb{P}(W > c_a(n)) \to 0$, we have

$$(n\mathbb{P}(W > c_a(n)))^{k-j-2} \leq (n\mathbb{P}(W > c_a(n)))^{k-i-2},$$

so, for each $j = 1, \dots, i$,

$$\mathbb{P}\left(C_n^{(k-j)} > \bar{a}m_n^{(k)}\right) = O((n\mathbb{P}(W > c_a(n)))^{k-i-2}).$$

Summing over $j = 1, \dots, i$ and combining with (48), we obtain

$$\mathbb{P}\left(C_n > (1+a)m_n^{(k)} \,\middle|\, L_n(\varepsilon c_a(n)) = i\right) = O((n\mathbb{P}(W > c_a(n)))^{k-i-2}),$$

which is exactly (47). This completes the induction and the proof of the proposition. □

Proof of Proposition 4.1 We decompose

$$\mathbb{P}\left(C_n > (1+a)m_n^{(k)}\right) = \sum_{i=0}^{n} \mathbb{P}\left(C_n > (1+a)m_n^{(k)}\,;\, L_n(\varepsilon c_a(n)) = i\right),$$

and estimate each term separately.

Case $i = 0$**.** By Proposition 4.5,

$$\mathbb{P}\left(C_n > (1+a)m_n^{(k)}\,;\, L_n(\varepsilon c_a(n)) = 0\right)$$

decays faster than any fixed polynomial in n.

Case $1 \leq i < k-2$**.** Proposition 4.6 gives

$$\mathbb{P}\left(C_n > (1+a)m_n^{(k)}\,;\ L_n(\varepsilon c_a(n)) = i\right) = O\left((n\mathbb{P}\left(W > c_a(n)\right))^{k-2}\right).$$

Case $i = k-2$**.** We use the trivial bounds

$$\begin{aligned}&\mathbb{P}\left(C_n > (1+a)m_n^{(k)}\,;\ L_n(c_a(n)) = k-2\right)\\ &\leq \mathbb{P}\left(C_n > (1+a)m_n^{(k)}\,;\ L_n(\varepsilon c_a(n)) = k-2\right)\\ &\leq \mathbb{P}\left(L_n(\varepsilon c_a(n)) = k-2\right).\end{aligned} \tag{51}$$

For the lower bound in (51), Lemma 4.4 applied with $b(n) = c_a(n)$ shows that the event $\{L_n(c_a(n)) = k-2\}$ already produces at least $(1+a)m_n^{(k)}$ cliques with probability asymptotically of order

$$(n\mathbb{P}\left(W > c_a(n)\right))^{k-2}.$$

Thus, the left-hand side of (51) is asymptotically bounded from below by a constant multiple of$(n\mathbb{P}\left(W > c_a(n)\right))^{k-2}$. For the upper bound, note that $L_n(\varepsilon c_a(n))$ is binomial with parameters n and

$$p_n := \mathbb{P}(W > \varepsilon c_a(n)).$$

$\mathbb{P}\left(W > c_a(n)\right)$ is regularly varying and

$$\mathbb{P}(W > \varepsilon c_a(n)) = (1+o(1))\mathbb{P}(W > c_a(n)),$$

because $\varepsilon > 0$ is fixed. Therefore,

$$\mathbb{P}\left(L_n(\varepsilon c_a(n)) = k-2\right) = \Theta\left((n\mathbb{P}\left(W > c_a(n)\right))^{k-2}\right).$$

Combining the upper and lower bounds in (51), we conclude that

$$\mathbb{P}\left(C_n > (1+a)m_n^{(k)}\,;\ L_n(\varepsilon c_a(n)) = k-2\right) = \Theta\left((n\mathbb{P}\left(W > c_a(n)\right))^{k-2}\right).$$

Case $i > k-2$**.** Since $L_n(\varepsilon c_a(n))$ is binomial with mean $n\mathbb{P}(W > \varepsilon c_a(n)) = o(1)$,

$$\mathbb{P}\left(L_n(\varepsilon c_a(n)) \geq k-3\right) = O\left((n\mathbb{P}\left(W > c_a(n)\right))^{k-3}\right).$$

Thus,

$$\mathbb{P}\left((C_n > (1+a)m_n^{(k)}\,;\ L_n(\varepsilon c_a(n)) > k-2\right)) \leq O\left((n\mathbb{P}\left(W > c_a(n)\right))^{k-3}\right).$$

Summing all contributions, the dominant term arises from $i = k-2$, and the proposition follows. □

5 LD for Number of Cliques (Proof of Theorem 2.2)

We now show that the probability that the number of cliques deviates from its expectation conditionally on the vertex weights decays exponentially fast. This relies on the fact that any k clique contains $k-2$ triangles sharing an edge. To do so, we first show that the probability that a single edge between two vertices with product of their weights $W_i W_j < \mu n$, appears in many k-cliques is small:

Lemma 5.1 *Suppose that* $1 < \alpha < 2$*. Then, when* $L > n^{(2-\alpha)(k-2)/2+\varepsilon}$ *for some* $\varepsilon > 0$*,*

$$\mathbb{P}\left(\{i, j\} \text{ appears in } \geq L \ k\text{-cliques} \mid W_i W_j < \mu n\right) \leq \exp(-c_1 L^{1/(k-2)}), \tag{52}$$

for some $c_1 > 0$*.*

Proof First of all, when an edge $\{i, j\}$ appears in L k-cliques, it has to appear in at least $L^{1/(k-2)}$ triangles. Indeed, any clique containing the edge $\{i, j\}$ contains $k-2$ triangles containing $\{i, j\}$. Therefore,

$$\begin{aligned} &\mathbb{P}\left(\{i, j\} \text{ appears in } \geq L \quad k-\text{cliques} \mid W_i W_j < \mu n\right) \\ &\leq \mathbb{P}\left(\{i, j\} \text{ appears in } \geq L^{1/(k-2)} \quad \text{triangles} \mid W_i W_j < \mu n\right) \\ &\leq \exp(-c_1 L^{1/(k-2)}), \end{aligned} \tag{53}$$

where the last step follows from [36, Lemma 4.4], which only holds for $L > n^{(2-\alpha)(k-2)/2+\varepsilon}$. □

Lemma 5.1 only bounds the number of shared edges in cliques for edges with incident vertices of relatively low degree. To bound the probability that the number of k-cliques is large, we also need to deal with overlapping cliques containing high-degree vertices. To do so, we define the event $E_w = \{W_1 = w_1, ..., W_n = w_n\}$ and we set $g_n = \sum_{i_1<i_2<\cdots<i_k} f_n(w_1, w_2, \ldots, w_k)$ to be the expected number of k-cliques conditional on the weights. Let $K_n^k(w)$ be the random variable describing the number of k-cliques conditionally on E_w. Finally,we set for $\zeta > 0$,

$$J(\eta) = (1+\zeta)\log(\zeta + 1/(1+\zeta))/3. \tag{54}$$

We now show that the probability that $K_n^k(w)$ is larger than g_n is small:

Lemma 5.2 *There exists an* $\varepsilon > 0$ *such that, for* $\zeta > 0$*, and all* $w = (w_1, ..., w_n)$*:*

$$\mathbb{P}\left(K_n^k(w) > (1+\zeta)g_n\right) \leq e^{-J(\zeta)g_n/n^{(2-\alpha)(k-2)/2+\delta}} + e^{-n^{\varepsilon}}. \tag{55}$$

Proof of Lemma 5.2 By Lemma 5.1, when $w_i w_j < \mu n$, and choosing $L = n^{(2-\alpha)(k-2)/2+\delta}$,

$$\mathbb{P}\left(\{i, j\} \text{ in } \geq n^{(2-\alpha)(k-2)/2+\delta} \quad k\text{-cliques}\right) \leq \exp(-K_1 n^{(2-\alpha)/2+\delta/(k-2)}), \tag{56}$$

for some $K_1 > 0$. This indicates that with probability at least $1-n^2\exp(-K_1 n^{(2-\alpha)/2+\delta/(k-2)}) \geq 1-\exp(-n^{\varepsilon})$, all edges between vertices of weights $w_i w_j < \mu n$ appear in at most $n^{(2-\alpha)(k-2)}$ k-cliques for some $\varepsilon > 0$. We now work on this event, which we call $\mathcal{E}'$.

We set $X_{i_1,...,i_k} = X'_{i_1,...,i_k}$ as the indicator that a k-clique is present on vertices $i_1, \ldots, i_k$. Furthermore, we set $X_{i_1,...,i_k(j_1,...,j_k)} = X_{i_1,...,i_k}$ when $|\{i_1, \ldots, i_k, j_1, \ldots, j_k\}| \geq 2k-2$. This corresponds to the setting in which two cliques overlap at a vertex or do not overlap at all. When $|\{i_1, \ldots, i_k, j_1, \ldots, j_k\}| < 2k-2$, we set $X_{i_1,...,i_k(j_1,...,j_k)} = X_{i_1,...,i_k}$ when the overlap of $i_1, \ldots, i_k$ and $j_1, \ldots, j_k$ occurs only at edges with $w_u w_v > \mu n$, and we set $X_{i_1,...,i_k(j_1,...,j_k)} = 0$ otherwise. Then, $\sum_{i_1,...,i_k} X_{i_1,...,i_k(j_1,...,j_k)}$ and $X_{j_1,...,j_k}$ are independent. Indeed, when two k-cliques do not overlap, or only overlap at a single vertex, their presence is independent, conditionally on the weights, as the edge indicators are independent conditionally on the weights. When the edge overlap(s) occur only at edges that are present with probability one, the presence of the two k-cliques is also independent conditionally on the weights. In all other cases, $X_{i_1,...,i_k(j_1,...,j_k)} = 0$, which is also independent of $X_{j_1,...,j_k}$.

We now bound the number of possible $X_{i_1,\ldots,i_k(j_1,\ldots,j_k)}$ that are manually put to zero. There are $\binom{k}{l}$ possible sets of l vertices of the k-cliques that can be part of an l-overlap with $j_1,\ldots,j_k$. As there is at least one edge with $w_u w_v < \mu n$ when $X_{i_1,\ldots,i_k(j_1,\ldots,j_k)}$ that is manually put to zero, there can be at most $n^{(3-\tau)/2+\delta/(k-2)}$ cliques that overlap with the l-set of vertices.

Summing over l, on the event $\mathcal{E}'$,

$$
\begin{aligned}
k!K_n^k(w) &= \sum_{i_1,\ldots,i_k} X_{i_1,\ldots,i_k} \\
&\leq \sum_{i_1,\ldots,i_k} X_{i_1,\ldots,i_k(j_1,\ldots,j_k)} + \left(\binom{k}{2} + \binom{k}{3} + \cdots + \binom{k}{k}\right) n^{(2-\alpha)(k-2)/2+\delta}. \qquad (57)
\end{aligned}
$$

Then, by [23, Lemma Theorem 3.1] with $a = \left(\binom{k}{2} + \binom{k}{3} + \cdots + \binom{k}{k}\right) n^{(2-\alpha)(k-2)/2+\delta}$, $t = (1+\zeta)g_n$ and $\lambda = g_n$ finishes the proof. □

Proof of Theorem 2.2 We begin with an asymptotic upper bound. Using Lemma 5.2, for $\zeta > 0$

$$
\begin{aligned}
\mathbb{P}\left(\mathcal{K}_n^{(k)} > (1+a)m_n^{(k)}\right) &\leq \mathbb{P}\left(\mathcal{K}_n^{(k)} \geq (1+\zeta)C_n\right) + \mathbb{P}\left(C_n/(1+\zeta) \geq \mathcal{K}_n^{(k)} \geq (1+a)m_n^{(k)}\right) \\
&\leq e^{-n^{\varepsilon}} + \mathbb{P}\left(C_n \geq (1+\zeta)(1+a)m_n^{(k)}\right) \\
&\leq H\left(n\mathbb{P}\left(W > c_a(n)\right)\right)^{k-2}, \qquad (58)
\end{aligned}
$$

for some constant H, where the last line follows from Proposition 4.1.

To obtain an asymptotic lower bound, we condition on the event that $k-2$ vertices have sufficiently large weights. In particular, let $\mathcal{K}_n^{(k)}(\delta, a)$ be the number of k-cliques in IRG, such that:

- either all nodes have weights at most $n^{1/2+\delta}$,
- or, two nodes have weights at most $n^{1/2+\delta}$ and the rest have weights at least $c_a(n)$.

Then,

$$
\mathbb{P}\left(\mathcal{K}_n^{(k)} \geq (1+a)m_n^{(k)}\right) \geq \left(n\mathbb{P}\left(W > c_a(n)\right)\right)^{k-2}\mathbb{P}\left(\mathcal{K}_n^{(k)}(\delta,a) \geq (1+a)m_n^{(k)}\right). \qquad (59)
$$

From here, it is sufficient to show that $\mathcal{K}_n^{(k)}(\delta,a)/m_n^{(k)}$ converges in probability to $(1+a)$. To this end, we employ a second-moment method. First, observe that

$$
\mathbb{E}\left[\mathcal{K}_n^{(k)}(\delta,a)\right] = \binom{n-(k-2)}{k} \int_0^{n^{1/2+\delta}} \cdots \int_0^{n^{1/2+\delta}} f_n(x_1,\ldots,x_k)\, \mathrm{d}F(x_1)\cdots\, \mathrm{d}F(x_k) \qquad (60)
$$

$$
+\frac{1}{2}(n-(k-2))^2 \int_0^{n^{1/2+\delta}} \int_0^{n^{1/2+\delta}} f_n(x_1,x_2,c_a(n),\ldots,c_a(n))\, \mathrm{d}F(x_1)\, \mathrm{d}F(x_2) \qquad (61)
$$

From the proof of Lemma 2.1, the right-hand side term in (60) asymptotically equals $m_n^{(k)}$. Instead, from Lemma 4.4, it follows by taking $b(n) = c_a(n)$ that the term in (61) is asymptotically equal to $am_n^{(k)}$. Summing up, $\mathbb{E}\left[\mathcal{K}_n^{(k)}(\delta,a)\right] = (1+o(1))(1+a)m_n^{(k)}$. Second, we investigate the concentration properties of $\mathcal{K}_n^{(k)}(\delta,a)$. We can write

$$
\mathcal{K}_n^{(k)}(\delta,a) = \binom{n-(k-2)}{k} \int_0^{n^{1/2+\delta}} \cdots \int_0^{n^{1/2+\delta}} f_n(x_1,\ldots,x_k)\, \mathrm{d}F_n(x_1)\cdots\, \mathrm{d}F_n(x_k) \quad (62)
$$

$$+ (n-(k-2))^2 \int_0^{n^{1/2+\delta}} \int_0^{n^{1/2+\delta}} f_n(x_1, x_2, c_a(n), ..., c_a(n))\, \mathrm{d}F_n(x_1)\, \mathrm{d}F_n(x_2) \tag{63}$$

where we recall that F_n denotes the empirical distribution function of weights. The distribution function F_n concentrates around F, that is, for any ζ and δ there exists ε such that (cfr. Proposition 2.2 in [36])

$$\mathbb{P}\left(\sup_{x<n^{1/2+\delta}} \left|\frac{\bar{F}_n(x)}{\bar{F}(x)} - 1\right| > \zeta\right) \le e^{-n^{\varepsilon}}. \tag{64}$$

On this event, $\mathcal{K}_n^{(k)}(\delta, a) \le (1+\zeta)^k \mathbb{E}\left[\mathcal{K}_n^{(k)}(\delta, a)\right] = (1+o(1))(1+\zeta)^k(1+a)m_n^{(k)}$, otherwise, we trivially bound $\mathcal{K}_n^{(k)}(\delta, a) \le n^k$. Thus,

$$\mathbb{E}\left[(\mathcal{K}_n^{(k)}(\delta, a))^2\right] \le (1+o(1))((1+\zeta)^k(1+a)m_n^{(k)})^2 + n^{2k} e^{-n^{\varepsilon}} = O((1+\zeta)^k(1+a)m_n^{(k)})^2 \tag{65}$$

Then, by Chebyshev's inequality, for any δ

$$\limsup_{n\to\infty} \mathbb{P}\left(\left|\frac{\mathcal{K}_n^{(k)}(\delta, a)}{\mathbb{E}\left[\mathcal{K}_n^{(k)}(\delta, a)\right]} - 1\right| > \delta\right) \le \frac{O(1+\zeta)^{2k} - 1}{\delta^2} \tag{66}$$

and the proof is concluded by taking the limit $\delta \to 0$. □

6 Proofs for General Subgraph Occurrence (Theorem 2.4)

We now investigate the impact of non-typical large hubs on subgraph counts in IRG. The appearance of a hub of size n^{β} with $\beta > 1/\alpha$ is a rare event. Indeed, the probability of a vertex weight to be of the order n^{β} is proportional to $n^{\alpha\beta}$, hence the expected number of vertices with such weights is $\Theta(n^{1-\beta\alpha}) \ll 1$.

By (7), $\mathbb{P}\left(N(H) \le n^{B(H)+\delta}\right) \to 1$, for any $\delta > 0$. Thus, $N(H)$ concentrates around $n^{B(H)}$. Still, $N(H)$ can be large when extra-large hubs appear in IRG. To formalize this, we investigate the extra contribution to the expected value $\mathbb{E}[N(H)]$ provided by extra-large hubs. Formally, let $\boldsymbol{\beta} = (\beta_1, ..., \beta_k) \in [0, 1]^k$ be a sequence where some entries might have value larger than $1/\alpha$, corresponding to extra-large hubs.

Proof of Theorem 2.4 **Lower bound.** Recall that

$$R(H) := \max_{\beta} \sum_{i\in[k]} \min(1-\alpha\beta_i, 0) \tag{67}$$

$$\text{s.t.} \sum_{i\in V_H} \max(1-\alpha\beta_i, 0) + \sum_{\{i,j\}\in E_H} \min\left(\beta_i + \beta_j - 1, 0\right) \ge \gamma. \tag{68}$$

Let $\boldsymbol{\beta}^*$ be the vector that optimizes $R(H)$. We now consider the number of copies of H on vertices with degrees proportional to $\boldsymbol{\beta}^*$, and will show that they cause n^{γ} copies of H with high probability. To be more precise, for fixed $0 < \varepsilon < 1$, we define

$$M^{(\boldsymbol{\beta})}(\varepsilon) := \left\{(v_1, ..., v_k) : w_{v_i} \in [1, 1/\varepsilon]n^{\beta_i}, \forall i \in [k]\right\}, \tag{69}$$

for any given $\boldsymbol{\beta} = (\beta_1, ..., \beta_k)$. Let $N(H, M^{(\boldsymbol{\beta})}(\varepsilon))$ be the conditional expectation of the number of subgraphs H appearing on vertex k-tuples in $M^{(\boldsymbol{\beta})}(\varepsilon)$, conditionally on the weights. Then,

$$N(H, M^{(\boldsymbol{\beta})}(\varepsilon)) = \sum_{\boldsymbol{v}\in M^{(\boldsymbol{\beta})}(\varepsilon)} \prod_{(i,j)\in E_H} \min\Big(\frac{w_{v_i}w_{v_j}}{n}, 1\Big). \tag{70}$$

For $\boldsymbol{\beta}^*$, this yields

$$N(H, M^{(\boldsymbol{\beta}^*)}(\varepsilon)) \geq |M^{(\boldsymbol{\beta}^*)}(\varepsilon)| \prod_{(i,j)\in E_H} n^{\min(\beta_i^*+\beta_j^*-1,0)}. \tag{71}$$

The number of vertices of weight within $[1, 1/\varepsilon]n^\beta$ is a Binomial random variable with mean $nn^{\beta\alpha}(1-\varepsilon^{(\alpha)})$. When $\beta \leq 1/\alpha$, then this mean is constant or tends to infinity when n grows, so that the number of vertices of weight within $[1, 1/\varepsilon]n^\beta$ is $\Theta_p(n^{1-\beta\alpha})$.

Let $\mathcal{T} = \{i \in [k] : \beta_i^* > 1/\alpha\}$. We then define the event

$$\mathcal{E} = \{\exists v_1, \dots, v_{|\mathcal{T}|} \in [n] : w_{v_i} \geq \varepsilon n^{\beta_i^*} \quad \forall i \in [|\mathcal{T}|]\}, \tag{72}$$

which implies that there is as least one set of $|\mathcal{T}|$ vertices with extra large degrees. Conditionally on $\mathcal{E}$,

$$\begin{aligned} |M^{(\boldsymbol{\beta}^*)}(\varepsilon)| &\geq \prod_{i:\beta_i^*\leq 1/\alpha} \Theta_p(n^{1-\alpha\beta_i^*}) \\ &= \Theta_p\Big(n^{\sum_{i\in[k]}\max(1-\alpha\beta_i^*,0)}\Big). \end{aligned} \tag{73}$$

Combining this with (71) yields that on $\mathcal{E}$, with high probability, for some $c_2 > 0$

$$N(H, M^{(\boldsymbol{\beta}^*)}(\varepsilon)) \geq c_2 n^{\sum_{i\in[k]}\max(1-\alpha\beta_i^*,0)+\sum_{\{i,j\}\in E_H}\min(\beta_i^*+\beta_j^*-1,0)} \geq c_2 n^\gamma, \tag{74}$$

where the last inequality follows from the constraint in the optimization problem for $R(H)$ in (68).

We now compute the probability of the event $\mathcal{E}$. When $\beta > 1/\alpha$,

$$\mathbb{P}\left(L_n(\varepsilon)n^\beta = k\right) \geq c_\varepsilon (n^{1+\alpha\beta})^k, \tag{75}$$

for some $c_\varepsilon > 0$. Thus, for some $\tilde{c} > 0$,

$$\mathbb{P}(\mathcal{E}) \geq \tilde{c} \prod_{i:\beta_i^*>1/\alpha} n^{1-\beta_i^*\alpha} = \tilde{c}n^{\sum_{i\in[k]}\min(1-\beta_i^*\alpha,0)} = \tilde{c}n^{R(H)}. \tag{76}$$

Therefore

$$\begin{aligned} \lim_{n\to\infty} \frac{\log(\mathbb{P}\left(C^{(n)}(H) > n^\gamma\right))}{\log(n)} &\geq \lim_{n\to\infty} \frac{\log(\mathbb{P}\left(N(H, M^{(\boldsymbol{\beta}^*)}(\varepsilon)) > n^\gamma\right))}{\log(n)} \\ &\geq \lim_{n\to\infty} \frac{\log(\mathbb{P}(\mathcal{E}))}{\log(n)} = R(H). \end{aligned} \tag{77}$$

Upper bound. We cover the interval $[1, n]$ with intervals of the type $I_\varepsilon(n^\zeta) = [\varepsilon, 1/\varepsilon]n^\zeta$ with $\zeta \in [0, 1]$. This is the same as covering $[0, \log(n)]$ with intervals of the type $[\zeta \log(n) + \log(\varepsilon), \zeta \log(n) - \log(\varepsilon)]$, which have length $-2\log(\varepsilon)$. Thus, to cover the interval $[1, n]$, we need $-2\log(n)/\log(\varepsilon)$ such logarithmic intervals. Therefore, $[0,1]^k$ can be covered with a covering $\mathcal{C}$ of size $(-2\log(n)/\log(\varepsilon))^k = \Theta(\log(n)^k)$. Then,

$$\begin{aligned}\mathbb{P}\left(C^{(n)}(H) > n^\gamma\right) &= \mathbb{P}\left(\sum_{\boldsymbol{\beta}\in\mathcal{C}} N(H, M^{(\boldsymbol{\beta})}(\varepsilon)) > n^\gamma\right)\\ &\le \mathbb{P}\left(\exists \boldsymbol{\beta}\in\mathcal{C} : N(H, M^{(\boldsymbol{\beta})}(\varepsilon)) > n^\gamma/\log(n)^k\right)\\ &\le \sum_{\boldsymbol{\beta}\in\mathcal{C}} \mathbb{P}\left(N(H, M^{(\boldsymbol{\beta})}(\varepsilon)) > n^\gamma/\log(n)^k\right), \end{aligned}\tag{78}$$

where the last step follows from the union bound. For $\beta < 1/\alpha$, the number of vertices with weight within $[\varepsilon, 1/\varepsilon]n^\beta$ is a Binomial random variable with diverging mean. Therefore, there exist $C_\varepsilon, c_\varepsilon > 0$ such that

$$\mathbb{P}\left(\sum_{i\in[n]} \mathbb{1}_{w_i\in[\varepsilon,1/\varepsilon]n^\beta} \ge C_\varepsilon n^{1-\beta\alpha}\right) \le \exp(-n^{c_\varepsilon}), \tag{79}$$

$\forall \boldsymbol{\beta}\in\mathcal{C}$. Let

$$\mathcal{E}_2 = \Big\{ \sum_{i\in[n]} \mathbb{1}_{w_i\in[\varepsilon,1/\varepsilon]n^\beta} \le C_\varepsilon n^{1-\beta\alpha} \quad \forall \boldsymbol{\beta}\in\mathcal{C}\Big\}. \tag{80}$$

Then,

$$\begin{aligned}&\mathbb{P}\left(C^{(n)}(H) > n^\gamma\right)\\ &\quad\le \sum_{\boldsymbol{\beta}\in\mathcal{C}} \mathbb{P}\left(|M^{(\boldsymbol{\beta})}(\varepsilon)| > \tilde{c}n^{\gamma-\sum_{\{i,j\}\in E_H}\min(\beta_i+\beta_j-1,0)}/\log(n)^k \mid \mathcal{E}_2\right) + \mathbb{P}(\mathcal{E}_2)\\ &\quad\le \sum_{\boldsymbol{\beta}\in\mathcal{C}} \mathbb{P}\left(\prod_{i:\beta_i>1/\alpha} |M^{(\beta_i)}(\varepsilon)| > \tilde{c}n^{\gamma-\sum_{\{i,j\}\in E_H}\min(\beta_i+\beta_j-1,0)-\sum_i(1-\alpha\beta_i)}/\log(n)^k\right)\\ &\qquad + \exp(-n^c).\end{aligned}$$

Now, when $\gamma - \sum_{\{i,j\}\in E_H}\min(\beta_i+\beta_j-1,0) - \sum_i(1-\alpha\beta_i) > 0$, then this calculates the probability of a polynomially growing number of vertices of degree more than $n^{1/\alpha}$, which happens with exponentially small probability. On the other hand, when $\gamma - \sum_{\{i,j\}\in E_H}\min(\beta_i+\beta_j-1,0) - \sum_i(1-\alpha\beta_i) \le 0$, then only a constant $|M^{(\beta_i)}(\varepsilon)|$ is required, which can be upper bounded by the probability of $M^{(\beta_i)}(\varepsilon)$ being non-empty.

As long as $\gamma - \sum_{\{i,j\}\in E_H}\min(\beta_i+\beta_j-1,0) - \sum_i(1-\alpha\beta_i) \ge 0$,

$$\mathbb{P}\left(M^{(\boldsymbol{\beta})}(\varepsilon) \neq \emptyset\right) \le \prod_{i:\beta_i>1/\alpha} (n^{1-\beta_i\alpha}) \le n^{R(H)}, \tag{81}$$

where the last inequality follows from (68). Now the requirement that $N(H, M^{(\boldsymbol{\beta})}(\varepsilon)) > n^\gamma/\log(n)^k$ rather than n^γ allows for vertices of weight $n^{\beta_i}/\log(n)$ rather than n^{β_i}, which adjusts the final probability to $n^{R(H)}\log(n)^{k\alpha}$. Combining this with (78) yields

$$\mathbb{P}\left(C^{(n)}(H) > n^\gamma\right) \le Cn^{-R(H)}\log(n)^{2k\alpha}, \tag{82}$$

finishing the proof. □

Funding Open access funding provided by Gran Sasso Science Institute - GSSI within the CRUI-CARE Agreement.

Data Availability Not applicable. No data was used for the preparation of this manuscript.

Declarations

Competing interests Nothing to disclose.

References

1. Vázquez, A., Pastor-Satorras, R., Vespignani, A.: Large-scale topological and dynamical properties of the internet. Phys. Rev. E **65**, 066130 (2002). https://doi.org/10.1103/PhysRevE.65.066130
2. Dhara, S., Hofstad, R., Leeuwaarden, J.S.H.: Critical percolation on scale-free random graphs: new universality class for the configuration model. Commun. Math. Phys. **382**(1), 123–171 (2021). https://doi.org/10.1007/s00220-021-03957-8
3. Friedrich, T., Krohmer, A.: Parameterized clique on inhomogeneous random graphs. Discrete Appl. Math **184**, 130–138 (2015). https://doi.org/10.1016/j.dam.2014.10.018
4. Heydari, H., Taheri, S.M.: Distributed maximal independent set on inhomogeneous random graphs. In: 2017 2nd Conference on Swarm Intelligence and Evolutionary Computation (CSIEC), pp. 77–82 (2017). https://doi.org/10.1109/CSIEC.2017.7940152 . IEEE
5. Janssen, A.J.E.M., Leeuwaarden, J.S.H., Shneer, S.: Counting cliques and cycles in scale-free inhomogeneous random graphs. J. Stat. Phys. **175**(1), 161–184 (2019). https://doi.org/10.1007/s10955-019-02248-w
6. der Hofstadvan, R., Hoorn, P., Litvak, N., Stegehuis, C.: Limit theorems for assortativity and clustering in null models for scale-free networks. Adv. Appl. Probab. **52**(4), 1035–1084 (2020). https://doi.org/10.1017/apr.2020.42
7. Yin, H., Benson, A.R., Ugander, J.: Measuring directed triadic closure with closure coefficients. Network Science **8**(4), 551–573 (2020). https://doi.org/10.1017/nws.2020.20
8. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U.: Network motifs: simple building blocks of complex networks. Science **298**(5594), 824–827 (2002). https://doi.org/10.1126/science.298.5594.824
9. Alon, U.: Network motifs: theory and experimental approaches. Nat. Rev. Genet. **8**(6), 450–461 (2007). https://doi.org/10.1038/nrg2102
10. Picard, F., Daudin, J.-J., Koskas, M., Schbath, S., Robin, S.: Assessing the exceptionality of network motifs. J. Comput. Biol. **15**(1), 1–20 (2008). https://doi.org/10.1089/cmb.2007.0137
11. der Hofstadvan, R., Janssen, A.J.E.M., Leeuwaarden, J.S.H., Stegehuis, C.: Local clustering in scale-free networks with hidden variables. Phys. Rev. E **95**(2), 022307 (2017). https://doi.org/10.1103/PhysRevE.95.022307
12. Sönmez, E., Stegehuis, C.: On the distances within cliques in a soft random geometric graph. J. Stat. Phys. **191**(3), 38 (2024). https://doi.org/10.1007/s10955-024-03254-3
13. Bianconi, G., Marsili, M.: Emergence of large cliques in random scale-free networks. Europhys. Lett. **74**(4), 740–746 (2006). https://doi.org/10.1209/epl/i2005-10574-3
14. Bianconi, G., Marsili, M.: Number of cliques in random scale-free network ensembles. Physica D **224**(1–2), 1–6 (2006). https://doi.org/10.1016/j.physd.2006.09.013
15. Bianconi, G., Marsili, M.: Loops of any size and hamilton cycles in random scale-free networks. J. Stat. Mech: Theory Exp. **2005**(06), 06005 (2005). https://doi.org/10.1088/1742-5468/2005/06/P06005

16. Bianconi, G., Marsili, M.: Effect of degree correlations on the loop structure of scale-free networks. Phys. Rev. E **73**(6), 066127 (2006). https://doi.org/10.1103/PhysRevE.73.066127
17. Bianconi, G., Capocci, A.: Number of loops of size *h* in growing scale-free networks. Phys. Rev. Lett. **90**(7), 078701 (2003). https://doi.org/10.1103/PhysRevLett.90.078701
18. Janson, S., Oleszkiewicz, K., Ruciński, A.: Upper tails for subgraph counts in random graphs. Israel J. Math. **142**(1), 61–92 (2004). https://doi.org/10.1007/bf02771528
19. Janson, S., Ruciński, A.: The deletion method for upper tail estimates. Combinatorica **24**(4), 615–640 (2004). https://doi.org/10.1007/s00493-004-0038-3
20. Kim, J.H., Vu, V.H.: Divide and conquer martingales and the number of triangles in a random graph. Random Struct. Algorithms **24**(2), 166–174 (2004). https://doi.org/10.1002/rsa.10113
21. DeMarco, B., Kahn, J.: Upper tails for triangles. Random Struct. Algorithms **40**(4), 452–459 (2012)
22. Chakrabarty, A., Hazra, R.S., Hollander, F.D., Sfragara, M.: Large deviation principle for the maximal eigenvalue of inhomogeneous erdős-rényi random graphs. J. Theor. Probab. **35**(4), 2413–2441 (2022). https://doi.org/10.1007/s10959-021-01138-w
23. Chatterjee, S.: The missing log in large deviations for triangle counts. Random Struct. Algorithms **40**(4), 437–451 (2012)
24. Coghi, F., Morand, J., Touchette, H.: Large deviations of random walks on random graphs. Phys. Rev. E **99**, 022137 (2019). https://doi.org/10.1103/PhysRevE.99.022137
25. Altarelli, F., Braunstein, A., Dall'Asta, L., Zecchina, R.: Large deviations of cascade processes on graphs. Phys. Rev. E **87**, 062115 (2013). https://doi.org/10.1103/PhysRevE.87.062115
26. Markering, M.: The large deviation principle for inhomogeneous erdős-rényi random graphs. J. Theor. Probab. **36**(2), 711–727 (2023). https://doi.org/10.1007/s10959-022-01181-1
27. Andreis, L., König, W., Langhammer, H., Patterson, R.I.: A large-deviations principle for all the components in a sparse inhomogeneous random graph. Probab. Theory Relat. Fields **186**(1), 521–620 (2023). https://doi.org/10.1007/s00440-022-01180-7
28. Dommers, S., Giardinà, C., Giberti, C., Hofstad, R.v.d.: Large deviations for the annealed ising model on inhomogeneous random graphs: spins and degrees. Journal of Statistical Physics 173(3), 1045–1081 (2018) https://doi.org/10.1007/s10955-018-2027-8
29. Stegehuis, C.: Degree correlations in scale-free random graph models. J. Appl. Probab. **56**(3), 672–700 (2019). https://doi.org/10.1017/jpr.2019.45
30. Yao, D., Hoorn, P., Litvak, N.: Average nearest neighbor degrees in scale-free networks. Internet Mathematics 1, (2018). https://doi.org/10.24166/im.02
31. Oliveira, R.I., Reis, G.H.: Interacting diffusions on random graphs with diverging average degrees: hydrodynamics and large deviations. J. Stat. Phys. **176**(5), 1057–1087 (2019). https://doi.org/10.1007/s10955-019-02332-1
32. Olvera-Cravioto, M.: PageRank's behavior under degree correlations. Ann. Appl. Probab. **31**(3), 1403–1442 (2021). https://doi.org/10.1214/20-aap1623
33. Kerriou, C., Mörters, P.: The fewest-big-jumps principle and an application to random graphs. Bernoulli **31**(3), 2525–2543 (2025). https://doi.org/10.3150/24-BEJ1816
34. Stegehuis, C., Zwart, B.: Scale-free graphs with many edges. Electron. Commun. Probab. **28**, 1–11 (2023). https://doi.org/10.1214/23-ECP567
35. Jorritsma, J., Zwart, B.: Large deviations of the giant component in scale-free inhomogeneous random graphs (2024). https://arxiv.org/abs/2407.01224
36. Stegehuis, C., Zwart, B.: Large deviations for triangles in scale-free random graphs. arXiv:2303.09198 (2023) https://doi.org/10.48550/ARXIV.2303.09198
37. Chung, F., Lu, L.: The average distances in random graphs with given expected degrees. Proc. Natl. Acad. Sci. **99**(25), 15879–15882 (2002). https://doi.org/10.1073/pnas.252631999
38. Stegehuis, C., Hofstad, R., Leeuwaarden, J.S.H., Janssen, A.J.E.M.: Clustering spectrum of scale-free networks. Phys. Rev. E **96**(4), 042309 (2017). https://doi.org/10.1103/physreve.96.042309
39. Britton, T., Deijfen, M., Martin-Löf, A.: Generating simple random graphs with prescribed degree distribution. J. Stat. Phys. **124**(6), 1377–1397 (2006). https://doi.org/10.1007/s10955-006-9168-x
40. Hofstad, R., Leeuwaarden, J.S.H., Stegehuis, C.: Optimal subgraph structures in scale-free configuration models. Ann. Appl. Probab. **31**(2), 501–537 (2021). https://doi.org/10.1214/20-AAP1580
41. Stegehuis, C., Hofstad, R.v.d., Leeuwaarden, J.S.: Variational principle for scale-free network motifs. Scientific reports 9(1), 6762 (2019) https://doi.org/10.1038/s41598-019-43050-8
42. Bakhshizadeh, M.: Exponential tail bounds and large deviation principle for heavy-tailed u-statistics. arXiv preprint arXiv:2301.11563 (2023) https://doi.org/10.48550/arXiv.2301.11563