

RESEARCH ARTICLE

Journal of the London
Mathematical Society

Local limit theorems for subgraph counts

Ashwin Sah | Mehtaab Sawhney

Department of Mathematics,
Massachusetts Institute of Technology,
Cambridge, Massachusetts, USA

Correspondence

Ashwin Sah, Department of Mathematics,
Massachusetts Institute of Technology, 182
Memorial Dr, Cambridge, MA 02139, USA.
Email: asah@mit.edu

Abstract

We introduce a general framework for studying anti-concentration and local limit theorems for random variables, including graph statistics. Our methods involve an interplay between Fourier analysis, decoupling, hypercontractivity of Boolean functions, and transference between ‘fixed-size’ and ‘independent’ models. We also adapt a notion of ‘graph factors’ due to Janson. As a consequence, we derive a local central limit theorem for connected subgraph counts in the Erdős–Renyi random graph $G(n, p)$, building on work of Gilmer and Kopparty and of Berkowitz. These results improve an anticoncentration result of Fox, Kwan, and Sauermaann and partially answer a question of Fox, Kwan, and Sauermaann. We also derive a local central limit theorem for induced subgraph counts, as long as p is bounded away from a set of ‘problematic’ densities, partially answering a question of Fox, Kwan, and Sauermaann. We then prove these restrictions are necessary by exhibiting a disconnected graph for which anticoncentration for subgraph counts at the optimal scale fails for all constant p , and finding a graph H for which anticoncentration for induced subgraph counts fails in $G(n, 1/2)$. These counterexamples resolve anticoncentration conjectures of Fox, Kwan, and Sauermaann in the negative. Finally, we also examine the behavior of counts of k -term arithmetic progressions in subsets of $\mathbb{Z}/n\mathbb{Z}$ and deduce a local limit theorem wherein the behavior is Gaussian at a global

scale but has nontrivial local oscillations (according to a Ramanujan theta function). These results improve on results of and answer questions of the authors and Berkowitz, and answer a question of Fox, Kwan, and Sauermann.

MSC (2020)

05C80, 60F05 (primary)

1 | INTRODUCTION

Random graph models have been studied in depth since their introduction by Erdős and Rényi [9]. Among the most important statistics of the random variables $G(n, p)$ and $G(n, m)$ are subgraph counts. Their study has been used, for instance, to construct large graphs avoiding certain subgraphs, among numerous other applications. In this paper, we will exclusively be focused on the regime where p is fixed and does not vary with n .

One natural question concerns the limiting distribution of the count X_H of some fixed graph, say H , within such a graph model. In the 1980s, several mathematicians studied this problem, for instance, showing that for any subgraph H we have convergence to a Gaussian in $G(n, p)$ for $p \in (0, 1)$. In particular, this proves that if μ_H, σ_H are the mean and standard deviation of X_H , the random variable counting the number of appearances of H , we have

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[a \leq \frac{X_H - \mu_H}{\sigma_H} \leq b \right] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx.$$

Note that these results imply asymptotically the probability that X_H lies in a given interval of size $\approx \sigma_H$. Using various methods, more quantitative forms of the above have resulted, with error terms of quality $O(n^{-\varepsilon})$. This allows one to control the probability that X_H lives in intervals of size $\sigma_H n^{-\varepsilon}$ for some $\varepsilon > 0$ (see, for example, [1, 25 28]).

A natural question to ask is if one can push this distributional control to pointwise control over probabilities. In particular, is it true that

$$\mathbb{P}[X_H = x] = \frac{1}{\sqrt{2\pi}\sigma_H} \exp \left[-\frac{(x - \mu_H)^2}{2\sigma_H^2} \right] + o_{n \rightarrow \infty} \left(\frac{1}{\sigma_H} \right)?$$

A result of this form is referred to as a *local central limit theorem*. Local limit theorems have a long history with the first central limit theorem, the De Moivre–Laplace central limit theorem in fact being a local central limit theorem. For independent integer-valued random variables, the seminal result of Gnedenko [13], specializes to that as long as there are no obvious modulus obstructions one in fact has a local central limit theorem. Local central limit theorems are now known in a large number of combinatorial situations including the size of the giant component of a random graph [3] (extended to hypergraphs in [2]) and the number of comparisons for merge-sort of a random permutation [15].

In recent studies of subgraph counts, more emphasis has been placed on the idea of *anticoncentration*. The first result in this direction was to Meka, Nyugen, and Vu [24] who proved, as a

consequence of a more general anticoncentration result, that each point probability is $n^{-1+o(1)}$. This was vastly improved for connected graphs by the recent work of Fox, Kwan, and Sauermann [11] in which they prove that

$$\sup_x \mathbb{P}[X_H = x] \leq n^{o(1)} \sigma_H^{-1}$$

for any connected subgraph H , using combinatorial methods. However, by design, anticoncentration on its own does not point toward a derivation of a local limit theorem.

Using careful analysis of characteristic functions on different regimes, Gilmer and Kopparty [12] showed that the triangle count in $G(n, p)$ indeed exhibits a local central limit theorem. This was improved by Berkowitz [5], who additionally proved the result for r -cliques [4]. We note that these works were preceded by earlier, closely related works, of Loeb, Matoušek, and Pangrác [23] demonstrating that triangle counts in $G(n, 1/2)$ are equidistributed modulo every fixed prime and the generalization to all connected graphs by Kolaitis and Kopparty [22].

We introduce a general framework toward proving such anticoncentration and local limit results, synthesizing many of the advances referenced above along with a notion of ‘graph factors’ used in work of Janson [18], and introduce a method of transferring results from $G(n, m)$ to $G(n, p)$. This transfer between ‘fixed-size’ and ‘independent’ models, as we see in the case of k -APs, will allow us to establish local limit theorems even when the pointwise behavior is not purely Gaussian.

Using our framework, we demonstrate optimal anticoncentration for connected subgraphs H by establishing a local central limit theorem, improving on the anticoncentration results of Fox, Kwan, and Sauermann [11]. In particular, we demonstrate an analogous result for the more delicate $G(n, m)$ model for $m/\binom{n}{2} \in (\lambda, 1 - \lambda)$, a generalization suggested in [11]. (This then transfers to the $G(n, p)$ model.) These results answer a question of Fox, Kwan, and Sauermann [11] in the connected case. As we will see later with k -APs, beyond simply being a natural generalization, analyzing the fixed-size model is key in establishing anticoncentration and local limit theorems in broader situations.

We then use similar techniques to study induced subgraph counts in random graphs of constant density. We demonstrate that as long as the density is sufficiently far away from a set of ‘problematic’ densities $P_{\text{crit}, H}$, the count of induced copies of H exhibit a local central limit theorem as well. This applies to both the $G(n, m)$ model if m is $\Theta(n^2)$ far from $p_{\text{crit}} \binom{n}{2}$ for $p_{\text{crit}} \in P_{\text{crit}, H}$ as well as the $G(n, p)$ model for constant $p \notin P_{\text{crit}, H}$. This again partially answers a conjecture of Fox, Kwan, and Sauermann [11] regarding the anticoncentration of induced subgraph counts.

We then demonstrate that in a certain sense the previous results are optimal by exhibiting a disconnected graph for which anticoncentration at the optimal scale fails (and, in fact, by a polynomial amount) for all constant p , as well as an induced graph H for which anticoncentration at the optimal scale fails (by a polynomial amount) in $G(n, 1/2)$. These counterexamples resolve conjectures of Fox, Kwan, and Sauermann [11] in the negative.

We also use our techniques to attack a structurally similar problem, that of length k arithmetic progressions (for fixed $k \geq 3$) in a random subset of $\mathbb{Z}/n\mathbb{Z}$. Significant attention has been given to understanding the large deviation behavior of $k\text{AP}(S)$ for random subsets S , particularly in the regime where the probability p that each element is chosen tends to 0. Here $k\text{AP}(S)$ denotes the number of k -term arithmetic progressions with all elements in S . Recently Harel, Mousset, and Samotij [14] (improving on earlier works of Warnke [30] and Bhattacharya, Ganguly, Shao, and Zhao [7]) found precise upper tail bounds for $k\text{AP}(S)$ in the sparse regime, while Janson and Warnke [20] proved lower tail bounds.

We prove a local central limit theorem for $k\text{AP}(S)$ where S is chosen to be uniform over sets of a fixed size $m \in (\lambda n, (1 - \lambda)n)$ and $k\text{AP}(S)$ denotes the number of k -term arithmetic progressions fully contained in S . For the model in which each element is picked independently, however, a local central limit theorem does not hold. Recent work by the authors and Berkowitz [6] demonstrates in a quantitative sense that the distribution behaves similarly to a convolution of discrete Gaussians at two scales, but does not prove a local limit theorem. By transferring our result in the fixed-size model, we can achieve this control, hence answering a question raised in [6]. This also demonstrates optimal anticoncentration for k -APs, a problem suggested in [11]. In this case, the local limit proven has Gaussian ‘large-scale’ structure, but exhibits nontrivial oscillations at a slightly smaller scale that are ultimately given in some sense by a theta function[†].

Notation

Throughout, we use $f \lesssim g$ to mean $|f| \leq Cg$ for some constant C , and o, O have their usual meanings. Subscripts denote dependence of the implicit constant.

1.1 | Main results

We now state the main results of this work. The first main result is a local central limit theorem for connected subgraphs in $G(n, p)$.

Theorem 1.1. *Fix a connected graph H . Choose $n \geq 1$ with $p \in (\lambda, 1 - \lambda)$ and sample a graph G from $G(n, p)$. Let the number of times H appears as a subgraph of G be X_H . Let μ_H be the mean and σ_H the standard deviation of X_H . Furthermore, define $Z_H = (X_H - \mu_H)/\sigma_H$ and $\mathcal{N}(z) = e^{-z^2/2}/\sqrt{2\pi}$. Then we have for any $\varepsilon > 0$ that*

$$\sup_{z \in (\mathbb{Z} - \mu_H)/\sigma_H} |\sigma_H \mathbb{P}[Z_H = z] - \mathcal{N}(z)| \lesssim_{\lambda, \varepsilon} n^{\varepsilon-1/2},$$

$$\sum_{z \in (\mathbb{Z} - \mu_H)/\sigma_H} |\mathbb{P}[Z_H = z] - \mathcal{N}(z)/\sigma_H| \lesssim_{\lambda, \varepsilon} n^{\varepsilon-1/2}.$$

Remark. This phrasing with $\varepsilon > 0$ can be made more quantitative via adding a large number of logarithm terms. We avoid specifying this dependence for the sake of clarity.

Our second main result is an anticoncentration result for induced subgraphs provided that they are sufficiently far away from a problematic set of densities.

Theorem 1.2. *Fix a graph H . There is an explicit set $\mathcal{P}_{\text{crit}}$ (of size at most $v(H)^2$) such that the following holds. Choose $n \geq 1$ with $p \in (\lambda, 1 - \lambda)$ and λ -separated from the set $\mathcal{P}_{\text{crit}}$ (that is, $|p - q| > \lambda$ for all $q \in \mathcal{P}_{\text{crit}}$) and sample a graph G from $G(n, p)$. Let the number of times H appears as an induced subgraph of G be X_H . Let μ_H be the mean and σ_H the standard deviation of X_H . Furthermore, define*

[†] In the sense of Ramanujan, for example, a function of the form $f(x) = \sum_{n \in \mathbb{Z}} e^{-(x+n)^2}$.

$Z_H = (X_H - \mu_H)/\sigma_H$ and $\mathcal{N}(z) = e^{-z^2/2}/\sqrt{2\pi}$. Then we have for any $\varepsilon > 0$ that

$$\sup_{z \in (\mathbb{Z} - \mu_H)/\sigma_H} |\sigma_H \mathbb{P}[Z_H = z] - \mathcal{N}(z)| \lesssim_{\lambda, \varepsilon} n^{\varepsilon-1/2},$$

$$\sum_{z \in (\mathbb{Z} - \mu_H)/\sigma_H} |\mathbb{P}[Z_H = z] - \mathcal{N}(z)/\sigma_H| \lesssim_{\lambda, \varepsilon} n^{\varepsilon-1/2}.$$

Remark. In fact we can show $|\mathcal{P}_{\text{crit}}| \lesssim v(H)$ and the truth is likely smaller, possibly even constant order. However, as we see next, the existence of this problematic set cannot be avoided.

As mentioned, we can prove analogous results to Theorems 1.1 and 1.2 for $G(n, m)$; we defer the statements to Theorems 4.2 and 4.4.

Given these two results there is a natural question: do we need to exclude disconnected subgraphs in Theorem 1.1 and the set of problematic densities in Theorem 1.2? In both cases, the answer is, surprisingly, yes. These examples resolve conjectures Fox, Kwan, and Sauermann [11] in the negative.

Theorem 1.3. *Let H be the disjoint union of 2 edges and fix $p \in (0, 1)$. Sample a graph G from $G(n, p)$, and let X_H count subgraphs of G isomorphic to H , with μ_H, σ_H the mean and standard deviation. Then*

$$\sup_x \mathbb{P}[X_H = x] \gtrsim n^{1/2} \sigma_H^{-1}.$$

Furthermore, there is a graph H' on 64 vertices such that if G is sampled from $G(n, 1/2)$, and if $X_{H'}, \mu_{H'}, \sigma_{H'}$ are defined analogously to before with respect to induced copies of H , then

$$\sup_x \mathbb{P}[X_{H'} = x] \gtrsim n^{1/2} \sigma_{H'}^{-1}.$$

Note that X_H for $H = K_2 + K_2$ (the union of 2 disjoint edges) is counting subgraphs and not homomorphisms. For homomorphism counts, it is easy to see that $H = K_2 + K_2$ forms a counterexample to anticoncentration, as every homomorphism count is a square number. The first part of Theorem 1.3 is not so straightforward, although this observation that the number of edges ‘mostly determines’ the number of counts offers a useful intuition for the failure of anticoncentration in this case.

Finally, we adapt our methods to handle the case of random k -term arithmetic progressions.

Theorem 1.4. *Fix $k \geq 3$ and choose $n \geq 1$ with $\gcd(n, (k-1)!) = 1$. Let $m/n \in (\lambda, 1-\lambda)$ and choose a uniformly random subset $\mathbb{Z}/n\mathbb{Z}$ of size m among all sets of this size; let \mathbf{x} be the indicator vector. Furthermore, let $\mu_k = \mathbb{E}[\text{kAP}(\mathbf{x})]$ and $\sigma_k = \text{Var}[\text{kAP}(\mathbf{x})]$. Finally, define $Z_k = (\text{kAP}(\mathbf{x}) - \mu_k)/\sigma_k$ and $\mathcal{N}(z) = e^{-z^2/2}/\sqrt{2\pi}$. Then we have for any $\varepsilon > 0$ that*

$$\sup_{z \in (\mathbb{Z} - \mu_k)/\sigma_k} |\sigma_k \mathbb{P}[Z_k = z] - \mathcal{N}(z)| \lesssim_{\lambda, \varepsilon} n^{\varepsilon-1/4}, \quad \sum_{z \in (\mathbb{Z} - \mu_k)/\sigma_k} |\mathbb{P}[Z_k = z] - \mathcal{N}(z)/\sigma_k| \lesssim_{\lambda, \varepsilon} n^{\varepsilon-1/4}.$$

As mentioned earlier, by transferring our result in the fixed-size model, we can also achieve a local limit theorem in the model where each element is chosen with a probability $p \in (0, 1)$,

hence answering a question raised by the authors and Berkowitz [6] and providing an optimal anticoncentration result, a direction suggested by Fox, Kwan, and Sauermaun [11]. We stress here that the distribution in the case where each element is chosen with probability p is not a pointwise Gaussian as a local *central* limit theorem in this model is false due to the results of Berkowitz and the authors [6]. Instead the distribution is a mixture of an infinite ensemble of Gaussians, reflected by a theta series, as hypothesized in [6] and discussed further in the final section of the paper. We defer the statement of this result to Theorem 7.8.

1.2 | Overview of methods

The methods of this work are Fourier analytic. For the sake of concreteness, we will restrict our attention to the case of connected subgraph counts although the proof for the remaining results are closely related. The main calculation is (essentially) demonstrating that

$$\mathbb{E}[e^{itZ_H}] \approx \mathbb{E}[e^{itZ}] \text{ for } |t| \leq \pi \cdot \sigma_H,$$

where $Z \sim \mathcal{N}(0, 1)$ is a standard Gaussian. Then the Fourier inversion formula for lattices gives an expression for the desired pointwise probabilities via Lemmas 2.4 and 2.5. We have three different ranges of $|t|$.

- (1) For $|t| \leq n^\epsilon$ we derive a sufficiently strong quantitative central limit theorem which can be used to provide bounds on the characteristic function of Z_H . In particular, Stein's method and the method of dependency graphs give a quantitative bound (Lemma 3.2) on the Wasserstein distance between Z_H and Z (in the independent model), which ultimately allows control of this range. This is closely related to the proof of a quantitative CLT (central limit theorem) given in [1]. A 'repair' argument coupling $G(n, m)$ and $G(n, p)$ (in the proof of Lemma 3.3) allows us to transfer bounds from $G(n, p)$ to $G(n, m)$. This argument is carried out to completion in Subsection 3.2.
- (2) For $n^\epsilon \leq |t| \leq \sigma_H n^{-\epsilon}$ we use decoupling arguments generalizing proofs of Berkowitz [4], relying on hypercontractive estimates to bound the typical sizes of coefficients of certain characteristic functions. However, since our random variables are constrained to live on a slice of the hypercube, and since we deal with non-complete H , various modifications are necessary. (By the slice of the hypercube we are referring to sampling a uniform point of $\{0, 1\}^n$ conditioned on having a fixed sum s .) In particular, we prove a decoupling lemma suitable for this situation (Lemma 2.10) and prove cancellation for characteristic functions of linear combinations of such random variables. These initial decoupling lemmas as well as several other useful preliminaries are discussed in Section 2.

It is in this range of $|t|$ where the connected assumption on the graph H is first invoked. The key point is that for a connected graph H , one can apply a decomposition into 'graph factors' and each 'graph factor' fluctuates at different scales. To demonstrate the necessary cancellation in this range we decompose the range $n^\epsilon \leq |t| \leq \sigma_H n^{-\epsilon}$ into a sequence of ranges (seen in the proof of Lemma 3.6). The key point is that for a connected graph H one has a graph factor for each necessary scale and this allows one to show cancellation of the Fourier coefficient (see the remark following Lemma 3.6) via the decoupling developed in Section 2. However, in order to adequately handle the decoupled Fourier coefficients arising in these

ranges various technical bounds are required and this technical work constitutes the meat of Subsection 3.4, including Lemma 3.5.

Finally, we note that the failure to have components ‘which fluctuate independently’ is ultimately the reason that a local central limit theorem does not necessarily hold for disconnected graphs (or for k -APs [6]). This argument is carried out in Section 6 and in particular the short Theorem 6.1 may serve a motivating example of why the proof isolates various ‘fluctuating’ components at each stage.

- (3) Finally, for $\sigma_H n^{-\varepsilon} \leq |t| \leq \pi \sigma_H$ we again use decoupling arguments related to those given in Berkowitz [4] (which in this range is closely related to the work of Kolaitis and Kopparty [22]), contained in Lemma 4.1 for subgraph counts and Lemma 4.3 for induced subgraph counts. However, in our case certain gymnastics are necessary in order to set up the decoupling method in order to hit the very top of the range, with complications arising which are not present in [4]. At this stage it is better to think about counting copies of the subgraphs directly, rather than decomposing into graph factors (as the integrality of the random variable comes heavily into play, since these Fourier coefficients, for example, can encode information such as equidistribution modulo small numbers).

When performing such analysis in general, we see that a notion of ‘graph factors’ stemming from work of Janson [18] is crucial, as it allows us to capture possible degeneracies as well as failures of local central limit theorems or even anticoncentration.

Finally, we remark that although a local central limit theorem fails for k -APs in the independent model, we are able to show one in the fixed-sum model (Theorem 1.4). Judicious transference arguments allows us to pin down a local limit theorem for the independent model (Theorem 7.8), down to the lower order fluctuations (which, as discussed earlier, behave somewhat like a theta function).

1.3 | Structure of the paper

In Section 2, we introduce necessary preliminaries for our estimates, including hypercontractivity, bounds for characteristic functions, decoupling techniques, and a notion of ‘graph factors’ that decompose graph statistics. In Section 3, we prove bounds for characteristic functions of graph statistics in $G(n, m)$ in a high degree of generality; we also explain how that is already enough to deduce optimal anticoncentration for very general graph polynomials. In Section 4, we specialize to connected subgraph counts and induced subgraph counts, proving $G(n, m)$ versions of Theorems 1.1 and 1.2. In Section 5, we transfer those versions to $G(n, p)$, establishing Theorems 1.1 and 1.2. In Section 6, we prove Theorem 1.3. Finally, in Section 7 we prove Theorem 1.4 as well as transfer to the corresponding result in the independent model.

2 | PRELIMINARIES

2.1 | Hypercontractivity

We will repeatedly require hypercontractive estimates which for us serve as tail bounds in a number of applications. These follow directly from theorems stated in O’Donnell’s book [26], although the results are originally due to Bonami, Beckner, Borell, and others.

Theorem 2.1 [26, Theorem 10.24]. *Let f be a polynomial in n variables of degree at most d , and let $X = (X_i)_{1 \leq i \leq n}$ be a sequence of mutually independent Boolean random variables such that each value is taken with probability at least λ . Then for any $t \geq (2e/\lambda)^{d/2}$,*

$$\mathbb{P}_X[|f(X)| \geq t \|f\|_2] \leq \lambda^d \exp\left(-\frac{d}{2e} \lambda t^{2/d}\right).$$

Here $\|f\|_2^2 = \mathbb{E}_X f(X)^2$.

Theorem 2.2 [26, Theorem 10.21]. *With the same hypotheses as above, if $q \geq 1$,*

$$\mathbb{E}_X[|f(X)|^{2q}] \leq (2q-1)^{dq} \lambda^{d(1-q)} \|f\|_2^{2q}.$$

Remark. We note that this theorem as stated in [4] has an incorrect exponent on λ , but it does not affect the results or proofs in any nontrivial fashion.

We often deal with a model in which the variables to which we wish to apply hypercontractivity are not independent, but constrained to have a fixed sum. Rather than use hypercontractivity on the slice, we use a trick of Jain [16] which allows one to transfer bounds from the independent model to a fixed sum model via a simple conditioning argument. See the proof of [16, Lemma 5.4] for an example of this trick.

2.2 | Converting from characteristic function to distributional control

Definition 2.3. Let X be a random variable. Then its *characteristic function* $\varphi_X : \mathbb{R} \rightarrow \mathbb{C}$ is defined to be $\varphi_X(t) := \mathbb{E}_X[e^{itX}]$.

Characteristic functions are very well-studied objects, and for sufficiently nice random variables the associated characteristic functions completely determine the random variable (for example, due to Lévy's continuity theorem). In particular, we will use the following inversion formula which bounds the L^∞ distance between the probability distribution of a lattice-valued random variable and the standard Gaussian in terms of characteristic functions. Let $\mathcal{N}(x)$ be the probability density function of the standard normal.

Lemma 2.4 [4, Lemma 2]. *Let X_n be a sequence of random variables supported in the lattices $\mathcal{L}_n = b_n + h_n\mathbb{Z}$, then*

$$\sup_{x \in \mathcal{L}_n} |h_n \mathcal{N}(x) - \mathbb{P}[X_n = x]| \leq h_n \left(\int_{-\frac{\pi}{h_n}}^{\frac{\pi}{h_n}} |\varphi_{\mathcal{N}(0,1)}(t) - \varphi_{X_n}(t)| dt + e^{-\frac{\pi^2}{2h_n^2}} \right).$$

We also use the following conversion to an L^1 distance estimate from the standard normal.

Lemma 2.5 [4, Lemma 3]. *Let X_n be a sequence of random variables supported in the lattice $\mathcal{L}_n := b_n + h_n\mathbb{Z}$, and with characteristic functions φ_n . Assume that there is $A > 0$ such that the following hold:*

- (1) $\sup_{x \in \mathcal{L}_n} |\mathbb{P}[X_n = x] - h_n \mathcal{N}(x)| < \delta_n h_n$;
 (2) $\mathbb{P}[|X_n| > A] \leq \epsilon_n$.

Then $\sum_{x \in \mathcal{L}_n} |\mathbb{P}[X_n = x] - h_n \mathcal{N}(x)| \leq 2A\delta_n + \epsilon_n + \frac{h_n}{\sqrt{2\pi A}} e^{-\frac{A^2}{2}}$.

2.3 | Estimates for characteristic functions

We will need a variety of estimates which will be used repeatedly in order to bound the characteristic functions of the random variables which we encounter. The first is essentially a well-known elementary estimate on the cosine function and although this precise result is not required for our setting it is present for comparison with the following estimate which obtains cancellation over a Boolean slice $\{\sum_{j=1}^n x_j = s\}$.

Lemma 2.6. *Let $Y \sim \text{Ber}(p)$. For any $|t| \leq \pi$ we have*

$$|\mathbb{E}[e^{itY}]| \leq 1 - \frac{2p(1-p)t^2}{\pi^2}.$$

Proof. Note that

$$\begin{aligned} |\mathbb{E}[e^{itY}]| &= (1 - 2p(1-p) + 2p(1-p)\cos t)^{1/2} \\ &\leq 1 - p(1-p)(1 - \cos t) \\ &\leq 1 - \frac{2p(1-p)t^2}{\pi^2}, \end{aligned}$$

where we have used that $1 - \cos(t) \geq 2t^2/\pi^2$ for $|t| \leq \pi$. □

The more difficult bound we will need is on characteristic functions when restricted to a slice of the hypercube. This is the critical estimate as it allows us to control a number of characteristic functions which will come up.

Lemma 2.7. *Let x_j be drawn with (x_1, \dots, x_n) uniform on $\{0, 1\}^n$ subject to $\sum_{j=1}^n x_j = s$. Furthermore, suppose that $p = s/n$ and $t \in \mathbb{R}$ is such that $|(a_j - a_k)t| \leq \pi$ for all $1 \leq j, k \leq n$. Then*

$$|\mathbb{E}[e^{it \sum_{j=1}^n a_j x_j}]| \leq (n+1) \exp[-2p(1-p)t^2 \text{Var}[a_j]n/\pi^2],$$

where $\text{Var}[a_j]$ is the variance of the random variable a_j , if J is an index uniformly drawn from $[n]$.

Proof. First note that $\binom{n}{k} \left(\frac{s}{n}\right)^k \left(\frac{n-s}{n}\right)^{n-k}$ is maximized at $k = s$ and since these values sum to 1 we have

$$\binom{n}{s} \left(\frac{s}{n}\right)^s \left(\frac{n-s}{n}\right)^{n-s} \geq \frac{1}{n+1}.$$

The key idea is that

$$\begin{aligned}
 |\mathbb{E}[e^{it \sum_{j=1}^n a_j x_j}]| &= \left| \frac{1}{2\pi i} \oint_{|z|=1} \frac{\prod_{j=1}^n (pe^{ita_j z} + (1-p))}{\binom{n}{s} \left(\frac{s}{n}\right)^s \left(\frac{n-s}{n}\right)^{n-s} z^{s+1}} dz \right| \\
 &\leq \frac{n+1}{2\pi} \left(2\pi \max_{|z|=1} \left| \prod_{j=1}^n (pe^{ita_j z} + (1-p)) \right| \right) \\
 &\leq (n+1) \left(\max_{|z|=1} \frac{1}{n} \sum_{j=1}^n |pe^{ita_j z} + (1-p)|^2 \right)^{n/2} \\
 &= (n+1) \left(\max_{|z|=1} \frac{1}{n} \sum_{j=1}^n (p^2 + (1-p)^2 + p(1-p)e^{ita_j z} + p(1-p)e^{-ita_j \bar{z}}) \right)^{n/2} \\
 &\leq (n+1) \left(1 - 2p(1-p) + 2p(1-p) \left| \frac{1}{n} \sum_{j=1}^n e^{ita_j} \right| \right)^{n/2} \\
 &= (n+1) \left(1 - 2p(1-p) + 2p(1-p) \left(\frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n \cos((a_j - a_k)t) \right)^{1/2} \right)^{n/2}.
 \end{aligned}$$

We now use the elementary facts that $\cos x \leq 1 - 2(x/\pi)^2$ for $|x| \leq \pi$, $1 + x \leq e^x$ for all $x \in \mathbb{R}$, and $\sqrt{1-t} \leq 1 - t/2$ for $t \in [-1, 1]$. Then it follows that

$$\begin{aligned}
 &(n+1) \left(1 - 2p(1-p) + 2p(1-p) \left(\frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n \cos((a_j - a_k)t) \right)^{1/2} \right)^{n/2} \\
 &\leq (n+1) \left(1 - 2p(1-p) + 2p(1-p) \left(\frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n 1 - \frac{2(a_j - a_k)^2 t^2}{\pi^2} \right)^{1/2} \right)^{n/2} \\
 &\leq (n+1) \left(1 - 2p(1-p) + 2p(1-p) \left(\frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n 1 - \frac{(a_j - a_k)^2 t^2}{\pi^2} \right) \right)^{n/2} \\
 &= (n+1) \left(1 - \left(\frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n \frac{2p(1-p)(a_j - a_k)^2 t^2}{\pi^2} \right) \right)^{n/2} \\
 &= (n+1) \left(1 - \frac{4p(1-p) \text{Var}[a_j] t^2}{\pi^2} \right)^{n/2} \\
 &\leq (n+1) \exp \left[\frac{-2p(1-p) \text{Var}[a_j] t^2 n}{\pi^2} \right]
 \end{aligned}$$

and the result follows. It is worth noting that in the second line, the expression within the square root can be verified to still be nonnegative, so the application of the inequalities above is valid. \square

2.4 | Decoupling methods

Definition 2.8. Define the operator α on functions of the form $f(X, Y_1, \dots, Y_k)$, which outputs the function $\alpha(f)$ given by

$$\alpha(f)(X, Y_1^0, Y_1^1, \dots, Y_k^0, Y_k^1) := \sum_{\mathbf{v} \in \{0,1\}^k} (-1)^{|\mathbf{v}|} f(X, Y^{\mathbf{v}}),$$

where $Y^{\mathbf{v}} = (Y_1^{v_1}, \dots, Y_k^{v_k})$. For the sake of notational simplicity define $\mathbf{Y} = (Y_1^0, Y_1^1, \dots, Y_k^0, Y_k^1)$.

This definition initially may seem opaque. However, a van der Corput-style Cauchy–Schwarz argument will allow us to utilize this definition in a critical way. Note that any function which is not dependent on all components of the random variables Y_i is in the kernel of the operator; following [4], we often refer to the remaining such functions as *rainbow* functions or terms. The key lemma we need will be a modification of the one in [4] and thus we repeat the original version; the proof is identical to the one given in [4].

Lemma 2.9. Let $k \geq 0$ and let (X, Y_1, \dots, Y_k) be mutually independent random variables. Let $\varphi(t) = \mathbb{E}_{X, Y_1, \dots, Y_k} e^{itf(X, Y_1, \dots, Y_k)}$. Then

$$|\varphi(t)|^{2^k} \leq \mathbb{E}_{\mathbf{Y}} |\mathbb{E}_X e^{it\alpha(f)(X, \mathbf{Y})}|,$$

where $\mathbf{Y} = (Y_1^0, Y_1^1, \dots, Y_k^0, Y_k^1)$, with Y_i^0 and Y_i^1 being independent samples of Y_i for all $i \in [k]$.[†]

Proof. We prove this by induction on k ; note that $k = 0$ is trivial. For induction define $X' = (X, Y_{k+1})$ and $f^*(X', Y_1, \dots, Y_k) = f(X, Y_1, \dots, Y_{k+1})$. Note that, applying the inductive hypothesis,

$$\begin{aligned} |\varphi(t)|^{2^k} &\leq \mathbb{E}_{Y_1^0, Y_1^1, \dots, Y_k^0, Y_k^1} |\mathbb{E}_{X'} e^{it\alpha(f^*)(X', Y_1^0, Y_1^1, \dots, Y_k^0, Y_k^1)}| \\ &\leq \mathbb{E}_{Y_1^0, Y_1^1, \dots, Y_k^0, Y_k^1} \mathbb{E}_X |\mathbb{E}_{Y_{k+1}} e^{it\alpha(f^*)(X', Y_1^0, Y_1^1, \dots, Y_k^0, Y_k^1)}|. \end{aligned}$$

By Cauchy–Schwarz and the triangle inequality we thus have

$$\begin{aligned} |\varphi(t)|^{2^{k+1}} &\leq \mathbb{E}_{Y_1^0, Y_1^1, \dots, Y_k^0, Y_k^1} \mathbb{E}_X |\mathbb{E}_{Y_{k+1}} e^{it\alpha(f^*)(X', Y_1^0, Y_1^1, \dots, Y_k^0, Y_k^1)}|^2 \\ &= \mathbb{E}_{Y_1^0, Y_1^1, \dots, Y_k^0, Y_k^1} \mathbb{E}_{(X, Y_{k+1}^0, Y_{k+1}^1)} e^{it\alpha(f^*)((X, Y_{k+1}^0), Y_1^0, Y_1^1, \dots, Y_k^0, Y_k^1)} e^{-it\alpha(f^*)((X, Y_{k+1}^1), Y_1^0, Y_1^1, \dots, Y_k^0, Y_k^1)} \\ &\leq \mathbb{E}_{Y_1^0, Y_1^1, \dots, Y_k^0, Y_k^1} \mathbb{E}_X e^{it\alpha(f)(X, Y_1^0, Y_1^1, \dots, Y_{k+1}^0, Y_{k+1}^1)} \end{aligned}$$

and the result follows. \square

[†] The inner expectation is technically conditional on a value of \mathbf{Y} , but since X and \mathbf{Y} are independent here and in the future we will abbreviate such expressions.

The need for a more complicated version of the above lemma stems from the fact that we will need to consider $f(X, Y_1, \dots, Y_k)$ where X, Y_j are tuples of Bernoulli random variables whose total sum has been conditioned on. Let A and B_j for $j \in [k]$ be sets which partition the index set $[n]$. Take independent Bernoulli random variables $(x_i)_{i \in [n]}$, then condition on the event $\sum_{i \in [n]} x_i = S$. Define random tuples $X = (x_i)_{i \in A}$ and $Y_j = (x_i)_{i \in B_j}$ and random variables $Z_0 = |X| = \sum_{i \in A} x_i$ and $Z_j = |Y_j| = \sum_{i \in B_j} x_i$. (Here $|X|$ is the sum of the vector, or equivalently the size of the corresponding set of nonzero indices in the Bernoulli setting.) Note that $Z_0 + \sum_{j \in [k]} Z_j = S$. The key idea is that given $(Z_j)_{1 \leq j \leq k}$ the random variables X, Y_j are conditionally independent.

Lemma 2.10. *Let $(x_i)_{i \in [n]}$ be a sequence of independent, identically distributed, Bernoulli random variables conditioned on the event $\sum_{i \in [n]} x_i = S$. Let A, B_j be sets as above with associated random variables X, Y_j, Z_0, Z_j and let $\varphi(t) = \mathbb{E}_{X, Y_1, \dots, Y_k} e^{itf(X, Y_1, \dots, Y_k)}$. Then*

$$|\varphi(t)|^{2k} \leq \mathbb{E}_{Z_1, \dots, Z_k, Y_1^0, Y_1^1, \dots, Y_k^0, Y_k^1} \left| \mathbb{E}_X \left[e^{it\alpha(f)(X, \mathbf{Y})} \middle| Z_1, \dots, Z_k \right] \right|,$$

where the variables Y_j^ℓ are conditionally independent $\{0, 1\}$ -vectors given $(Z_j)_{1 \leq j \leq k}$, equal to a uniform random vector with Z_j values of 1.

Remark. Note that the presence of Z_1, \dots, Z_k in the outer expectation is irrelevant, although it serves to clarify the joint distribution of the Y_j^b variables and is useful to the proof method.

Proof. By iterated Cauchy–Schwarz or Hölder’s inequality,

$$\begin{aligned} |\varphi(t)|^{2k} &= \left| \mathbb{E}_{Z_1, \dots, Z_k} \mathbb{E}_{X, Y_1, \dots, Y_k} \left[e^{itf(X, Y_1, \dots, Y_k)} \middle| Z_1, \dots, Z_k \right] \right|^{2k} \\ &\leq \mathbb{E}_{Z_1, \dots, Z_k} \left| \mathbb{E}_{X, Y_1, \dots, Y_k} \left[e^{itf(X, Y_1, \dots, Y_k)} \middle| Z_1, \dots, Z_k \right] \right|^{2k}. \end{aligned}$$

Since Z_0 is determined by Z_1, \dots, Z_k , we see that X, Y_1, \dots, Y_k are mutually conditionally independent given Z_1, \dots, Z_k . Hence, by Lemma 2.9, we conclude

$$|\varphi(t)|^{2k} \leq \mathbb{E}_{Z_1, \dots, Z_k} \left[\mathbb{E}_{Y_1^0, Y_1^1, \dots, Y_k^0, Y_k^1} \left[\left| \mathbb{E}_X [e^{it\alpha(f)(X, \mathbf{Y})} | Z_1, \dots, Z_k] \right| \middle| Z_1, \dots, Z_k \right] \right],$$

which equals the desired. □

2.5 | Graph factors

Finally, we define a notion of graph factors that will be critical for the remainder of our analysis regarding graphs. These notions are critical in previous work by Janson [17] and other results concerning the method of projections. Suppose a random graph is sampled, with the indicator of edge $e \in \binom{[n]}{2}$ denoted x_e . Let $\chi_e = (x_e - p)/\sqrt{p(1-p)}$. Note that if $x_e \sim \text{Ber}(p)$ then χ_e has mean 0 and variance 1.

Definition 2.11. Fix a graph H with no isolated vertices and an integer $n \geq |V(H)|$. Then define

$$\gamma_H(\mathbf{x}) = \sum_{\substack{E \subseteq \binom{[n]}{2} \\ E \simeq H}} \chi_E,$$

where $\chi_S = \prod_{e \in S} \chi_e$. Here \simeq denotes graph isomorphism, specifically between H and the graph spanned by the edges E . We call $\gamma_H(\mathbf{x})$ the *graph factor* corresponding to the graph H .

Remark. The empty graph K_0 has no isolated vertices, and appears as a subgraph of $\binom{[n]}{2}$ exactly once, so $\gamma_{K_0}(\mathbf{x}) = 1$.

The key property of this collection of functions is that they are orthogonal when the graph is sampled via $G(n, p)$, that is,

$$\mathbb{E}_{G(n,p)}[\chi_e \chi_{e'}] = \mathbb{1}_{e=e'}. \quad (2.1)$$

Additionally, the various graph theoretic functions we consider will be expressible in this basis. One of the key motivating results of Janson [17] is that in the $G(n, p)$ model, for a set of distinct connected graphs H_1, H_2, \dots, H_a , the vector $(\gamma_{H_b}(\mathbf{x}))_{b \in [a]}$ (scaled appropriately) approaches a vector of independent Gaussians.

2.6 | Graph notations

Given a graph G , we write $V(G)$ and $E(G)$ for the vertex and edge sets, $\bar{E}(G)$ for the set $\binom{V(G)}{2} \setminus E(G)$, and $v(G) = |V(G)|$, $e(G) = |E(G)|$, $\bar{e}(G) = |\bar{E}(G)|$.

3 | BOUNDS FOR GRAPH CHARACTERISTIC FUNCTIONS

In this section, we prove bounds for characteristic functions of quite general random variables associated to graphs, including connected subgraph counts and induced subgraph counts (aside from some potentially problematic edge-counts, as discussed in Section 1). We will prove these results in the $G(n, m)$ model.

3.1 | Setup

Let $\ell \geq 3$ be an integer and $\lambda \in (0, 1/2)$ be a real parameter. We define the notion of a well-behaved graph statistic of ‘degree ℓ ’, which will be the central object for our general analysis in this section.

Definition 3.1. An (ℓ, λ) -factor system is the following data. Let $\mathcal{H} = \{H_1, \dots, H_a\}$ be a set of (nonisomorphic) graphs on at most ℓ vertices with no vertex isolated. Further suppose that $\mathcal{H}' \subseteq \mathcal{H}$ is such that for each $3 \leq k \leq \ell$ it contains a connected graph on k vertices. Let \mathcal{H}_k for $0 \leq k \leq \ell$ denote the subset of \mathcal{H} with k vertices. Now, given such a factor system, let x_e for $e \in \binom{[n]}{2}$ be a

random graph drawn according to $G(n, m)$ such that $p = m/\binom{n}{2} \in (\lambda, 1 - \lambda)$, and let $\chi_e = (x_e - p)/\sqrt{p(1 - p)}$ as before. Now consider a linear combination of graph factors,

$$W = \sum_{H \in \mathcal{H}} n^{\ell - v(H)} \Delta_H \gamma_H(\mathbf{x}), \quad (3.1)$$

where the functions Δ_H are arbitrarily reals chosen such that $|\Delta_H| \leq 1/\lambda$ for $H \in \mathcal{H}$ and $|\Delta_H| \geq \lambda$ for $H \in \mathcal{H}'$. We call W an (ℓ, λ) -graph statistic with an associated factor system, or an (ℓ, λ) -statistic for short.

Remark. In applications, in particular counting subgraphs and induced subgraphs, we will choose the functions Δ_H to be specific functions of p (and to a lesser extent n), and $p = m/\binom{n}{2}$ will be constrained, so that the $H \in \mathcal{H}'$ terms satisfy a uniform lower bound as $n \rightarrow \infty$ as required by the definition. The purpose of \mathcal{H}' is that sometimes we may have some terms ‘self-cancel’ (for example, for certain induced subgraph counts) and be exactly zero in the $G(n, m)$ setting. \mathcal{H}' merely contains guaranteed non-canceling terms that we will need to establish the desired bounds.

Associated to an (ℓ, λ) -statistic W with factor system $(\mathcal{H}, \mathcal{H}')$, we will typically write

$$W_k = \sum_{H \in \mathcal{H}_k} n^{\ell - v(H)} \Delta_H \gamma_H(\mathbf{x}),$$

the *order k portion* of W . We now define a normalized version of the statistic W . Write σ for the standard deviation of W_3 in the $G(n, p)$ model. We easily see by orthogonality in the $G(n, p)$ model (2.1) that, since W_3 has a term from \mathcal{H}' , $\sigma/n^{\ell-3} = \Theta_\lambda(n^{3/2})$. Subsequently, we will drop all asymptotic dependence on $\mathcal{H}, \mathcal{H}'$ but keep the λ dependence, as it helps clarify the instances where we need coefficients of certain terms γ_H to be ‘of the correct size’.

As it turns out, σ and the standard deviation of W in the $G(n, m)$ model are essentially the same. In particular, $\sigma = \sigma_W(1 + O_{\lambda, \varepsilon}(n^{\varepsilon-1/2}))$ for any $\varepsilon > 0$. We will comment further on this phenomenon in Subsection 3.3; however, this is ‘as expected’ if we believe that the ‘only real effect’ of the $G(n, m)$ model is to constrain the sum of x_e (and hence χ_e). For technical reasons, it will be more convenient for us to work with σ initially, and to later use its closeness to σ_W to transfer any necessary results. Now we define

$$\mathcal{K} = \frac{W - W_0 - W_2}{\sigma}$$

to be the *normalized version* of the statistic W and let

$$\varphi_{\mathcal{K}}(t) = \mathbb{E} e^{it\mathcal{K}}$$

(note there is no graph on 1 vertex without isolated vertices, so W_1 never exists). This is the characteristic function that we will study in depth and is the key object of study. It is worth noting that \mathcal{K} does not even have mean zero in the $G(n, m)$ model, only in the $G(n, p)$ model. The true mean will also be discussed in Subsection 3.3.

To see this last remark, note that \mathcal{K} is a multilinear polynomial in the functions χ_e without a constant term, and the functions χ_e are independent mean 0 variance 1 random variables under

$G(n, p)$. We will use these facts about the $G(n, p)$ model often without comment. Finally, it is worth noting that on occasion we may switch between the probabilistic models $G(n, m)$ and $G(n, p)$. The switches between models will be clearly marked in the exposition; the results of this section are ultimately about the $G(n, m)$ model and \mathcal{K} is understood to be drawn from that model.

3.2 | Bounds for $|t| \leq n^\varepsilon$

\mathcal{K} should limit toward the Gaussian $\mathcal{N}(0, 1)$ as $n \rightarrow \infty$, which morally means the desired characteristic function bounds will hold for small t . For example, [17] shows such a central limit theorem for subgraph counts of $G(n, m)$. However, we need quantitative convergence in order to deduce the desired bounds. Furthermore, in $G(n, m)$, the expressions γ_H are not orthogonal unlike in $G(n, p)$, which complicates the usual techniques for proving effective central limit theorems.

Therefore, we will first quantitatively show that W_3/σ tends to $\mathcal{N}(0, 1)$ in the $G(n, p)$ model, and then transfer over to the $G(n, m)$ setting. Along the way we will also deal with the deviations introduced by the terms W_k for $k \geq 4$; we note that they will be lower order in size. The proof we give that W_3/σ limits to a Gaussian is similar to earlier work on a quantitative central limit theorem [1]. Also, the qualitative version follows from earlier work of [17]. However, we could not find work making the convergence in our setting quantitative and hence we provide a proof.

Lemma 3.2. *Let W be an (ℓ, λ) -statistic (with an associated factor system). Suppose the χ_e are drawn as in the $G(n, p)$ model, so that they are independent and identically distributed. Then*

$$\text{Wass}\left(\frac{W_3}{\sigma}, \mathcal{N}(0, 1)\right) \lesssim_\lambda \frac{1}{\sqrt{n}},$$

where $\text{Wass}(A, B) := \sup_{\text{Lip}(f) \leq 1} |\mathbb{E}[f(A) - f(B)]|$ denotes the Wasserstein metric.

Proof. We model our proof on the method of dependency graphs for proving quantitative central limit theorems. Let $Y = W_3/\sigma$. Let I be an indexing set for the monomials R_j , $j \in I$ in the expansion of W_3 , with the constant coefficients included. There are only three possibilities for I depending on the associated factor system: either I is the set of triangles within a complete graph on n vertices, the set of length 2 paths within a complete graph on n vertices, or the union of the two. An example of one of the monomials R_j would be $c\chi_{12}\chi_{23}$ for some $c \in \mathbb{R}$, if 1, 2, 3 are indices for vertices. We will treat the three cases for I uniformly.

For $j \in I$, let N_j be the set of $i \in I$ such that R_j, R_i are supported on sets of variables χ_e that overlap (that is, some χ_e is in both). Note $j \in N_j$. Let $Y_j = 1/\sigma \sum_{k \notin N_j} R_k$. Note that R_j, Y_j are independent. We also record that $\mathbb{E}[Y] = 0$ and $\sigma/n^{\ell-3} = \Theta_\lambda(n^{3/2})$, already noted earlier. This stems from the fact that the χ_e are centered and independent and identically distributed.

The key fact we will use is a version of Stein's lemma (see, for example, [27, Theorem 3.1]) which states

$$\text{Wass}(S, \mathcal{N}(0, 1)) \leq \sup\{|\mathbb{E}[f'(S) - Sf(S)]| : \|f\|_\infty \leq 1, \|f'\|_\infty \leq \sqrt{2/\pi}, \|f''\|_\infty \leq 2\}.$$

Now

$$\begin{aligned}
 \left| \mathbb{E}[Yf(Y) - f'(Y)] \right| &= \left| \frac{1}{\sigma} \sum_{j \in I} \mathbb{E}[R_j f(Y)] - \mathbb{E}[f'(Y)] \right| \\
 &\leq \frac{1}{\sigma} \left| \sum_{j \in I} \mathbb{E}[R_j(f(Y) - f(Y_j)) - R_j(Y - Y_j)f'(Y)] \right| \\
 &\quad + \left| \frac{1}{\sigma} \sum_{j \in I} \mathbb{E}[R_j(Y - Y_j)f'(Y)] - \mathbb{E}[f'(Y)] \right|
 \end{aligned}$$

since $\mathbb{E}[R_j] = 0$ and $R_j, f(Y_j)$ are independent. We now bound each of the terms separately. For the first,

$$\begin{aligned}
 &\frac{1}{\sigma} \left| \sum_{j \in I} \mathbb{E}[R_j(f(Y) - f(Y_j)) - R_j(Y - Y_j)f'(Y)] \right| \\
 &\leq \frac{1}{2\sigma} \left| \sum_{j \in I} \mathbb{E}[\|f''\|_{\infty} R_j(Y - Y_j)^2] \right| \lesssim_{\lambda} \frac{n^{\ell-3}}{\sigma} \left| \sum_{j \in I} \mathbb{E}[(Y - Y_j)^2] \right| \\
 &= \frac{n^{\ell-3}}{\sigma} \sum_{j \in I} \frac{1}{\sigma^2} \mathbb{E} \left[\left(\sum_{k \in N_j} R_k \right)^2 \right] \lesssim_{\lambda} n^4 (n^{\ell-3} \sigma^{-1})^3 \lesssim_{\lambda} n^{-\frac{1}{2}}.
 \end{aligned}$$

We used that $|I| = O(n^3)$ and $|N_j| = O(n)$, so that the inner expectation has $O(n)$ nonzero terms (by orthogonality (2.1), we can deduce $\mathbb{E}[R_i R_j] = 0$ for $i \neq j$). Here we also used that each R_j has coefficients of size $O_{\lambda}(n^{\ell-3})$.

For the second term note that

$$\frac{1}{\sigma} \sum_{j \in I} \mathbb{E}[R_j(Y - Y_j)] = \frac{1}{\sigma} \sum_{j \in I} \mathbb{E}[R_j Y] = \mathbb{E}[Y^2] = 1$$

since R_j, Y_j are independent. Thus,

$$\begin{aligned}
 &\left| \frac{1}{\sigma} \sum_{j \in I} \mathbb{E}[R_j(Y - Y_j)f'(Y)] - \mathbb{E}[f'(Y)] \right| = \left| \mathbb{E} \left[f'(Y) \left(\frac{1}{\sigma} \sum_{j \in I} R_j(Y - Y_j) - 1 \right) \right] \right| \\
 &\lesssim \left(\frac{1}{\sigma^2} \text{Var} \left[\sum_{j \in I} R_j(Y - Y_j) \right] \right)^{\frac{1}{2}} = \left(\frac{1}{\sigma^4} \sum_{\substack{k, m \in I \\ t \in N_k, s \in N_m}} (\mathbb{E}[R_k R_t R_m R_s] - \mathbb{E}[R_k R_t] \mathbb{E}[R_m R_s]) \right)^{\frac{1}{2}}
 \end{aligned}$$

$$= \frac{1}{\sigma^2} \left(\sum_{\substack{k,m \in I \\ t \in N_k \setminus k, s \in R_m \setminus m}} (\mathbb{E}[R_k R_t R_m R_s] - \mathbb{E}[R_k R_t] \mathbb{E}[R_m R_s]) \right. \\ \left. + 2 \sum_{\substack{k,m \in I \\ s \in N_m \setminus m}} (\mathbb{E}[R_k^2 R_m R_s] - \mathbb{E}[R_k^2] \mathbb{E}[R_m R_s]) + \sum_{k,m \in R} (\mathbb{E}[R_k^2 R_m^2] - \mathbb{E}[R_k^2] \mathbb{E}[R_m^2]) \right)^{\frac{1}{2}}$$

Note that $\mathbb{E}[R_m R_s] = 0$ if $m \neq s$ so the above simplifies to

$$\lesssim_{\lambda} \frac{1}{n^{2\ell-3}} \left(\sum_{\substack{k,m \in I \\ t \in N_k \setminus k, s \in N_m \setminus m}} \mathbb{E}[R_k R_t R_m R_s] + 2 \sum_{\substack{k,m \in I \\ s \in N_m \setminus m}} \mathbb{E}[R_k^2 R_m R_s] + \sum_{k,m \in I} (\mathbb{E}[R_k^2 R_m^2] - \mathbb{E}[R_k^2] \mathbb{E}[R_m^2]) \right)^{\frac{1}{2}}.$$

Now

$$\sum_{k,m \in I} \mathbb{E}[R_k^2 R_m^2] - \mathbb{E}[R_k^2] \mathbb{E}[R_m^2] \lesssim_{\lambda} n^4 (n^{\ell-3})^4 \quad (3.2)$$

as R_m^2 and R_k^2 are independent unless they intersect in an edge and there are $O(n^4)$ such configurations. Next,

$$\sum_{\substack{k,m \in I \\ s \in N_m \setminus m}} \mathbb{E}[R_k^2 R_m R_s] \lesssim_{\lambda} n^4 (n^{\ell-3})^4. \quad (3.3)$$

This is because in order for the term $\mathbb{E}[R_k^2 R_m R_s]$ to be nonzero, since $m \neq s$, either m or s must share an edge with k , say m . If m shares 2 edges or more, then it is contained within the same vertices as k , and we see that s must have this property as well, leading to $O(n^3)$ configurations. If it shares exactly 1 edge with k , then we see k, m span 4 vertices, and it is easy to see that s cannot introduce a new vertex else some edge will have multiplicity 1 in the term $R_k^2 R_m R_s$. This leads to $O(n^4)$ configurations, and hence justifies (3.3).

Finally,

$$\sum_{\substack{k,m \in I \\ t \in N_k \setminus k, s \in N_m \setminus m}} \mathbb{E}[R_k R_t R_m R_s] \lesssim_{\lambda} n^5 (n^{\ell-3})^4. \quad (3.4)$$

The reasoning is as follows. Note that k, t overlap on an edge as do m, s , and since $t \neq k$ and $s \neq m$ we have that $R_k R_t$ and $R_m R_s$ share some edge as well. So the total graph spanned is connected. Additionally, each edge must be covered at least twice to be nonzero. Now consider how many edges, with multiplicity, are spanned by a term $R_k R_t R_m R_s$, or equivalently the degree of the term.

If the degree is at most 9 for a nonzero term, each distinct edge is covered at least twice so we must have at most 4 total edges in the resulting graph. The configuration is connected so has at most 5 vertices. Hence, there are $O(n^5)$ configurations.

For degrees 10 and 11, in a nonzero term there must be at least one triangle present among the R factors, and there are at most 5 edges in the support. Again, we see that this leads to at most 5 vertices. So, we have $O(n^5)$ configurations again.

Finally, for degree 12, every R must correspond to a triangle. Hence, in a nonzero term R_k, R_t must be two triangles attached on an edge, and same for R_m, R_s . The only way to have a nonzero term is to superimpose these in some way. Thus, there are again 4 vertices, so $O(n^4)$ configurations. This justifies (3.4).

Using (3.2) to (3.4), we obtain the result. \square

Since $x \mapsto \exp(itx)$ is t -Lipschitz, this allows for a comparison of the characteristic functions of W_3/σ and $\mathcal{N}(0, 1)$ when drawing from the $G(n, p)$ model, which is essentially what we need in this range as outlined in Subsection 1.2. However, we need to transfer this information about the $G(n, p)$ model into estimates on the $G(n, m)$ model.

Lemma 3.3. *Let W be an (ℓ, λ) -statistic with normalized version \mathcal{K} . Then for all $\varepsilon > 0$ and $t \in \mathbb{R}$ we have*

$$|\mathbb{E}[e^{it\mathcal{K}}] - e^{-t^2/2}| \lesssim_{\lambda, \varepsilon} \frac{|t|}{n^{1/2-\varepsilon}} + \exp(-\Omega_\lambda((\log n)^2)).$$

Remark. Recall that for \mathcal{K} the edges are drawn from $G(n, m)$.

Proof. We first couple the independent and fixed-size models in order to compare their information. Sample $\chi' = (\chi'_{ij})_{1 \leq i < j \leq n}$ from the $G(n, p)$ model (identifying sets $\{i, j\}$ with edges). Then adjust a uniformly random subset of the edges (or non-edges, depending on if there are too many or too few edges) to obtain exactly m edges. Call the resulting random variable χ . By symmetry considerations, χ is distributed as if it came from the $G(n, m)$ model.

By Azuma–Hoeffding (see, for example, [19, Theorem 2.25]), with probability $1 - \exp(-\Omega_\lambda((\log n)^2))$ the number of edges adjusted is $O_\lambda(n \log n)$. Now let $Y = W_3/n^{\ell-3}$ (different than in the proof of Lemma 3.2), so that the coefficients are constant-order. We claim that for any $\varepsilon > 0$,

$$\mathbb{P}[|Y[\chi] - Y[\chi']| \geq n^{1+\varepsilon}] \lesssim \exp(-\Omega_\lambda((\log n)^2)). \quad (3.5)$$

By $Y[\chi']$ we abusively mean the polynomial Y evaluated at the (random) input χ' . Let S_0 be the set of edges that are changed and S_1 be the direction they changed in (we will later condition on ‘good’ realizations of this information to proceed). Given these two variables, note that χ_e, χ'_e are determined for $e \in S_0$, and $\chi_e = \chi'_e$ for $e \notin S_0$. We can therefore expand $Y[\chi] - Y[\chi']$ as a degree at most 3 polynomial in the variables $\chi_e, e \notin S_0$. The coefficients are functions of S_0, S_1 . Note that the monomials χ_T are supported only on $|T| \leq 3$. We therefore can write

$$Y[\chi] - Y[\chi'] = \sum_{|T| \leq 3, T \cap S_0 = \emptyset} c_T \chi_T. \quad (3.6)$$

We show that the sum of the squares of the coefficients of this random polynomial is small, with high probability over the randomness of S_0, S_1 . As noted, with high probability we have $|S_0| \lesssim_\lambda n \log n$. Next, the polynomial (3.6) only has terms that have interacted with χ_e for $e \in S_0$, that is,

is supported on χ_T for which there is a nonempty set $U \subseteq S_0$ of edges with $T \cup U$ a triangle or path of length 2. Thus, in fact (3.6) is degree at most 2.

Each $e \in S_0$ is in $O(n)$ triangles or paths of length 2, so the total number of nonzero coefficients in (3.6) is $O_\lambda(n^2 \log n)$ with high probability. We now bound the sum of squares of the coefficients c_T over various classes of set T . It is worth noting before we do this that conditional on $|S_0|, S_1$, we have that the set S_0 of edges is uniformly distributed among all possible subsets of the given size.

- (1) For $|T| = 2$, since the coefficients of Y are bounded by λ , we obtain a total contribution of $O_\lambda(n^2 \log n)$ with high probability since (3.6) has that many terms with high probability.
- (2) For $|T| = 1$, say $T = \{e\}$, this coefficient is (up to a constant depending on λ) bounded by the number of triangles or length 2 paths which contain e and such that the remaining edges are in S_0 . Now we fix some value $|S_0| \lesssim_\lambda n \log n$ and look over the randomness of S_0 . We care only about edges incident to e , and in particular, the desired coefficient is linearly bounded by how many edges there are.

Now, we need to bound the probability that there are too many such edges in S_0 incident to e . Note that S_0 is uniformly random conditional on its size, and $|S_0| \lesssim_\lambda n \log n$. Consider the random edge set S'_0 which is obtained by selecting each of the $\binom{n}{2}$ edges with probability $(\log n)/n$, say. By Chernoff, with probability at least $1 - \exp(-\Omega((\log n)^2))$ the number of resulting edges is greater than the upper bound on S_0 fixed above. Therefore, due to the symmetry of S_0 and S'_0 , we can create a coupling of these two variables such that $S_0 \subseteq S'_0$ except for $\exp(-\Omega((\log n)^2))$ of the probability space. Thus, it suffices to study the independent model, since in the coupling going from S_0 to S'_0 can only increase the edges incident to e .

Finally, in the independent model, by Chernoff, with probability at least $1 - \exp(-\Omega((\log n)^2))$ the desired edge count is $O_\lambda((\log n)^2)$ hence the coefficient is $O_\lambda((\log n)^2)$. The total contribution to the sum of squares is $O_\lambda(n^2(\log n)^4)$ with high probability.

- (3) For the constant term we are bounding the number of length 2 paths and triangles which appear in S_0 . Note that with high probability any edge of S_0 appears in $O_\lambda((\log n)^2)$ triangles or length 2 paths in total by the analysis in item 2 and thus this coefficient is bounded by $|S_0|(\log n)^2 \lesssim_\lambda n(\log n)^3$ with high probability, giving a contribution of $O_\lambda(n^2(\log n)^6)$ to the sum of squares.

In conclusion, using a union bound, we have shown that over the randomness of S_0, S_1 , with probability $1 - \exp(-\Omega_\lambda((\log n)^2))$, the sum of squares of the coefficients in (3.6) is $O_\lambda(n^2(\log n)^6)$, and also $|S_0| = O_\lambda(n \log n)$.

Now fix S_0, S_1 in such a case. We then sample every edge $e \notin S_0$ with a fixed sum depending on $|S_0|$ and S_1 , since this has the correct distribution for χ . If instead we sampled $e \notin S_0$ from $G(n, p)$, by hypercontractivity (Theorem 2.1) on the polynomial (3.6) we have $|Y[\chi'] - Y[\chi]| \geq n^{1+\varepsilon}$ with probability $\exp(-\Omega_\lambda(n^{\varepsilon'}))$. Since $|S_0| \lesssim_\lambda n \log n$, we see that we sample the correct total number of edges $e \notin S_0$ with probability $\exp(-O_\lambda((\log n)^2))$, which is much bigger than the failure probability above. Therefore, if we condition on drawing the right amount of edges, the quality of this estimate is preserved. Thus,

$$\mathbb{P}[|Y[\chi'] - Y[\chi]| \geq n^{1+\varepsilon} | S_0, S_1] \leq \exp(-\Omega_\lambda(n^{\varepsilon'}))$$

for every S_0, S_1 in one of the cases delineated above. The remainder of cases occur with probability $\exp(-\Omega_\lambda((\log n)^2))$ over the randomness of S_0, S_1 .

Overall, therefore, we have proven the desired (3.5):

$$\mathbb{P}[|Y[\chi'] - Y[\chi]| \geq n^{1+\varepsilon}] \lesssim \exp(-\Omega_\lambda((\log n)^2)).$$

To complete the proof we note that

$$\mathcal{K} = \frac{n^{\ell-3}Y}{\sigma} + \frac{\sum_{k \geq 4} W_k}{\sigma}.$$

Letting \mathcal{K}_{rem} denote the latter term, thus

$$\begin{aligned} |\mathbb{E}[e^{it\mathcal{K}}] - e^{-t^2/2}| &\leq \mathbb{E}[\min(|t\mathcal{K}_{\text{rem}}|, 2)] + |\mathbb{E}[e^{itn^{\ell-3}Y/\sigma}] - e^{-t^2/2}| \\ &\leq \mathbb{E}[\min(|t\mathcal{K}_{\text{rem}}|, 2)] + |\mathbb{E}[e^{itn^{\ell-3}Y/\sigma} - e^{itn^{\ell-3}Y'/\sigma}]| + |\mathbb{E}[e^{itn^{\ell-3}Y'/\sigma}] - e^{-t^2/2}| \\ &\lesssim_{\lambda, \varepsilon} |t|/n^{1/2-\varepsilon} + |t|/n^{1/2-\varepsilon} + |t|/\sqrt{n} + \exp(-\Omega_\lambda((\log n)^2)), \end{aligned}$$

where $Y = Y[\chi]$, $Y' = Y[\chi']$ are coupled together as described above. The last inequality is derived as follows. The third term comes from the fact that $x \mapsto e^{itx}$ is t -Lipschitz and using Lemma 3.2 (technically we have to separate into real and imaginary parts). The second term comes from $\sigma/n^{\ell-3} = \Theta_\lambda(n^{3/2})$ and the distance estimate (3.6) on Y versus Y' .

The first term comes from hypercontractivity (Theorem 2.1) once more: $\sigma\mathcal{K}_{\text{rem}}$ has standard deviation $O_\lambda(n^{\ell-2})$ in the $G(n, p)$ model, hence is at least n^ε times that with probability $\exp(-\Omega_\lambda(n^{\varepsilon'}))$. With probability $\Theta_\lambda(1/n)$ we draw exactly m edges, hence the integrity of this concentration estimate is maintained even if we condition on drawing m edges. The failure probability is absorbed into the fourth error term when multiplied by $\min(|t\mathcal{K}_{\text{rem}}|, 2) \in [0, 2]$. In the non-failure cases, $|t\mathcal{K}_{\text{rem}}| \lesssim_{\lambda, \varepsilon} |t|/n^{1/2-\varepsilon}$.

The final bound easily implies the result. \square

We will see later that for $|t| \leq n^\varepsilon$, Lemma 3.3 immediately establishes estimates of the desired quality.

3.3 | Mean and standard deviation considerations

This is the promised discussion of the difference between σ and σ_W , which is the true standard deviation in the model, and the difference between 0 and $\mathbb{E}[\sum_{3 \leq k \leq \ell} W_k | G(n, m)]$.

Lemma 3.4. *Let W be an (ℓ, λ) -statistic with $p = m/\binom{n}{2} \in (\lambda, 1 - \lambda)$ and define $\sigma^2 = \text{Var}[\sum_{k=3}^{\ell} W_k | G(n, p)]$ and $\sigma_W^2 = \text{Var}[W | G(n, m)]$. Then*

$$\sigma = \sigma_W(1 + O_{\lambda, \varepsilon}(n^{\varepsilon-1/2})).$$

Furthermore, if \mathcal{K} is the normalized version of W then

$$|\mathbb{E}[\mathcal{K} | G(n, m)]| \lesssim_{\lambda, p} n^{-1/2}.$$

Proof. We will first show that $\sigma = \sigma_W(1 + O_{\lambda,\varepsilon}(n^{\varepsilon-1/2}))$ using the coupling between $G(n, m)$ and $G(n, p)$ given in the previous subsection. First note

$$\text{Var}[W|G(n, m)] = \text{Var} \left[\sum_{3 \leq k \leq \ell} W_k \middle| G(n, m) \right]$$

as fixing the number of edges fixes W_2 . (Here we have abusively used the notation of conditional variance to express that the underlying variables in W are being sampled from $G(n, m)$ in this situation.) Now, in $G(n, p)$, by a trivial calculation we have

$$\text{Var} \left[\sum_{3 \leq k \leq \ell} W_k \middle| G(n, p) \right] = (1 + \Theta_\lambda(n^{-1})) \text{Var}[W_3|G(n, p)].$$

Therefore, to prove our claim it suffices to show that

$$\text{Var} \left[\sum_{3 \leq k \leq \ell} W_k \middle| G(n, p) \right] = (1 + O_{\lambda,\varepsilon}(n^{\varepsilon-1/2})) \text{Var} \left[\sum_{3 \leq k \leq \ell} W_k \middle| G(n, m) \right].$$

To prove this consider the coupling between $G(n, m)$ and $G(n, p)$ in the previous subsection; that is, sample $G(n, p)$ and then adjust the number of edges to be exactly m . Now define

$$Y_k = W'_k - W_k,$$

where W'_k has distribution corresponding to that in $G(n, p)$ and W_k the appropriate distribution in $G(n, m)$ for $3 \leq k \leq \ell$. The key claim in the previous subsection was essentially that

$$\mathbb{P}[|Y_3| \geq n^{\varepsilon-1/2}\sigma] \lesssim \exp(-\Omega_{\lambda,\varepsilon}((\log n)^2))$$

and this coupling almost immediately gives the desired result. In particular,

$$\begin{aligned} & \text{Var} \left[\sum_{3 \leq k \leq \ell} W'_k \right] - \text{Var} \left[\sum_{3 \leq k \leq \ell} W_k \right] \\ &= \mathbb{E} \left[\left(\sum_{3 \leq k \leq \ell} W'_k \right)^2 - \left(\sum_{3 \leq k \leq \ell} W_k \right)^2 \right] + \mathbb{E} \left[\sum_{3 \leq k \leq \ell} W_k \right]^2 - \mathbb{E} \left[\sum_{3 \leq k \leq \ell} W'_k \right]^2 \\ &= \mathbb{E} \left[\left(\sum_{3 \leq k \leq \ell} Y'_k \right) \left(\sum_{3 \leq k \leq \ell} W_k + W'_k \right) \right] - \mathbb{E} \left[\sum_{3 \leq k \leq \ell} Y'_k \right] \mathbb{E} \left[\sum_{3 \leq k \leq \ell} W_k + W'_k \right]. \end{aligned}$$

By using that W_i are mean zero in $G(n, p)$ and hypercontractivity (Theorem 2.1) with subsampling it follows that

$$\mathbb{P} \left[\left| \sum_{3 \leq k \leq \ell} W'_k \right| \geq n^\varepsilon \sigma \right] \lesssim \exp(-\Omega_{\lambda,\varepsilon}((\log n)^2)), \quad \mathbb{P} \left[\left| \sum_{3 \leq k \leq \ell} W_k \right| \geq n^\varepsilon \sigma \right] \lesssim \exp(-\Omega_{\lambda,\varepsilon}((\log n)^2)).$$

Furthermore, since $|\sum_{4 \leq k \leq \ell} Y_k| \leq |\sum_{4 \leq k \leq \ell} W_k| + |\sum_{4 \leq k \leq \ell} W'_k|$ it follows, again using hypercontractivity (Theorem 2.1) with subsampling, that

$$\mathbb{P} \left[\left| \sum_{4 \leq k \leq \ell} Y_k \right| \geq n^{\varepsilon-1/2} \sigma_W \right] \lesssim \exp(-\Omega_{\lambda, \varepsilon}((\log n)^2))$$

as the standard deviation of $\sum_{4 \leq k \leq \ell} W_k$ is $\Theta_{\lambda}(n^{-1/2})$ smaller than $\sum_{3 \leq k \leq \ell} W_k$. Finally, using that W_k, W'_k are polynomially bounded random variables, substituting in the above analysis gives that

$$\left| \mathbb{E} \left[\left(\sum_{3 \leq k \leq \ell} Y_k \right) \left(\sum_{3 \leq k \leq \ell} W_k + W'_k \right) \right] - \mathbb{E} \left[\sum_{3 \leq k \leq \ell} Y_k \right] \mathbb{E} \left[\sum_{3 \leq k \leq \ell} W_k + W'_k \right] \right| \lesssim_{\lambda, \varepsilon} n^{2\varepsilon-1/2} \sigma^2,$$

and rearranging and taking square roots the desired claim that $\sigma = \sigma_W(1 + O_{\lambda, \varepsilon}(n^{\varepsilon-1/2}))$ follows.

We now derive that $|\mathbb{E}[\sum_{3 \leq k \leq \ell} W_k | G(n, m)]| \lesssim_{\lambda} n^{-1/2} \sigma_W$ which is equivalent to the second estimate we wish to derive. To see this, we simply use linearity of expectation. All we need is that any multilinear degree k or less monomial in the χ_e has expectation bounded in absolute value by $O_{\lambda}(n^{-2})$. To see this, we compute

$$\mathbb{E}[\chi_{e_j} | \chi_{e_1}, \dots, \chi_{e_{j-1}}] \in \left[(m - j + 1) / \binom{n}{2} - j + 1 - p, m / \binom{n}{2} - j + 1 - p \right] / \sqrt{p(1-p)},$$

which is $O_{\lambda}(n^{-2})$. Using this estimate directly, it follows that

$$\mathbb{E} \left[\sum_{3 \leq k \leq \ell} W_k \right] \lesssim_{\lambda, p} \sum_{3 \leq k \leq \ell} n^{\ell-k} (n^k) O_{\ell, p}(1/n^2) \lesssim_{\lambda} n^{\ell-2} \lesssim_{\lambda} n^{-1/2} \sigma_W. \quad \square$$

3.4 | Bounds for $n^{\varepsilon} \leq |t| \leq \sigma n^{-\varepsilon}$

This subsection is by far the most elaborate in the paper due to various technical computations. At first reading the reader is recommended to take various probability and concentration claims at face value and not delve deeply into the calculations. It may also be useful to think of the $G(n, p)$ case as a model for calculations. Throughout, we will fix a factor system with data $k, \mathcal{H}, \mathcal{H}'$ as well as an (ℓ, λ) -graph statistic W with standard deviation σ and normalized version $\mathcal{K} = (W - W_0 - W_2)/\sigma$. Our ultimate goal is to show the bound

$$|\mathbb{E} e^{it\mathcal{K}} - e^{-t^2/2}| \lesssim_{\lambda, \varepsilon} n^{-\Omega_{\lambda, \varepsilon}(\log \log n)}$$

for all $|t| \in [n^{\varepsilon}, \sigma n^{-\varepsilon}]$ (see Lemma 3.6).

For this section, consider the following decoupling. Choose some $1 \leq k \leq \ell - 2$. Partition the vertex set into U_1, \dots, U_k of size n^{β} and a remainder set U_0 , for some $\beta \in (0, 1)$. We now separate the edge set into $k + 1$ classes B_0, \dots, B_k , where an edge between a vertex of U_i and U_j is put in $B_{\max(i, j)}$. We will require that β is bounded away from 0 and 1 by a constant depending only on \mathcal{H} . Therefore, B_0 has $\Theta(n^2)$ edges and each B_i for $i \geq 1$ has $\Theta(n^{1+\beta})$ edges. This decoupling is closely related to that in [4, Sections 9 and 10].

Now sample $(Z_i)_{0 \leq i \leq k}$, the number of edges chosen in each B_i , as if it is coming from $G(n, m)$, and then sample X , the actual vector of edges of B_0 (conditional on the Z_i). Then sample two independent copies Y_i^0, Y_i^1 of the vector of edges in B_i for $1 \leq i \leq k$, conditional on the previous information. Equivalently, we sampled from $G(n, m)$ and then resampled the edges in $(B_i)_{1 \leq i \leq k}$ but preserved the number of edges in each B_i .

Define a *suitable* outcome of the values Z_i to be if $|Z_i - p|B_i|| \leq \sqrt{|B_i|} \log |B_i|$, say. The key point is that there is an overwhelming probability that all Z_i are suitable by Azuma–Hoeffding and union bounding over a fixed number of events $k + 1 \leq \ell$. Indeed, the probability of failure is $\exp(-\Omega_\lambda((\log n)^2))$. (Here one must use a version of tail bounds for hypergeometric distributions, which can be deduced via Azuma–Hoeffding and subsampling, or see, for example, [19, Theorem 2.10]; we will not remark on this further.)

If we sample the edges of B_i with probability p independently (sampling $i \geq 1$ twice) then we attain any particular suitable vector of edge counts over the B_i with probability at least $\exp(-O_\lambda((\log n)^2))$. Therefore, if in this independent model an event has probability at most $\exp(-\Omega_\lambda((\log n)^3))$, then even in the $G(n, m)$ model within suitable outcomes it occurs with this probability, perhaps weakening the constants in the exponent. (This is a version of the transference trick we used in the small $|t|$ regime as well.) Then we must add back in the unsuitable cases, which account for a probability of at most $\exp(-\Omega_\lambda((\log n)^2))$ by the above application of Azuma–Hoeffding.

Now define coefficients δ , which are functions of \mathbf{Y} , via

$$\alpha(W)(X, \mathbf{Y}) = \delta_\emptyset + \sum_{e \in B_0} \delta_e \chi_e + \sum_{\substack{S \subseteq B_0 \\ |S| \geq 2}} \delta_S \chi_S.$$

Recall the definition of α as an iterated difference from Definition 2.8

We now proceed to prove an absurd number of bounds on these coefficients with extremely high probability in the $G(n, p)$ model, such that the above argument applies to transfer the high probability to the $G(n, m)$ model. Note that the δ are polynomials in χ_e^b for $e \notin B_0$ and $b \in \{0, 1\}$, where the superscript refers to the distinction between the two samples Y_i^0, Y_i^1 . In fact, we have

$$\delta_S = \sum_{H' \simeq H \in \mathcal{H}} \Delta_H n^{\ell - v(H)} \prod_{j=1}^k \left(\prod_{\substack{e \in E(H') \setminus S \\ e \in B_j}} \chi_e^1 - \prod_{\substack{e \in E(H') \setminus S \\ e \in B_j}} \chi_e^0 \right), \quad (3.7)$$

where the sum is over subgraphs H' of $\binom{[n]}{2}$ isomorphic to a graph in \mathcal{H} such that H' contains all $e \in S$, no other edges of B_0 , and at least 1 edge in each B_i for $i \geq 1$. Therefore, there is at least 1 vertex in each U_i with $i \geq 1$, and all the vertices of S are included. We will find it convenient to extract a ‘main term’ from $\delta_e, e \in B_0$, namely

$$\delta'_e = \sum_{H' \simeq H \in \mathcal{H}_{k+2}} \Delta_H n^{\ell - k - 2} \prod_{j=1}^k \left(\prod_{\substack{e \in E(H') \setminus S \\ e \in B_j}} \chi_e^1 - \prod_{\substack{e \in E(H') \setminus S \\ e \in B_j}} \chi_e^0 \right). \quad (3.8)$$

To be clear, the sum is over subgraphs H' of $\binom{[n]}{2}$ isomorphic to a graph in \mathcal{H}_{k+2} such that H' contains e , no other edges of B_0 , and at least 1 edge in each B_i for $i \geq 1$. Thus, it has $k+2$ vertices, which by the above considerations is the smallest number of vertices H' could have. We can easily show by induction on $i \geq 1$ that every vertex is connected within H' to the edge e , hence the H' considered must be connected. And by hypothesis of being a factor system, $H' \cap \mathcal{H}_{k+2}$ has a connected graph, so the above considerations suggest δ'_e should be nontrivial (though we defer a formal statement of what that means and a proof to Lemma 3.5(3)). Let $r_e = \delta_e - \delta'_e$ be the remainder. We will now prove the following set of bounds on the sizes of these coefficients.

Lemma 3.5. *Let $X, \mathbf{Y}, \delta_S, \delta_e, \delta'_e$, and r_e be as above. Let C be a suitably large constant. Then have the following concentration bounds (in the $G(n, p)$ model).*

(1) *We have that*

$$\mathbb{P}[\sup_{e \in B_0} |\delta_e| \geq n^{\ell-k-2+k\beta/2} (\log n)^C] \leq \exp(-\Omega_\lambda((\log n)^3)).$$

(2) *We have that*

$$\mathbb{P}[\sup_{e \in B_0} |r_e| \geq n^{\ell-k-5/2+k\beta/2} (\log n)^C] \leq \exp(-\Omega_\lambda((\log n)^3)).$$

(3) *We have that*

$$\mathbb{E} \left[\sum_{e \in B_0} \delta_e^2 \right] = \Theta_\lambda(n^{2(\ell-k-1)+k\beta})$$

and

$$\text{Var} \left[\sum_{e \in B_0} \delta_e^2 \right] = O_\lambda(n^{4(\ell-k-1)+(2k-1)\beta}).$$

(4) *We have that*

$$\mathbb{P} \left[\left| \sum_{e \in B_0} \delta_e \right| \leq n^{(\ell-k-1/2)+k\beta/2} (\log n)^C \right] \leq \exp(-\Omega_\lambda((\log n)^3)).$$

(5) *We have that*

$$\mathbb{P} \left[\left| \sum_{\substack{S \subseteq B_0 \\ |S| \geq 2}} \delta_S^2 \right| \leq n^{2(\ell-k)-3+k\beta} (\log n)^{2C} \right] \leq \exp(-\Omega_\lambda((\log n)^3)).$$

We now prove each of the parts of Lemma 3.5 in order with each item corresponding to a separate subsection.

3.4.1 | Proof of Lemma 3.5(1)

We wish to show that for all $e \in B_0$ we have that

$$|\delta_e| \leq n^{\ell-k-2+(k\beta/2)}(\log n)^C$$

with probability $1 - \exp(-\Omega_\lambda((\log n)^3))$ for suitable C . To see this note δ_e is a polynomial of bounded degree in the χ_e^b for $e \notin B_0$ and $b \in \{0, 1\}$, with sum of squares of coefficients $O_\lambda(n^{2(\ell-k-2)+k\beta})$. Thus, applying hypercontractivity (Theorem 2.1) gives the result. Here C must be chosen large enough in terms of the degree of the polynomial, which is bounded by ℓ , so can be taken to depend only on \mathcal{H} .

The sum of squares estimate is derived as follows. The contributing terms to δ_e are subgraphs H' as delineated after (3.7), with $S = \{e\}$. Say it has v vertices, and w vertices outside of U_0 . Note that $v \geq w + 2$ as we have at least 2 vertices in U_0 . Then the coefficient is $O_\lambda(n^{\ell-v})$ and there are $n^{\beta w + (v-w-2)}$ choices for the location of the remaining vertices, since two are fixed by e . Hence, the contribution is $O_\lambda(n^{2(\ell-v)+\beta w + (v-w-2)})$. As w increases this decreases, so the major contribution is from $w = k$, the minimum, and as v increases the resulting expression decreases, so the major contribution is from $v = k + 2$, the minimum, yielding the desired bound. It is worth noting for later that the main contributors are those with $v = k + 2$ and $w = k$ only, and the next highest term is from $v = k + 3$ and $w = k$, which is n^{-1} times smaller.

3.4.2 | Proof of Lemma 3.5(2)

We next show that for all $e \in B_0$, $|r_e| \leq n^{\ell-k-5/2+(k\beta/2)}(\log n)^C$ with high probability (of the same quality as before). Indeed, the only point is that δ'_e contains all the main contributors discussed above (as is evident from (3.8)), and thus the sum of squares of coefficients in $r_e = \delta_e - \delta'_e$ is $O_\lambda(n^{2(\ell-k-2)+k\beta-1})$. Hypercontractivity (Theorem 2.1) finishes.

3.4.3 | Proof of Lemma 3.5(3)

We now prove that $\sum_{e \in B_0} \delta_e^2$ concentrates on a value of size $\Theta_\lambda(n^{2(\ell-k-1)+k\beta})$. First we compute the expectation. Note that $\mathbb{E}[\chi_{e'}^1 - \chi_{e'}^0] = 0$, hence the same is true of the products making up δ_e (see (3.7)) by independence. Further, we see that if we expand δ_e^2 into a sum of products of two terms based on (3.7), then the product of two terms coming from H'_1, H'_2 has zero expectation unless they have the same edge set, in which case it is constant. Summing over $H'_1 = H'_2$ with v vertices and w outside U_0 , we obtain an expectation of size $O_\lambda(n^{2(\ell-v)} \cdot n^{\beta w + (v-w-2)})$ for such terms. This is maximized when $w = k$ and $v = k + 2$. Furthermore, if we take a connected graph $H \in \mathcal{H}' \cap \mathcal{H}_{k+2}$ and then look at its embeddings H' containing e with a vertex in each U_i for $i \geq 1$, we find that the terms $H'_1 = H'_2 = H'$ contribute $\Omega_\lambda(n^{2(\ell-k-2)+\beta k})$ in the above. So we obtain $\mathbb{E}\delta_e^2 = \Theta_\lambda(n^{2(\ell-k-2)+\beta k})$. Summing over $\Theta(n^2)$ edges in B_0 , we obtain the first claimed result.

Next, writing $\delta_e^2 = (\delta'_e)^2 + 2\delta_e r_e - r_e^2$, and using the L^∞ bounds from Lemma 3.5(1) and (2) above, we see with high probability that $\sum_{e \in B_0} \delta_e^2$ and $\sum_{e \in B_0} (\delta'_e)^2$ differ by $O_\lambda(n^{2 \cdot n^{2(\ell-k)-9/2+k\beta}}(\log n)^{2C})$. This is smaller in magnitude than the expectation, that is, with high probability this deviation is (relatively) small.

But note that adversarially changing a random variable by t in any direction can increase its variance from V to at most $2V + 2t^2$ (since $\text{Var}(A + B) \leq 2\text{Var}(A) + 2\text{Var}(B)$). Therefore, since $O_\lambda((n^2 \cdot n^{2(\ell-k)-9/2+k\beta}(\log n)^{2C})^2) = o_\lambda(n^{4(\ell-k-1)+(2k-1)\beta})$ (as $\beta < 1$), it remains to study the variance of $\sum_{e \in B_0} (\delta'_e)^2$ instead.

Now it remains to show the standard deviation of $\sum_{e \in B_0} (\delta'_e)^2$ is smaller in magnitude by some power of n compared to the expectation (explicitly, by a factor of $n^{\beta/2}$ or so). Note that the variance is

$$\sum_{e_1, e_2 \in B_0} (\mathbb{E}[\delta'_{e_1} \delta'_{e_2}] - \mathbb{E}[\delta'_{e_1}] \mathbb{E}[\delta'_{e_2}]).$$

First, if e_1, e_2 share a vertex, there are $O(n^3)$ choices for them. Using the L^∞ bounds on δ_e and r_e Lemma 3.5(1) and (2) (which yields a bound on δ'_e) we see that the contribution to the sum above is $O_\lambda(n^3 \cdot n^{4(\ell-k-2)+2k\beta}(\log n)^{4C})$. This bound is acceptable, by a factor of approximately $n^{-1/2}$ against the standard deviation.

Now consider the $O(n^4)$ cases where e_1, e_2 do not share a vertex. We write out $(\delta'_{e_j})^2$ as a sum over $H_{j,1}, H_{j,2}$:

$$(\delta'_{e_j})^2 = \sum_{\substack{H'_{j,1} \simeq H_{j,1} \in H \\ H'_{j,2} \simeq H_{j,2} \in H}} \Delta_{H_{j,1}} \Delta_{H_{j,2}} \prod_{t=1}^k \left(\prod_{\substack{e \in E(H_{j,1}) \setminus S \\ e \in B_t}} \chi_e^1 - \prod_{\substack{e \in E(H_{j,1}) \setminus S \\ e \in B_t}} \chi_e^0 \right) \left(\prod_{\substack{e \in E(H_{j,2}) \setminus S \\ e \in B_t}} \chi_e^1 - \prod_{\substack{e \in E(H_{j,2}) \setminus S \\ e \in B_t}} \chi_e^0 \right).$$

Therefore, we can write the above covariance $\mathbb{E}[\delta'_{e_1} \delta'_{e_2}] - \mathbb{E}[\delta'_{e_1}] \mathbb{E}[\delta'_{e_2}]$ as a further sum of covariances, with terms indexed by a choice of $H'_{j,b}$ for $j, b \in \{1, 2\}$.

Since we are dealing with δ' (as opposed to δ) these graphs are connected, with 2 vertices in B_0 forming the prescribed edge and 1 vertex in each U_i with $i \geq 1$. Consider the union of all these graphs $H'_{j,b}$, $j, b \in \{1, 2\}$ (within K_n). If any of its edges is only covered once, then we easily see the corresponding covariance will be zero (recall we are currently in the $G(n, p)$ model).

Suppose the union graph has at least 3 vertices in some U_i with $i \geq 1$. Then one of the vertices is hit by a unique $H'_{j,b}$, which implies some edge is only hit by one. Thus, these terms are zero. Therefore, the remaining union graphs have at most 2 vertices in each U_i for $i \geq 1$. Now suppose that for some U_i there is only 1 vertex. Then the number of configurations that could give rise to this situation is $O_\lambda(n^4 \cdot n^{(2k-1)\beta})$, with coefficient of size $O_\lambda(n^{4(\ell-k-2)})$. This gives an acceptable bound as well, by a factor of $n^{-\beta/2}$ against the standard deviation.

Now consider the case where there are exactly 2 vertices in each U_i with $i \geq 1$. We claim that the remaining terms are zero. It can be nonzero only if every edge of $H'_{1,1}, H'_{1,2}, H'_{2,1}, H'_{2,2}$ is covered more than once in the union of these graphs. However, our graphs $H'_{j,b}$ are connected with 1 vertex in each U_i for $i \geq 1$. We easily prove by induction on U_i for $i \geq 1$ that to satisfy the edge covering condition, the graphs $H'_{j,1}, H'_{j,2}$ have the same vertex set. (Carrying this out requires

e_1, e_2 to have disjoint vertices.) Now, this implies the vertex sets of $H'_{1,1}, H'_{1,2}$ versus $H'_{2,1}, H'_{2,2}$ are disjoint in any remaining term. Therefore, the edge sets are disjoint so the corresponding variables are independent, leading to a zero term once more.

Overall, we obtain a bound on the variance of quality $O_\lambda(n^{4(\ell-k-1)+(2k-1)\beta})$, so the standard deviation is $O_\lambda(n^{2(\ell-k-1)+k\beta-(\beta/2)})$, which is the desired bound.

3.4.4 | Proof of Lemma 3.5(4)

Next we show

$$\left| \sum_{e \in B_0} \delta_e \right| \lesssim_\lambda n^{\ell-k-1/2+(k\beta/2)} (\log n)^C$$

with probability $1 - \exp(-\Omega_\lambda((\log n)^3))$. To bound the sum of squares of coefficients of $\sum_{e \in B_0} \delta_e$, imagine writing each $\delta_e = \sum_{I \in \mathcal{I}} a_{e,I} M_I$, where M_I is a bounded degree monomial in the χ_e^b . Each has a sum of squares $\sum_{I \in \mathcal{I}} a_{e,I}^2$ which corresponds to its individual variance.

On the other hand, $\sum_{e \in B_0} \delta_e = \sum_{I \in \mathcal{I}} (\sum_{e \in B_0} a_{e,I}) M_I$ has sum of squares $\sum_{I \in \mathcal{I}} (\sum_{e \in B_0} a_{e,I})^2$. In particular, every monomial M_I (which is a product over $E(H') \setminus e$) is in at most n^2 polynomials $\delta_{e'}$ trivially (namely because $|B_0| \leq n^2$). Therefore, after combining terms in $\sum_{e \in B_0} \delta_e$, by Cauchy-Schwarz, the new sum of squares of coefficients is at most n^2 times what we get by not combining:

$$\sum_{I \in \mathcal{I}} \left(\sum_{e \in B_0} a_{e,I} \right)^2 \leq \sum_{I \in \mathcal{I}} n^2 \left(\sum_{e \in B_0} a_{e,I}^2 \right).$$

This new expression in turn has $|B_0|$ values (based on e) which were bounded in Subsection 3.4.1. This gives $O_\lambda(n^4 \cdot n^{2(\ell-k-2)+k\beta})$, which is not good enough.

But in fact, for monomials that contribute the most, namely those corresponding to the H' with $v = k + 2$ vertices and $w = k$ of them outside B_0 , we see that $E(H') \setminus e$ can be completed to a valid monomial in some $\delta_{e'}$ only if e' is incident to 1 of the 2 vertices in $V(H') \cap B_0$, which yields $2n$ possible polynomials a given term is in. The remaining monomials have a better bound (by a factor of n) as proved in Subsection 3.4.1. Therefore, we obtain $O_\lambda(n^3 \cdot n^{2(\ell-k-2)+k\beta} + n^4 \cdot n^{2(\ell-k-2)+k\beta-1})$, which yields the result directly upon using hypercontractivity (Theorem 2.1).

3.4.5 | Proof of Lemma 3.5(5)

Finally, we prove

$$\sum_{\substack{S \subseteq B_0 \\ |S| \geq 2}} \delta_S^2 \lesssim_\lambda n^{2(\ell-k)-3+k\beta} (\log n)^{2C}$$

with probability $1 - \exp(-\Omega_\lambda((\log n)^3))$. We first consider each term δ_S individually. As above, it is a polynomial of bounded degree. The contributing monomials correspond to subgraphs H' of $\binom{[n]}{2}$ isomorphic to some graph in \mathcal{H} such that H' contains the edges of S , no other edges of B_0 ,

and at least 1 edge in each B_i for $i \geq 1$. Again, suppose it has v vertices, with a of them spanned by the edges in S , and w outside of U_0 . There is at least 1 vertex in each U_i for $i \geq 1$. Then we obtain an estimate of $O_\lambda(n^{\ell-v})$ for the coefficient corresponding to H' , over $n^{\beta w + (v-w-a)}$ different possible such H' . Summing over $w \geq k$ we find that the sum of squares of coefficients therefore is $O_\lambda(n^{2(\ell-k-a)+\beta k})$, similar to earlier. Again we have $|\delta_S| \leq n^{\ell-k-a+(k\beta)/2}(\log n)^C$ with probability $1 - \exp(-\Omega_\lambda((\log n)^3))$. There are $O(n^a)$ possible sets S spanning a vertices, and summing the squares of the above gives $O_\lambda(n^{2(\ell-k)-a+k\beta}(\log n)^{2C})$. Then summing over $3 \leq a \leq f$ (since $|S| \geq 2$ guarantees $a \geq 3$) gives a total sum of squares of size $O_\lambda(n^{2(\ell-k)-3+k\beta}(\log n)^{2C})$ whp (with high probability), as claimed.

Remark. Note that in the case $k = \ell - 2$, there in fact are no higher terms as such a term would require H' to have at least $\ell + 1$ vertices, but all $H \in \mathcal{H}$ satisfy $v(H) \leq \ell$. This will be used later.

This (finally) concludes the proof of Lemma 3.5. Now we use these bounds to conclude our argument in the intermediate range of $|t|$.

3.4.6 | Deriving characteristic function bounds

Overall, we showed the above statements with high probability in the $G(n, p)$ model. As noted, this transfers to a statement with high probability in the $G(n, m)$ model via naive conditioning. Now we claim the following bound on the characteristic function, in the $G(n, m)$ model.

Lemma 3.6. *Let W be an (ℓ, λ) -statistic with normalized version \mathcal{K} . Then for all $\varepsilon > 0$ and $|t| \in [n^\varepsilon, \sigma n^{-\varepsilon}]$ we have*

$$|\mathbb{E}[e^{it\mathcal{K}} - e^{-t^2/2}]| \lesssim_{\lambda, \varepsilon} n^{-\Omega_{\lambda, \varepsilon}(\log \log n)}.$$

Proof. Note that $e^{-t^2/2}$ is sufficiently small in the necessary range to ignore. Let X, \mathbf{Y} be as at the beginning of Subsection 3.4, having already conditioned on ‘suitable’ Z_i . We use Lemma 2.10, obtaining

$$|\varphi_{\mathcal{K}}(t)|^{2^k} \leq \mathbb{E}_{\mathbf{Y}} |\mathbb{E}_X e^{it\alpha(W)(X, \mathbf{Y})/\sigma}|.$$

Now with probability $1 - \exp(-\Omega_\lambda((\log n)^2))$ over the randomness of \mathbf{Y} , we can assume all the claims regarding the δ coefficients in Lemma 3.5 are true. This leaves an error term of size $\exp(-\Omega_\lambda((\log n)^2))$ which we will be able to disregard. We can also impose the condition that $|\sum_{e \in B_0} \chi_e| \lesssim_\lambda B_0^{1/2} \log B_0$ by Azuma–Hoeffding upon revealing the elements of B_0 . (And since revealing \mathbf{Y} will deterministically fix the sum in question as we are in $G(n, m)$, the same holds when conditioning only on \mathbf{Y} .) This induces an error term of size $\exp(-\Omega_\lambda((\log n)^2))$, again acceptable.

Now condition on one of the suitable choices of \mathbf{Y} . Define

$$L = \sum_{e' \in B_0} \delta_{e'} \chi_{e'}, \quad U = \sum_{\substack{S \subseteq B_0 \\ |S| \geq 2}} \delta_S \chi_S,$$

which are random variables now depending only on X (as \mathbf{Y} is fixed). We need to bound

$$\mathbb{E}_X e^{it(L+U)/\sigma},$$

noting we can disregard δ_\emptyset as $|e^{it\delta_\emptyset/\sigma}| = 1$. To bound this quantity, we will adapt the method in [4, Theorem 3]. Fix some integer $d \geq 1$ that we will later send to infinity slowly. Now by Taylor's theorem with Lagrange error,

$$\left| e^{itU/\sigma} - \sum_{j=0}^d \frac{(itU/\sigma)^j}{j!} \right| \leq 2 \frac{|tU/\sigma|^{d+1}}{(d+1)!},$$

where the 2 comes from splitting into real and imaginary parts. Note that the interior sum is really a polynomial in the functions χ of degree bounded in terms of d , with coefficients at most some polynomial in n of degree bounded by d , noting that $|t| \leq \pi\sigma$. Therefore, we can write

$$\sum_{j=0}^d \frac{(itU/\sigma)^j}{j!} = \sum_{M \in \mathcal{M}} a_M \cdot M,$$

where \mathcal{M} is a set of bounded degree monomials in the χ variables, and in particular $\sum_{M \in \mathcal{M}} |a_M| = O(n^D)$ for some D depending on d . We see

$$|\mathbb{E}_X e^{it(L+U)/\sigma}| \lesssim_d \sum_{M \in \mathcal{M}} |a_M \mathbb{E}_X M e^{itL/\sigma}| + \mathbb{E}_X |tU/\sigma|^{d+1}.$$

Now note that tU/σ is a polynomial in the $\chi_{e'}$ for $e' \in B_0$ of bounded degree, and by our assumptions on \mathbf{Y} we control its sum of squares of coefficients. By hypercontractivity (Theorem 2.1) we have

$$\mathbb{P}(|tU/\sigma| \geq n^{-\varepsilon}) = \exp(-\Omega_\lambda(n^{\varepsilon'}))$$

as long as $(t^2 n^{2\varepsilon} / \sigma^2) n^{2(\ell-k)-3+k\beta} < n^{-\varepsilon}$ for some $\varepsilon' > 0$, using our L^2 control of the higher terms Lemma 3.5(5). Here hypercontractivity applies in the independent model, but again using our subsampling trick we can make it over the randomness of X , which constrains $\sum_{e \in B_0} \chi_e$.

Now the last term has good bounds, since $|tU/\sigma| \geq n^{-\varepsilon}$ occurs with very low probability and $|tU/\sigma|$ is bounded above by some fixed degree polynomial in n always. Indeed, this allows us to bound the last term by $O_d(n^{-\varepsilon(d+1)})$. Alternatively, we could have used the moment form (Theorem 2.2) of hypercontractivity.

Now each term

$$|\mathbb{E}_X M e^{itL/\sigma}| \leq \mathbb{E}_{e' \text{ of } \text{supp}(M)} |M| |\mathbb{E} e^{itL/\sigma}|,$$

where the outer expectation is over the randomness of each e' contained in the monomial M , and the inner expectation is only over the randomness of $e' \in B_0$ not contained in the monomial M . This is all but $O_d(1)$ of them. Now, the inner term is of a form with which we can apply Lemma 2.7 (say, shifting the $\chi_{e'}$ back to $x_{e'}$). The precise value of the conditioned sum $\sum_{e' \in B_0} \chi_{e'}$ that we chose at the beginning will change exactly what replaces p in the statement of the lemma, but

it is say in $(\lambda/2, 1 - \lambda/2)$ for n sufficiently large, hence bounded away from $\{0, 1\}$. Therefore, we obtain a bound of quality

$$O_\lambda(n^D) \cdot n^2 \exp(-\Omega_\lambda((t^2 n^2 / \sigma^2) \text{Var}[\delta_{e'}])),$$

where D is some constant depending on d . (Here Var is the variance of the sequence of numbers as in Lemma 2.7; $\delta_{e'}$ are constants given the conditioned information.) Now the point is we control $\text{Var}[\delta_{e'}]$ because of all the bounds from earlier. Indeed, the average of $\delta_{e'}^2$ concentrates on a value of size $\Theta_\lambda(n^{2(\ell-k-2)+k\beta})$ by Lemma 3.5(3) with super-polynomial probability using hypercontractivity (Theorem 2.1). On the other hand, the average of $\delta_{e'}$ is of size $O_\lambda(n^{\ell-k-5/2+(k\beta/2)}(\log n)^C)$ by Lemma 3.5(4). Since that is smaller in magnitude than the square root of above, we see that the variance $\text{Var}[\delta_{e'}]$ over all $e' \in B_0$ is of order $\Theta_\lambda(n^{2(\ell-k-2)+k\beta})$. The deletion of $O_d(1)$ terms from the $\delta_{e'}$ does not change the variance from this order of magnitude due to the L^∞ bounds on $\delta_{e'}$ established by Lemma 3.5(1) and (2). Therefore, if $(t^2 n^2 / \sigma^2) n^{2(\ell-k-2)+k\beta} > n^\varepsilon$ then this bound is acceptable. Additionally, to apply Lemma 2.7 we need $|t/\sigma| \cdot |\delta_{e'}| \lesssim 1$, hence $|t/\sigma| n^{\ell-k-2+(k\beta/2)}(\log n)^C \lesssim 1$ suffices.

In conclusion, fixing $\varepsilon > 0$, we have shown for any fixed d that

$$|\mathbb{E}_X e^{it(L+U)/\sigma}| \lesssim_{d,\lambda} n^{-\varepsilon(d+1)}$$

as long as

$$(t^2 n^{2\varepsilon} / \sigma^2) n^{2(\ell-k)-3+k\beta} < n^{-\varepsilon}, \quad (t^2 n^2 / \sigma^2) n^{2(\ell-k-2)+k\beta} > n^\varepsilon, \quad |t/\sigma| n^{\ell-k-2+(k\beta/2)}(\log n)^C \lesssim 1. \quad (3.9)$$

Now we send $d \rightarrow \infty$ slowly, finding ultimately that

$$|\varphi_K(t)| \lesssim n^{-\Omega_{\lambda,\varepsilon}(d(n))}$$

for some slow growing $d = d(n)$ that is monotonic and limits to infinity. Note that $d(n) = \log \log n$ surely suffices.

Now it remains to calculate which range of t is covered by this computation. The three bounds (3.9) show that the range

$$n^{k-(k\beta+1-\varepsilon)/2} \lesssim_\lambda |t| \lesssim_\lambda n^{k-(k\beta+3\varepsilon)/2}$$

certainly is valid. We are allowed to range $1 \leq k \leq \ell - 2$ and $0 < \beta < 1$, although remember the warning that β must be bounded away from $\{0, 1\}$. Restricting $\ell\beta \in (\varepsilon, 1 - \varepsilon)$ still allows us to cover the range $t \in [n^{(k-1)/2+2\varepsilon}, n^{k-3\varepsilon}]$ for each k , say. For $1 \leq k \leq \ell - 2$ these intervals overlap and hit the range $[n^{2\varepsilon}, n^{\ell-2-3\varepsilon}]$. This almost hits the entire range we want.

However, note that for $k = \ell - 2$, the top value, there are no higher order terms: see the remark following the proof of Lemma 3.5 (5). Hence, there is no U term and the above analysis is simplified. In particular, the first of the three conditions on t in (3.9) can be dropped. So for $k = \ell - 2$ we actually cover the larger range governed by

$$(t^2 n^2 / \sigma^2) n^{2(\ell-k-2)+k\beta} > n^\varepsilon, \quad |t/\sigma| n^{\ell-k-2+(k\beta/2)}(\log n)^C \lesssim 1,$$

which allows us to cover $n^{k-(k\beta+1-\varepsilon)/2} \lesssim_\lambda |t| \lesssim_\lambda n^{k-(k\beta-1+3\varepsilon)/2}$ when $k = \ell - 2$. This lets us cover $|t| \in [n^{(\ell-3)/2+2\varepsilon}, n^{\ell-3/2-3\varepsilon}]$, which gets the remaining portion of the range.

Therefore, we have hit every necessary t with a bound of the desired quality, taking ε sufficiently small. \square

Remark. Ensuring the ranges cover everything is where we use the hypothesis that $\mathcal{H}' \cap \mathcal{H}_k$ contains a connected graph for all $3 \leq k \leq \ell$. More specifically, this hypothesis is used in the proof of the first part of Lemma 3.5(3). Looking closely, we see that this can be weakened; the precise condition coming from our argument is that $\mathcal{H}' \cap \mathcal{H}_k$ contains a connected graph for all k in some set $\{k_1, \dots, k_a\}$, where $k_1 = 3$, $k_a = \ell$, and $k_{j+1} \leq 2k_j - 2$.

3.5 | Further comments

It is worth noting that the above proofs also work for the $G(n, p)$ model with the obvious alterations. In fact, there is significantly less headache because we have independence. The major difference is that \mathcal{H} should include K_2 , and now X_2 controls the standard deviation. Although this approach would allow us to conclude the necessary theorems about $G(n, p)$, we will instead demonstrate those results via transference from the $G(n, m)$ model, which is a more powerful technique as we will see from our study of k -APs as well as anticoncentration counterexamples.

We would also like to briefly address how the results up to this point are already sufficient to prove anticoncentration for statistics which satisfy the hypotheses of Subsection 3.1. In particular, using Esseen's concentration inequality [10] (see [29] for a modern treatment) one can convert the bounds we have derived into anticoncentration estimates, losing only a factor of $n^{o(1)}$ versus the optimal bound in $G(n, m)$ (which can bootstrapped to $G(n, p)$). In fact, using a variant of the decouplings provided in Section 4 one can establish Fourier control up to $|t| \leq c_{\mathcal{H}, \lambda} \sigma$ and thus establish optimal anticoncentration for $G(n, m)$, losing only constant factors. Again this can be bootstrapped to $G(n, p)$ (with some care being required).

Note here that these remarks extend to graph statistics such as two times the C_6 count plus the number of copies $P_1 + P_3$, the disjoint union of paths of length 1 and 3. This example may seem rather obscure but note that this statistic has a parity bias due to results of DeMarco and Redlich [8] and therefore a local central limit theorem fails. However, using our Fourier analytic methods, since this statistic ultimately takes the form (3.1), we can still obtain optimal anticoncentration for random variables that do not satisfy a local central limit theorem, overcoming a theoretical obstruction suggested in [11].

We remark that difficulties in establishing anticoncentration through Fourier analytic methods arise when statistics display longer scale fluctuations in the pointwise probabilities than simple parity biases. This will be due to 'degeneracy' in which certain terms in the expansion (3.1) of the graph statistic are missing or of a different magnitude than expected. However, even such situations are not insurmountable as we will later demonstrate in the case of k -term arithmetic progressions in the independent model. In future work, we intend on elaborating on these remarks and developing a systematic theory of anticoncentration for graph counts.

4 | LOCAL LIMIT THEOREMS FOR SUBGRAPH COUNTS

We now prove local limit theorems for subgraph counts and induced subgraph counts. Throughout this section, we will work in the $G(n, m)$ model. Specifically, we prove a local limit theorem for subgraph counts of H where H is connected, and induced subgraph counts for any H . We specify

that the number of edges must be such that $p = m/\binom{n}{2}$ is at least λ away from $\{0, 1\}$. In the induced case, there may also be up to around $v(H)^2$ ‘critical’ values p_{crit} that p is λ away from. We will see later that, with these caveats, the results in Section 3 can be applied directly. Therefore, it remains to bound the necessary characteristic functions in the top range $\sigma n^{-\varepsilon} \leq |t| \leq \pi\sigma$.

4.1 | Connected subgraph counts

As this is the simpler case, we do it first. Let H be a connected graph on $\ell \geq 2$ vertices. Write

$$W = \sum_{\substack{H' \subseteq \binom{[n]}{2} \\ H' \simeq H}} \prod_{e \in E(H')} x_e.$$

If we let $\chi_e = (x_e - p)/\sqrt{p(1-p)}$ as usual, then it will expand into a form such as (3.1). In particular,

$$W = \sum_{\substack{H' \subseteq \binom{[n]}{2} \\ H' \simeq H}} \prod_{e \in E(H')} (p + \sqrt{p(1-p)}\chi_e) = \sum_{S \subseteq H} p^{e(H)-e(S)} (\sqrt{p(1-p)})^{e(S)} c_{S,H} d_{S,H} \binom{n-v(S)}{\ell-v(S)} \gamma_S(\mathbf{x}), \quad (4.1)$$

where the sum is over subgraphs S (lacking isolated vertices) of H up to isomorphism. Here $c_{S,H}$ explicitly equals $(\ell - v(S))! \text{aut } S / \text{aut } H$ and $d_{S,H}$ equals the number of times S appears as a subgraph of H , for example, $d_{K_2,H} = e(H)$. (Here $\text{aut } S$ is the size of the automorphism group of the graph S .) For the empty graph, these values are taken to be $\ell! / \text{aut } H$ and 1, respectively. This follows from an easy double-counting argument.

In particular, for $p \in (\lambda, 1 - \lambda)$, we see the coefficient of $\gamma_S(\mathbf{x})$ is of size $\Theta_\lambda(n^{\ell-v(S)})$. Furthermore, H has a connected subgraph with k vertices for each $3 \leq k \leq \ell$ since H is connected (for example, take subtrees of a spanning tree). Thus, the results of Section 3 apply. In particular, define σ and \mathcal{K} from W in the same way as in Subsection 3.1. Then by Lemma 3.3, for $|t| \leq n^\varepsilon$ we have

$$|\varphi_{\mathcal{K}}(t) - e^{-t^2/2}| \lesssim_\lambda \frac{|t|}{n^{\frac{1}{2}-\varepsilon}} + \exp(-\Omega_\lambda((\log n)^2)) \quad (4.2)$$

and by Lemma 3.6, for $n^\varepsilon \leq |t| \leq \sigma n^{-\varepsilon}$ we have

$$|\varphi_{\mathcal{K}}(t) - e^{-t^2/2}| \lesssim n^{-\Omega_{\lambda,\varepsilon}(\log \log n)}. \quad (4.3)$$

Now we present a decoupling which handles the top range $\sigma n^{-\varepsilon} \leq |t| \leq \pi\sigma$.

Lemma 4.1. *Let W be as in (4.1), and define σ, \mathcal{K} as in Subsection 3.1. Then for $|t| \leq \pi\sigma$,*

$$|\varphi_{\mathcal{K}}(t)| \leq \exp(-\Omega_\lambda(n)) + \exp\left(-\Omega_\lambda\left(\frac{t^2 n}{\sigma^2}\right)\right). \quad (4.4)$$

Proof. Partition the vertex set $[n]$ into $\lfloor n/\ell \rfloor$ cliques of size ℓ , with at most $\ell - 1$ extra vertices that we will essentially ignore. Within each of the cliques take an isomorphic copy of H and label

its edges 0 to $e(H) - 1$ arbitrarily. Let \widetilde{B}_0 be the union of all unlabeled edges along with those labeled 0, and B_i be all the edges labeled i for $0 \leq i \leq e(H) - 1$. Thus, B_i for $1 \leq i \leq e(H) - 1$ and \widetilde{B}_0 partition the edges. Finally, define $\widetilde{X} \in \{0, 1\}^{\widetilde{B}_0}$ as the indicator vector of which edges are included in $G(n, m)$ and Z_i for $1 \leq i \leq e(H) - 1$ as the number of edges in each set B_i when sampling from $G(n, m)$. Then let Y_i^0, Y_i^1 for $1 \leq i \leq e(H) - 1$ be two independent samples of the edges within B_i given Z_i (that is, they are independent vectors $\{0, 1\}^{B_i}$ which are uniform of size Z_i , given Z_i). Also let Y_0 be the indicator of the edges of \widetilde{X} in B_0 only, and \widetilde{Y}_0 be the indicator of the edges in $\widetilde{B}_0 \setminus B_0$ only. Now note that, writing $\mathbf{Y} = (Y_i^0, Y_i^1)_{1 \leq i \leq e(H)-1}$ as in Lemma 2.10,

$$\alpha(W)(\widetilde{X}, \mathbf{Y}) = \sum_{e \in \widetilde{B}_0} \delta_e(\mathbf{Y}) x_e$$

with $\delta_e(\mathbf{Y}) \in \{0, \pm 1\}$ for all $e \in B_0$ which are labeled. Indeed, for all $e \in B_0$ we have

$$\delta_e(\mathbf{Y}) = \prod_{e' \in E(H_e) \setminus e} (x_{e'}^1 - x_{e'}^0),$$

where H_e is the unique isomorphic copy of H containing e that was embedded into one of the cliques.

Now we claim that with extremely high probability, the number of $e \in B_0$ such that $\delta_e(\mathbf{Y}) = 1$ is greater than $\lambda^{2(e(H)-1)} n / (2\ell)$ and the number such that $\delta_e(\mathbf{Y}) = 0$ satisfies the same. This is clear in the $G(n, p)$ model, as there are more than $n / (2\ell)$ edges $e \in B_0$, which have mutually independent coefficients which are easily seen to take on the desired values with positive probabilities. In particular, the probability of this event not occurring in the independent model is $\exp(-\Omega_\lambda(n))$.

In the $G(n, m)$ model, we repeatedly use Azuma–Hoeffding. First, it demonstrates that each Z_i for $1 \leq i \leq e(H) - 1$ is approximately pn/ℓ with high probability, say, within the interval $[pn/(2\ell), (1+p)n/(2\ell)]$ with probability $1 - \exp(-\Omega_\lambda(n))$. Conditional on a realization of the Z_i , the vectors Y_i^0, Y_i^1 are independent and uniform with a fixed sum. By Azuma–Hoeffding again, we can show that $x_{e'}^1 = 1$ and $x_{e'}^0 = 0$ for each $e' \in E(H_e) \setminus e$ in at least $\Omega_\lambda(n)$ of our cliques with probability $1 - \exp(-\Omega_\lambda(n))$. Similarly, we can show that $x_{e'}^1 = x_{e'}^0 = 1$ for some $e' \in E(H_3) \setminus e$ happens in at least $\Omega_\lambda(n)$ of our cliques with a similar probability.

We also control the number of edges among Y_0 . Note that its distribution is the same as looking at the number of edges in a specific subset of $G(n, m)$. By Azuma–Hoeffding, with a process revealing edges within B_0 one at a time, we see with probability $1 - \exp(-\Omega_\lambda(n))$ the fraction of edges chosen in this set is in $(p/2, (1+p)/2)$. Therefore, conditional on the randomness of Y_i^0 and \widetilde{Y}_0 , say, the number of edges in Y_0 is $\Theta_\lambda(n)$ (noting that this number is deterministic given the revealed information).

Now we are ready to apply Lemma 2.10. Let B'_0 be the set of $e \in B_0$ such that $\delta_e(\mathbf{Y}) \notin \{0, 1\}$ and let Y'_0 be the corresponding indicator vector (which is a restriction of Y_0). This is deterministic given $\mathbf{Y} = (Y_i^b)_{1 \leq i \leq e(H)-1}$. We obtain

$$\begin{aligned} |\varphi_{\mathcal{K}}(t)|^{2^{e(H)-1}} &\leq \mathbb{E}_{\mathbf{Y}} |\mathbb{E}_{\widetilde{X}} e^{it\alpha(W)(\widetilde{X}, \mathbf{Y})/\sigma}| \leq \mathbb{E}_{\mathbf{Y}, \widetilde{Y}_0} |\mathbb{E}_{Y_0} e^{it\alpha(W)(\widetilde{X}, \mathbf{Y})/\sigma}| \\ &\leq \mathbb{E}_{\mathbf{Y}, \widetilde{Y}_0, Y'_0} |\mathbb{E}_{Y_0 \setminus Y'_0} e^{(it/\sigma) \sum_{e \in B_0 \setminus B'_0} \gamma_e(\mathbf{Y}) x_e}| \leq \exp(-\Omega_\lambda(n)) + \exp\left(-\Omega_\lambda\left(\frac{t^2 n}{\sigma^2}\right)\right), \end{aligned}$$

the last inequality using that the function is bounded by 1 in the rare cases delineated above, and using Lemma 2.7 in the remaining cases in which we know that the inner x_e for $e \in Y_0 \setminus Y'_0$ are drawn uniformly with a fixed sum depending on $\mathbf{Y}, \widetilde{Y}_0, Y'_0$. That sum is $\Theta_\lambda(n)$ in size, and additionally we use that a positive fraction (in terms of λ, H) of coefficients $\delta_e(\mathbf{Y})$ are 1 as well as 0. Note that Lemma 2.7 only applies if $(t/\sigma) \cdot (1 - 0) \leq \pi$, which precisely hits the top of the range. \square

Now we are ready to prove a local limit theorem for $G(n, m)$.

Theorem 4.2. *Let H be a connected graph, and fix $\lambda > 0$. Choose $n \geq 1$ and m such that $p = m/\binom{n}{2} \in (\lambda, 1 - \lambda)$, and let X_H be the number of times H appears as a subgraph of the random graph $G(n, m)$. Let μ_H, σ_H be the mean and standard deviation of this random variable. Finally, define $Z_H = (X_H - \mu_H)/\sigma_H$. Then we have*

$$|\sigma_H \mathbb{P}[Z_H = z] - \mathcal{N}(z)| \lesssim_{H, \lambda, \varepsilon} n^{\varepsilon-1/2}$$

for all $z \in (\mathbb{Z} - \mu_H)/\sigma_H$ and

$$\sum_{z \in (\mathbb{Z} - \mu_H)/\sigma_H} |\mathbb{P}[Z_H = z] - \mathcal{N}(z)/\sigma_H| \lesssim_{H, \lambda, \varepsilon} n^{\varepsilon-1/2}$$

for all $\varepsilon > 0$.

Proof. Let W, \mathcal{K}, σ be defined as earlier. Then, by Lemma 2.4, we have for $z \in (\mathbb{Z} - W_0 - W_2)/\sigma$ that

$$|\sigma \mathbb{P}[\mathcal{K} = z] - \mathcal{N}(z)| \leq e^{-\pi^2 \sigma^2/2} + \int_{-\pi\sigma}^{\pi\sigma} |\varphi_{\mathcal{K}}(t) - \varphi_{\mathcal{N}(0,1)}(t)| dt \lesssim_{H, \lambda, \varepsilon} \frac{1}{n^{1/2-\varepsilon}}$$

for all $\varepsilon > 0$, combining (4.2), (4.3), and (4.4) for different integration ranges. This is a local central limit theorem, with one minor technical issue, which is that \mathcal{K} has neither mean 0 nor variance 1. In particular,

$$Z_H = \frac{\sigma}{\sigma_H}(\mathcal{K} - \mathbb{E}\mathcal{K}).$$

But now we recall $\sigma/\sigma_H = 1 + O_\varepsilon(n^{\varepsilon-1/2})$ and $\mathbb{E}\mathcal{K} \lesssim_\varepsilon n^{\varepsilon-1/2}$, which follow from Lemma 3.4. Thus,

$$\mathbb{P}[Z_H = z] = \mathbb{P}[\mathcal{K} = z(\sigma_H/\sigma) + \mathbb{E}\mathcal{K}]$$

is near $(1/\sigma)\mathcal{N}(z(\sigma_H/\sigma) + \mathbb{E}\mathcal{K})$, which is near $(1/\sigma_H)\mathcal{N}(z)$, and the necessary bounds follow using that $\mathcal{N}(z)$ is Lipschitz. To deduce the second statement, we use what we have already proved along with Lemma 2.5. We simply need to verify that $\mathbb{P}[|Z_H| > n^\varepsilon]$ is small. Since the standard deviation of Z_H is 1, this follows immediately by hypercontractivity (Theorem 2.1) along with our trick of transferring bounds to the slice. \square

4.2 | Induced subgraph counts

Let H be a graph, not necessarily connected, with $\ell \geq 3$ vertices. Let $q = -p/(1-p)$, which is negative and bounded away from zero as well as bounded in size in terms of λ . Write

$$W = \sum_{\substack{H' \subseteq \binom{[n]}{2} \\ H' \simeq H}} \prod_{e \in E(H')} x_e \prod_{e \in \bar{E}(H')} (1 - x_e),$$

where $\bar{E}(H')$ is the complement of $E(H')$ within the set of all possible edges $\binom{V(H')}{2}$. We expand

$$\begin{aligned} W &= \sum_{\substack{H' \subseteq \binom{[n]}{2} \\ H' \simeq H}} \prod_{e \in E(H')} (p + \sqrt{p(1-p)}\chi_e) \prod_{e \in \bar{E}(H')} (1 - p - \sqrt{p(1-p)}\chi_e) \\ &= p^{e(H)}(1-p)^{\bar{e}(H)} \sum_{S \subseteq K_\ell} p^{-e(S)/2}(1-p)^{e(S)/2} f_{S,H}(q) \binom{n-v(S)}{\ell-v(S)} \gamma_S(\mathbf{x}), \end{aligned} \quad (4.5)$$

where the sum is over subgraphs S (lacking isolated vertices) of K_ℓ up to isomorphism. Here $f_{S,H}(q)$ is a polynomial in q with positive coefficients, computed as the sum

$$f_{S,H}(q) = \sum_{\substack{S', H' \subseteq K_\ell \\ S' \simeq S, H' \simeq H}} q^{|E(S') \setminus E(H')|}$$

which is taken over embeddings of S, H into K_ℓ . In particular, $f_{S,H}$ is a nonzero polynomial for each subgraph S of K_ℓ . For the empty graph, we obtain the constant polynomial $\ell!/\text{aut } H$.

Now, in order for W to satisfy the hypotheses of Subsection 3.1, we need there to be a term γ_S on k vertices for each $3 \leq k \leq \ell$ which has the correct order of magnitude. To ensure this, we merely need q to be bounded away from a root of $f_{S,H}$. Simply let $S = K_{1,k}$. Then we see $f_{S,H}$ has degree at most k , hence has at most k roots. Therefore, as long as q is bounded away from a set of at most $3 + 4 + \dots + \ell < \ell^2$ values, or equivalently p is bounded away (say by λ) from at most ℓ^2 values as well as $\{0, 1\}$, the necessary hypotheses will be satisfied.

In particular, define σ, \mathcal{K} from W in the same way as in Subsection 3.1. Then for $|t| \leq n^\varepsilon$ we have by Lemma 3.3 that

$$|\varphi_{\mathcal{K}}(t) - e^{-t^2/2}| \lesssim_\lambda \frac{|t|}{n^{\frac{1}{2}-\varepsilon}} + \exp(-\Omega_\lambda((\log n)^2)) \quad (4.6)$$

and for $n^\varepsilon \leq |t| \leq \sigma n^{-\varepsilon}$ we have by Lemma 3.6 that

$$|\varphi_{\mathcal{K}}(t) - e^{-t^2/2}| \lesssim n^{-\Omega_{\lambda,\varepsilon}(\log \log n)}. \quad (4.7)$$

Now we present a decoupling which handles the top range $\sigma n^{-\varepsilon} \leq |t| \leq \pi\sigma$.

Lemma 4.3. *Let W be as in (4.5), and define σ, \mathcal{K} as in Subsection 3.1. Suppose p is bounded away by λ from a set of ℓ^2 values. Then for $|t| \leq \pi\sigma$,*

$$|\varphi_{\mathcal{K}}(t)| \leq \exp(-\Omega_{\lambda}(n)) + \exp\left(-\Omega_{\lambda}\left(\frac{t^2 n}{\sigma^2}\right)\right). \quad (4.8)$$

Proof. For simplicity, without loss of generality we assume H is connected. We can do this because replacing H with its complement in K_{ℓ} , replacing p by $1 - p$, and replacing W by $\ell! \binom{n}{\ell} / \text{aut } H - W$ keeps the random variable the same, and either H or its complement is connected.

After doing this, we use the same decoupling as in Subsection 4.1. Partition the vertex set $[n]$ into $\lfloor n/\ell \rfloor$ cliques of size ℓ , with at most $\ell - 1$ extra vertices. Label the vertices of H by $[\ell]$ arbitrarily, and its edges from 0 to $e(H) - 1$. Within each clique take an isomorphic copy of this labeled H . Let \widetilde{B}_0 be the union of all unlabeled edges along with those labeled 0, and B_i be all the edges labeled i for $0 \leq i \leq e(H) - 1$. Thus, B_i for $1 \leq i \leq e(H) - 1$ and \widetilde{B}_0 partition the edges. Finally, define $\widetilde{X} \in \{0, 1\}^{\widetilde{B}_0}$ as the indicator vector of which edges are included in $G(n, m)$ and Z_i for $1 \leq i \leq e(H) - 1$ as the number of edges in each set B_i when sampling from $G(n, m)$. Let Y_i^0, Y_i^1 for $1 \leq i \leq e(H) - 1$, similarly, be independent copies of the $G(n, m)$ draw given Z_i . Also let Y_0 be the indicator of the edges of \widetilde{X} in B_0 only, and \widetilde{Y}_0 be the indicator of the edges in $\widetilde{B}_0 \setminus B_0$ only. Though the decoupling is the same, the resulting decoupled function is more complex. We first show that we can write

$$\alpha(W)(\widetilde{X}, \mathbf{Y}) = \delta_{\emptyset}(\mathbf{Y}, \widetilde{Y}_0) + \sum_{e \in B_0} \delta_e(\mathbf{Y}, \widetilde{Y}_0) x_e$$

for polynomials $\delta_e, e \in B_0$. To prove this, we consider which terms in the definition of W (in the x basis) provide a term in $\alpha(W)$ dependent on x_e . For a term to not become zero, it must have an edge from each of B_i for $i \geq 1$, as well as the edge $e \in B_0$. Consider a term corresponding to a copy H' of H . We claim that $V(H') = V(H_e)$, where H_e is the copy of H within the clique that contains e . First, since the above shows we have an edge in B_i for each $i \geq 1$, and since H has no vertex isolated, we have that $V(H')$ has at least 1 vertex with each label from $[\ell]$. Since $v(H') = \ell$, this means it has each label exactly once. Now, if vertex labels a and b are connected by an edge labeled c within our labeled version of H , then to ensure an edge from B_c exists in H' , the unique elements of $V(H')$ labeled by a and b must be in the same one of the $\lfloor n/\ell \rfloor$ cliques. This fact, along with $e \in E(H')$ and the connectedness of H , immediately shows that $V(H') = V(H_e)$.

But now this means the only terms with a dependence on x_e are the terms

$$W_e = \sum_{\substack{H' \subseteq \binom{V(H_e)}{2} \\ H' \simeq H}} \prod_{e' \in E(H')} x_{e'} \prod_{e' \in \widetilde{E}(H')} (1 - x_{e'}).$$

These terms do not contain any other $x_{e'}$ for $e' \in B_0$, hence we obtain only linear terms in $\alpha(W)$ when collecting in the variable set $\{x_e\}_{e \in B_0}$. That is, $\alpha(W)$ is of the claimed form. Not only that, but we now know how to explicitly compute each δ_e .

In fact, we will merely compute $\delta_e(\mathbf{Y}, \mathbf{0})$. In this case,

$$W_e = \prod_{e' \in E(H_e)} x_{e'}$$

since the remaining terms corresponding to $H' \neq H_e$ have a factor of $x_{e'}$ for $e' \in \widetilde{B_0} \setminus B_0$, which were set to 0, and since the terms from $e' \in \widetilde{E}(H_e)$, equal to $1 - x_{e'}$, merely become 1. Therefore, $\delta_e(\mathbf{Y}, \mathbf{0})$ equals what it did in Subsection 4.1, namely

$$\delta_e(\mathbf{Y}, \mathbf{0}) = \prod_{e' \in E(H_e) \setminus e} (x_{e'}^1 - x_{e'}^0).$$

In fact, this formula still holds as long as just $x_{e'} = 0$ for all $e' \in \widetilde{E}(H_e)$ (rather than all of $\widetilde{Y_0}$).

Now we claim that with extremely high probability, the number of $e \in B_0$ such that $\delta_e(\mathbf{Y}) = 1$ is greater than $\lambda^{\ell^2} n / (2\ell)$ and the number such that $\delta_e(\mathbf{Y}) = 0$ satisfies the same. This is clear in the $G(n, p)$ model, as there are more than $n / (2\ell)$ edges $e \in B_0$, and so long as $x_{e'} = 0$ for all $e' \in \widetilde{E}(H_e)$ and

$$\prod_{e' \in E(H_e) \setminus e} (x_{e'}^1 - x_{e'}^0) = 1$$

(or 0, respectively) we have the desired event for e . Independence finishes: the probability of this event not occurring in the independent model is $\exp(-\Omega_\lambda(n))$.

In the $G(n, m)$ model, as in the proof of Lemma 4.1 we repeatedly use Azuma–Hoeffding. First, it demonstrates that each Z_i for $1 \leq i \leq e(H) - 1$ is approximately pn/ℓ with high probability, say, within the interval $[pn/(2\ell), (1+p)n/(2\ell)]$ with probability $1 - \exp(-\Omega_\lambda(n))$. Conditional on a realization of the Z_i , the vectors Y_i^0, Y_i^1 are independent and uniform with a fixed sum. By Azuma–Hoeffding again, we can show that $x_{e'}^1 = 1$ and $x_{e'}^0 = 0$ for each $e' \in E(H_e) \setminus e$ as well as $x_{e'} = 0$ for all $e' \in \widetilde{E}(H_e)$, simultaneously, in at least $\Omega_\lambda(n)$ of our cliques with probability $1 - \exp(-\Omega_\lambda(n))$. Similarly, we can show that $x_{e'}^1 = x_{e'}^0 = 1$ for some $e' \in E(H_3) \setminus e$ happens in at least $\Omega_\lambda(n)$ of our cliques with a similar probability.

We also control the number of edges among Y_0 . Note that its distribution is the same as looking at the number of edges in a specific subset of $G(n, m)$. By Azuma–Hoeffding, with a process revealing edges within B_0 one at a time, we see with probability $1 - \exp(-\Omega_\lambda(n))$ the fraction of edges chosen in this set is in $(p/2, (1+p)/2)$. Therefore, over the randomness of Y_i^0 and $\widetilde{Y_0}$, say, the number of edges in Y_0 is fixed to some value that is $\Theta_\lambda(n)$.

Now we are ready to apply Lemma 2.10. Let B'_0 be the set of $e \in B_0$ satisfying $\delta_e \notin \{0, 1\}$ (which depends on $\mathbf{Y}, \widetilde{Y_0}$). Furthermore, let Y'_0 be the vector (x_e) for $e \in B'_0$. We obtain

$$\begin{aligned} |\varphi_K(t)|^{2^{e(H)-1}} &\leq \mathbb{E}_{\mathbf{Y}} |\mathbb{E}_{\widetilde{X}} e^{it\alpha(W)(\widetilde{X}, \mathbf{Y})/\sigma}| \leq \mathbb{E}_{\mathbf{Y}, \widetilde{Y_0}} |\mathbb{E}_{Y_0} e^{it\alpha(W)(\widetilde{X}, \mathbf{Y})/\sigma}| \\ &\leq \mathbb{E}_{\mathbf{Y}, \widetilde{Y_0}, Y'_0} |\mathbb{E}_{Y_0 \setminus Y'_0} e^{(it/\sigma) \sum_{e \in B_0 \setminus B'_0} \delta_e(\mathbf{Y}, \widetilde{Y_0}) x_e}| \leq \exp(-\Omega_\lambda(n)) + \exp\left(-\Omega_\lambda\left(\frac{t^2 n}{\sigma^2}\right)\right), \end{aligned}$$

the last inequality using that the function is bounded by 1 in the rare cases delineated above, and using Lemma 2.7 in the remaining cases in which we know that the inner x_e for $e \in Y_0 \setminus Y'_0$ are drawn uniformly with a fixed sum depending on $\mathbf{Y}, \widetilde{Y_0}, Y'_0$. That sum is $\Theta_\lambda(n)$ in size, and additionally we use that a positive fraction (in terms of λ, H) of coefficients $\delta_e(\mathbf{Y})$ are 1 as well as 0. Note that Lemma 2.7 only applies if $(t/\sigma) \cdot (1 - 0) \leq \pi$, which precisely hits the top of the range. \square

Now, in exactly the same way as for Theorem 4.2, we deduce a local limit theorem for induced subgraphs.

Theorem 4.4. *Let H be a graph, and fix $\lambda > 0$. There is a set $\mathcal{P}_{\text{crit}}$ of size at most $v(H)^2$ such that the following holds. Choose $n \geq 1$ and m such that $p = m/\binom{n}{2}$ is λ -separated from $\{0, 1\} \cup \mathcal{P}_{\text{crit}}$, and let X_H be the number of times H appears as an induced subgraph of the random graph $G(n, m)$. Let μ_H, σ_H be the mean and standard deviation of this random variable. Finally, define $Z_H = (X_H - \mu_H)/\sigma_H$. Then we have*

$$|\sigma_H \mathbb{P}[Z_H = z] - \mathcal{N}(z)| \lesssim_{H, \lambda, \varepsilon} n^{\varepsilon-1/2}$$

for all $z \in (\mathbb{Z} - \mu_H)/\sigma_H$ and

$$\sum_{z \in (\mathbb{Z} - \mu_H)/\sigma_H} |\mathbb{P}[Z_H = z] - \mathcal{N}(z)/\sigma_H| \lesssim_{H, \lambda, \varepsilon} n^{\varepsilon-1/2}$$

for all $\varepsilon > 0$.

Remark. In fact, as noted in Section 1, we can reduce our bound on the number of critical values to $O(v(H))$. This follows from the remark at the end of Subsection 3.4, which demonstrates that we only need to ensure a dyadically separated set of k have nontrivial coefficients. Of course, since there are many more graphs on k vertices than merely $K_{1,k}$, and only one of them must be nonzero, it is likely that there are even fewer critical values than we can prove.

5 | INDEPENDENT MODELS

In this section, we deduce a local limit theorem for subgraph counts of connected graphs in $G(n, p)$ from the corresponding result for $G(n, m)$. It is worth remarking that all the earlier calculations done to prove the $G(n, m)$ case can be done analogously and with much more simplicity (as variables are actually independent) to directly prove the $G(n, p)$ case. However, the method of transfer is still important as it will allow us to prove more general results such as a local limit theorem in the k -AP case. For the sake of not belaboring the issue we prove the reduction only for Theorem 1.1, as the analysis for Theorem 1.2 is completely analogous.

Proof of Theorem 1.1. Let X_H denote the number of copies of H . Let $\sigma_H^2 = \text{Var}[X_H|G(n, p)]$ and $\mu_H = \mathbb{E}[X_H|G(n, p)]$. We now recall from Subsection 4.1 that

$$\begin{aligned} X_H &= \sum_{\substack{H' \subseteq \binom{[n]}{2} \\ H' \simeq H}} \prod_{e \in E(H')} (p + \sqrt{p(1-p)} \chi_e) \\ &= \sum_{S \subseteq H} p^{e(H)-e(S)} (\sqrt{p(1-p)})^{e(S)} c_{S,H} d_{S,H} \binom{n-v(S)}{\ell-v(S)} \gamma_S(\mathbf{x}), \end{aligned}$$

where the sum is over subgraphs S (lacking isolated vertices) of H up to isomorphism. Recall that as before, $c_{S,H}$ explicitly equals $(\ell - v(S))! \text{aut } S / \text{aut } H$ and $d_{S,H}$ equals the number of times S

appears as a subgraph of H , for example, $d_{K_2, H} = e(H)$. Let

$$X_2 = p^{e(H)-1}(\sqrt{p(1-p)})c_{S, K_2}e(H)\binom{n-2}{\ell-2}\gamma_{K_2}(\mathbf{x})$$

and

$$X_{\text{rem}} = X_H - p^{e(H)}\binom{n}{\ell}\frac{\ell!}{\text{aut } H} - X_2.$$

First note by direct computation that if $p \in (\lambda, 1 - \lambda)$ then

$$\text{Var}[X_{\text{rem}}|G(n, p)] = (1 + O_\lambda(\log n/n)) \text{Var}[X_{\text{rem}}|G(n, p')]$$

if $p' = (1 + \Theta(\log n/n))p$. Given this and the deductions in Lemma 3.4 that

$$\text{Var}[X_H|G(n, m)] = \text{Var}[X_{\text{rem}}|G(n, m)] = (1 + O_{\lambda, \varepsilon}(n^{\varepsilon-1/2})) \text{Var}[X_{\text{rem}}|G(n, q)]$$

(here $q = m/\binom{n}{2} \in (\lambda, 1 - \lambda)$), we find for any $m, m' \in [p\binom{n}{2} - n \log n, p\binom{n}{2} + n \log n]$ that

$$\begin{aligned} \text{Var}[X_H|G(n, m)] &= (1 + O_{\lambda, \varepsilon}(n^{\varepsilon-1/2})) \text{Var}[X_H|G(n, m')] \\ &= (1 + O_{\lambda, \varepsilon}(n^{\varepsilon-1/2})) \text{Var}[X_{\text{rem}}|G(n, p)]. \end{aligned}$$

From now we denote $\sigma^2 = \text{Var}[X_{\text{rem}}|G(n, p)]$. We now explicitly use that the expectation of X_H varies essentially linearly given the number of edges. In particular, for $m \in [p\binom{n}{2} - n \log n, p\binom{n}{2} + n \log n]$, note by linearity of expectation that

$$\begin{aligned} \mathbb{E}[X_H|G(n, m)] &= \binom{n}{\ell}\frac{\ell!}{\text{aut } H} \prod_{i=0}^{e(H)-1} \left(\frac{m-i}{\binom{n}{2}-i} \right) \\ &= \binom{n}{\ell}\frac{\ell!}{\text{aut } H} \left(p^{e(H)} + e(H)p^{e(H)-1} \left(\frac{m-p\binom{n}{2}}{\binom{n}{2}} \right) \right) (1 + \Theta_\lambda((\log n)^2/n^2)). \end{aligned}$$

An essentially similar estimate was derived in Lemma 3.4 for more general graph statistics. Now, for the sake of clarity define

$$f(m) = \binom{n}{\ell}\frac{\ell!}{\text{aut } H} \left(p^{e(H)} + e(H)p^{e(H)-1} \left(\frac{m-p\binom{n}{2}}{\binom{n}{2}} \right) \right).$$

Finally, we are in a position to explicitly calculate the distribution of X_H under $G(n, p)$. Let $\sigma_m^2 = \text{Var}[X_H|G(n, m)]$ and $\mu_m = \mathbb{E}[X_H|G(n, m)]$. Now note that

$$\begin{aligned} \mathbb{P}[X_H = x] &= \sum_{m \in \mathbb{Z}} \mathbb{P}[X_H = x|G(n, m)] \mathbb{P}\left[\sum x_e = m\right] \\ &= \sum_{m \in [p\binom{n}{2} - n \log n, p\binom{n}{2} + n \log n]} \mathbb{P}[X_H = x|G(n, m)] \mathbb{P}\left[\sum x_e = m\right] + \exp(-\Omega_\lambda((\log n)^2)) \end{aligned}$$

where we have used Chernoff to bound the probability that number of edges deviates too far from the mean. For the sake of clarity we will implicitly assume that x is within $\sigma_H(\log n)^C$ of the mean; for x outside this range and C sufficiently large the probability of attaining x is super-polynomially small by hypercontractivity (Theorem 2.1) so the desired statement is trivial. This assumption will be used implicitly later on. Now let \mathcal{M}_x denote the set of m such that

$$|x - f(m)| \leq \sigma(\log n)^C$$

and

$$m \in \left[p\binom{n}{2} - n \log n, p\binom{n}{2} + n \log n \right]$$

for a suitably large C . Now suppose that $m \in [p\binom{n}{2} - n \log n, p\binom{n}{2} + n \log n] \setminus \mathcal{M}_x$. Then

$$\begin{aligned} \mathbb{P}[X_H = x | G(n, m)] &\leq \mathbb{P}[|X_{\text{rem}}| \geq \sigma(\log n)^C / 2] / \mathbb{P}\left[\sum x_e = m\right] \\ &\lesssim \exp(-\Omega_\lambda((\log n)^2)), \end{aligned}$$

using that $\mathbb{P}[\sum x_e = m] \gtrsim \exp(-O_\lambda((\log n)^2))$ and then choosing C sufficiently large, so that the bound coming from hypercontractivity (Theorem 2.1) on the numerator is sufficiently strong. The key point is that since $f(m)$ is a linear function with slope $\Theta_\lambda(n^{\ell-2})$ we have $|\mathcal{M}_x| = \Theta_\lambda(n^{1/2}(\log n)^C)$. Thus, we have that

$$\begin{aligned} \mathbb{P}[X_H = x] &= \sum_{\substack{m \in [p\binom{n}{2} - n \log n, \\ p\binom{n}{2} + n \log n]}} \mathbb{P}[X_H = x | G(n, m)] \mathbb{P}\left[\sum x_e = m\right] + \exp(-\Omega_\lambda((\log n)^2)) \\ &= \sum_{m \in \mathcal{M}_x} \mathbb{P}[X_H = x | G(n, m)] \mathbb{P}\left[\sum x_e = m\right] + \exp(-\Omega_\lambda((\log n)^2)). \end{aligned}$$

Now using Theorem 4.2 and that σ_m is approximately equal to σ , the last summation equals

$$\begin{aligned} &\sum_{m \in \mathcal{M}_x} \left(\frac{1}{\sigma_m} \mathcal{N}\left(\frac{x - \mu_m}{\sigma_m}\right) + O_\lambda\left(\frac{n^{\frac{\varepsilon-1}{2}}}{\sigma}\right) \right) \mathbb{P}\left[\sum x_e = m\right] \\ &= \sum_{m \in \mathcal{M}_x} \frac{1}{\sigma_m} \mathcal{N}\left(\frac{x - \mu_m}{\sigma_m}\right) \mathbb{P}\left[\sum x_e = m\right] + O_\lambda\left(\frac{|\mathcal{M}_x|}{n^{\frac{3-\varepsilon}{2}} \sigma}\right), \end{aligned}$$

where we use that probability of having a given number of edges is $O_\lambda(1/n)$. Now note that σ_H is order $n^{1/2}$ larger than σ . Therefore, the error term can be seen to be $O_\lambda(n^{\varepsilon-1/2}\sigma_H^{-1})$, which is the correct magnitude. Now $\sigma_m = (1 + O_\lambda(n^{\varepsilon-1/2}))\sigma$ and $\mu_m = f(m) + O_\lambda((\log n)^2 n^{-1/2}\sigma)$ for all $m \in [p\binom{n}{2} - n \log n, p\binom{n}{2} + n \log n]$ by the remarks which began the section. It follows that

$$\begin{aligned} \sum_{m \in \mathcal{M}_x} \frac{1}{\sigma_m} \mathcal{N}\left(\frac{x - \mu_m}{\sigma_m}\right) \mathbb{P}\left[\sum x_e = m\right] \\ = \sum_{m \in \mathcal{M}_x} \frac{1}{\sigma} \mathcal{N}\left(\frac{x - f(m)}{\sigma}\right) \mathbb{P}\left[\sum x_e = m\right] + O_\lambda\left(\frac{n^{\varepsilon-1/2}}{\sigma_H}\right). \end{aligned}$$

At this point the rest is elementary calculation. Let m^* be the solution to $f(m^*) = x$ and note that $|m - m^*| \lesssim_\lambda n^{1/2}(\log n)^C$ since f has slope $\Theta_\lambda(\sigma n^{-1/2})$. This is enough to conclude that $\mathbb{P}[\sum x_e = m]$ is essentially constant over $m \in \mathcal{M}_x$, close enough to replace the above with

$$\begin{aligned} \mathbb{P}\left[\sum x_e = \lfloor m^* \rfloor\right] \sum_{m \in \mathcal{M}_x} \frac{1}{\sigma} \mathcal{N}\left(\frac{x - f(m)}{\sigma}\right) \\ = \mathbb{P}\left[\sum x_e = \lfloor m^* \rfloor\right] \sum_{m \in \mathbb{Z}} \frac{1}{\sigma} \mathcal{N}\left(\frac{x - f(m)}{\sigma}\right) + O_\lambda(\exp(-(\log n)^2)) \end{aligned}$$

without increasing the error term. Now since \mathcal{N} is continuous, unimodal, and integrable, and since f has slope

$$\eta = \binom{n}{\ell} \frac{\ell!}{\text{aut } H} \frac{e(H)p^{e(H)-1}}{\binom{n}{2}},$$

standard results on Riemann approximation show that this equals

$$\mathbb{P}\left[\sum x_e = \lfloor m^* \rfloor\right] \frac{1 + O_\lambda(n^{-1/2})}{\eta}.$$

Again the error term is acceptable, and using Stirling's approximation shows that this is approximately

$$\frac{1}{\sqrt{p(1-p)\binom{n}{2}}} \frac{1}{\eta} \mathcal{N}\left(\frac{m^* - p\binom{n}{2}}{\sqrt{p(1-p)\binom{n}{2}}}\right)$$

Finally, we note that by calculation that

$$\frac{m^* - p\binom{n}{2}}{\sqrt{p(1-p)\binom{n}{2}}} = \frac{x - \mathbb{E}[X_H | G(n, p)]}{\eta \cdot \sqrt{p(1-p)\binom{n}{2}}}, \text{ and } \eta \cdot \sqrt{p(1-p)\binom{n}{2}} = (1 + O_{H,p}(1/n))\sigma_H.$$

These two estimates, combined with the rest in this proof finally give that

$$\mathbb{P}[X_H = x] = \frac{1}{\sigma_H} \mathcal{N}\left(\frac{x - \mu_H}{\sigma_H}\right) + O_\lambda\left(\frac{n^{\varepsilon-1/2}}{\sigma_H}\right).$$

To deduce the necessary L^1 bound, use Lemma 2.5 and hypercontractivity (Theorem 2.1). \square

6 | COUNTEREXAMPLES

In this section, we establish counterexamples to some anticoncentration conjectures of Fox, Kwan, and Sauermann [11]. The main technical result is that the following class of graph-related polynomials do not exhibit anticoncentration. Note here that we are using the notation of graph factors established in Subsection 2.5.

Theorem 6.1. *Let $\chi_e = (x_e - p)/\sqrt{p(1-p)}$ where x_e are independent Bernoulli random variables with expectation $p \in (0, 1)$ for all $e \in \binom{[n]}{2}$. Suppose that*

$$F(\mathbf{x}) = \sum_{H: v(H) \leq \ell} \binom{n - v(H)}{\ell - v(H)} \Phi_H \gamma_H(\mathbf{x})$$

where Φ_H are constants independent of n and satisfy

- (1) $\Phi_H = 0$ for all connected graphs on 3 and 4 vertices;
- (2) $\Phi_H \neq 0$ for H being an edge and H being the disjoint union of 2 edges.

Furthermore, suppose that F is integer valued. Then there exists a sequence y_n such that

$$\mathbb{P}[F(\mathbf{x}) = y_n] \gtrsim_{H,p} n^{3/2-\ell}.$$

Remark. Recall that γ_H is only defined if H has no isolated vertices. Also, note that the standard deviation of F is of order $n^{\ell-1}$ since $\Phi_{K_2} \neq 0$ and therefore anticoncentration fails by order $n^{1/2}$. Furthermore, such polynomials are $\mathcal{U}_3 \cup \mathcal{U}_4^c$ -proportional (and not \mathcal{U}_2 -proportional) in the notation of Janson [18]. This result is ultimately derived from the results in [17] along with a conditioning argument.

Proof. Note that the probability that the number of edges is $\lfloor p \binom{n}{2} \rfloor$ is $\Omega_p(n^{-1})$. Due to results of Janson [17], statistics satisfying the hypothesis of this theorem converge to Gaussians of standard deviation $\Theta_p(n^{\ell-5/2})$ (see [17, Theorem III.8]) in $G(n, m)$. The reason for this is

$$\begin{aligned} 2\gamma_{K_2+K_2}(\mathbf{x}) &= \left(\sum_e \chi_e \right)^2 - \sum_e \chi_e^2 - 2 \sum_{i,j,k} \chi_{i,j} \chi_{i,k} \\ &= \gamma_{K_2}(\mathbf{x})^2 - \sum_e \chi_e^2 - 2\gamma_{K_{1,2}}(\mathbf{x}), \end{aligned}$$

hence the number of edges determines the first two terms. (This identity also appears as [18, Theorem 4].) Thus, given the number of edges the standard deviation coming from the $K_2 + K_2$ term is only $\Theta_p(n^{\ell-5/2})$ in $G(n, m)$ (for m near $p \binom{n}{2}$) instead of $\Theta_p(n^{\ell-2})$ as would be ‘typical’.

In particular, by the aforementioned theorem of Janson, we have a sequence α_n and $\beta_n = \Theta_{F,p}(n^{\ell-5/2})$ such that

$$\frac{F(\mathbf{x}) - \alpha_n}{\beta_n} \xrightarrow{d} \mathcal{N}(0, 1)$$

if the edges are sampled in $G(n, m)$ with $m = \lfloor p \binom{n}{2} \rfloor$. Therefore, we have that

$$\begin{aligned} \mathbb{P}[|F(\mathbf{x}) - \alpha_n| \leq n^{\ell-5/2}] &\geq \mathbb{P}\left[|F(\mathbf{x}) - \alpha_n| \leq n^{\ell-5/2} \cap \sum_e x_e = \left\lfloor p \binom{n}{2} \right\rfloor\right] \\ &\gtrsim_p (c_H + o(1))/n \gtrsim_{p,H} 1/n, \end{aligned}$$

where in the final step we have used convergence in distribution to a Gaussian. The theorem then follows as one of the integers in the range $[\alpha_n - n^{\ell-5/2}, \alpha_n + n^{\ell-5/2}]$ has the desired property. \square

This immediately disproves the conjecture of Fox, Kwan, and Sauermann regarding anticoncentration of subgraph counts as the polynomial counting the number of copies of $K_2 + K_2$ trivially satisfies the conditions of Theorem 6.1, so it in turn proves the first part of Theorem 1.3. We remark that in fact that any disjoint union of edges also trivially satisfies the conditions of Theorem 6.1.

We similarly disprove the conjecture of Fox, Kwan, and Sauermann regarding anticoncentration of induced subgraph counts (the second part of Theorem 1.3) by constructing a graph H for which the polynomial counting induced subgraphs of H satisfies the conditions of Theorem 6.1. However, the construction here is significantly more intricate. We sketch in greater detail how to arrive at such a graph. The construction here is closely related to the method used by Kärrmann [21]. The algorithm we use to search for an appropriate graph is based on that work.

The key idea here is to note that whether Φ_H vanishes or not is a condition that is independent of n and is only based on the density of various 4-vertex subgraphs present in H . In particular note that for a given graph H the number of induced copies of H is

$$X_H = \sum_{\substack{V \subseteq [n] \\ |V|=v(H)}} \sum_{\substack{E \subseteq \binom{V}{2} \\ E \simeq H}} \prod_{e \in E} x_e \prod_{e \in \binom{V}{2} \setminus E} (1 - x_e).$$

Letting $\chi_e = (x_e - p)/\sqrt{p(1-p)}$ we have

$$\frac{X_H}{p^{e(H)}(1-p)^{\bar{e}(H)}} = \sum_{\substack{V \subseteq [n] \\ |V|=v(H)}} \sum_{\substack{E \subseteq \binom{V}{2} \\ E \simeq H}} \prod_{e \in E} (1 + \sqrt{(1-p)/p} \chi_e) \prod_{e \in \binom{V}{2} \setminus E} (1 - \sqrt{p/(1-p)} \chi_e)$$

where we recall $\bar{e}(H) = \binom{v(H)}{2} - e(H)$. This can be rewritten as

$$\frac{X_H}{p^{e(H)}(1-p)^{\bar{e}(H)}} = \sum_{\substack{V \subseteq [n] \\ |V|=v(H)}} \sum_{\substack{E \subseteq \binom{V}{2} \\ E \simeq H}} \sum_{T \subseteq \binom{V}{2}} \prod_{e \in E \cap T} \left(\sqrt{\frac{1-p}{p}} \chi_e \right) \prod_{e \in T \setminus E} \left(-\sqrt{\frac{p}{1-p}} \chi_e \right).$$

Let δ_T be the coefficient of $\gamma_T(\mathbf{x})$ in the expansion of X_H . Let $q = -p/(1-p)$ and $\text{ind}(S, H)$ be the number of induced subgraphs of H isomorphic to S . We find that for any T without isolated vertices,

$$\delta_T = \left(\sqrt{\frac{1-p}{p}} \right)^{e(T)} \frac{\binom{n}{v(H)} v(H)! \text{aut } T}{\binom{n}{v(T)} v(T)! \text{aut } H} \sum_{v(S)=v(T)} \delta_{S,T}(q) \text{ind}(S, H)$$

K_4 -embedding:	1	6	12	4	3	12	1	6	12	4	3
Induced $S \setminus T$	K_0	K_2	P_2	K_3	$K_2 + K_2$	P_3	K_4	$K_4 - K_2$	$K_4 - P_2$	$K_{1,3}$	C_4
D_4	1	$6q$	$12q^2$	$4q^3$	$3q^2$	$12q^3$	q^5	$6q^5$	$12q^4$	$4q^3$	$3q^4$
$K_2 + D_2$	1	$5q + 1$	$8q^2 + 4q$	$2q^3 + 2q^2$	$2q^2 + q$	$6q^3 + 6q^2$	q^5	$q^5 + 5q^4$	$4q^4 + 8q^3$	$2q^3 + 2q^2$	$q^4 + 2q^3$
$P_2 + K_1$	1	$4q + 2$	$5q^2 + 6q + 1$	$q^3 + 2q^2 + q$	$q^2 + 2q$	$2q^3 + 8q^2 + 2q$	q^4	$2q^4 + 4q^3$	$q^4 + 6q^3 + 5q^2$	$q^3 + 2q^2 + q$	$2q^3 + q^2$
$K_2 + K_2$	1	$4q + 2$	$4q^2 + 8q$	$4q^2$	$2q^2 + 1$	$4q^3 + 4q^2 + 4q$	q^4	$2q^4 + 4q^3$	$8q^3 + 4q^2$	$4q^2$	$q^4 + 2q^2$
$K_3 + K_1$	1	$3q + 3$	$3q^2 + 6q + 3$	$3q^2 + 1$	$3q$	$6q^2 + 6q$	q^3	$3q^3 + 3q^2$	$3q^3 + 6q^2 + 3q$	$q^3 + 3q$	$3q^2$
P_3	1	$3q + 3$	$2q^2 + 8q + 2$	$2q^2 + 2q$	$q^2 + q + 1$	$q^3 + 5q^2 + 5q + 1$	q^3	$3q^3 + 3q^2$	$2q^3 + 8q^2 + 2q$	$2q^2 + 2q$	$q^3 + q^2 + q$
$K_{1,3}$	1	$3q + 3$	$3q^2 + 6q + 3$	$q^3 + 3q$	$3q$	$6q^2 + 6q$	q^3	$3q^3 + 3q^2$	$3q^3 + 6q^2 + 3q$	$3q^2 + 1$	$3q^2$
C_4	1	$2q + 4$	$8q + 4$	$4q$	$q^2 + 2$	$4q^2 + 4q + 4$	q^2	$4q^2 + 2q$	$4q^2 + 8q$	$4q$	$2q^2 + 1$
$-(P_2 + K_1)$	1	$2q + 4$	$q^2 + 6q + 5$	$q^2 + 2q + 1$	$2q + 1$	$2q^2 + 8q + 2$	q^2	$4q^2 + 2q$	$5q^2 + 6q + 1$	$q^2 + 2q + 1$	$q^2 + 2q$
$-(K_2 + D_2)$	1	$q + 5$	$4q + 8$	$2q + 2$	$q + 2$	$6q + 6$	q	$5q + 1$	$8q + 4$	$2q + 2$	$2q + 1$
K_4	1	6	12	4	3	12	1	6	12	4	3

FIGURE 1 Table of polynomials $\delta_{S,T}(q)$

for appropriate polynomials $\delta_{S,T}(q)$ with positive coefficients. Here the sum ranges over the isomorphism classes S of graphs on $v(T)$ vertices. Explicitly,

$$\delta_{S,T}(q) = \sum_{\substack{T' \subseteq \binom{V(S)}{2} \\ T' \simeq T}} q^{|E(T') \setminus E(S)|}.$$

This formula is derived by conditioning on which size $v(T)$ subset of V is used to embed T in. Such a subset yields an induced subgraph S of the copy of H considered. Then, summing over all embeddings of T gives the result.

Suppose $v(T) \leq 4$. Instead of conditioning on a size $v(T)$ subset of V , condition on a size 4 subset of V (possibly overcounting) that contains the copy of T we wish to count. A similar calculation yields

$$\delta_T \propto \sum_{v(S)=4} \delta_{S,T}(q) \text{ind}(S, H),$$

with $\delta_{S,T}$ defined in the same way as before. We do not bother computing the exact pre-factors, as we only ultimately care whether a term is zero or not. The following Figure 1 is a table of $\delta_{S,T}$. The graph D_k is the empty graph on k vertices, P_k the path on k edges, and $+$ denotes disjoint union while $-$ denotes complementation within K_4 . Thus, $-(K_2 + D_2)$, for instance, is K_4 without an edge. Now let $p = 1/2$, so that $q = -1$. We wish δ_T to be nonzero for $T = K_2$ but zero for connected graphs of size 3 and 4. This ultimately imposes eight linear conditions on $\text{ind}(S, H)$ when $v(S) = 4$, beyond the additional condition that the sum of all these is $\binom{v(H)}{4}$. Finally, induced subgraph counts of degree 4 satisfy an additional (quadratic) relation. The one further quadratic relation stems from the fact that $\text{ind}(K_2, H)^2$ can be expressed as a linear combination of $\text{ind}(S, H)$ for $v(S) = 4$ (with the resulting equations have a dependence on $v(H)$).

Thus, after we fix $v(H)$, the eleven counts

$$\text{ind}_4(H) = (\text{ind}(D_4, H), \text{ind}(K_2 + D_2, H), \dots, \text{ind}(K_4, H))$$

ought to be constrained by one parameter. (This is because there are eleven free parameters satisfying nine linear constraints and one quadratic constraint.) As it turns out, we end up needing 16 to divide $\binom{v(H)}{4}$, as well as satisfy some square root integrality constraints which constrain this one parameter. The fact that δ_{K_2} must be nonzero corresponds to the fact that H must not have exactly $\binom{v(H)}{2}/2$ edges.

Ultimately, the smallest possible example satisfies $v(H) = 64$ with

$$\text{ind}_4(H) = (11\ 835, 67\ 163, 126\ 632, 31\ 723, 39\ 646, 119\ 198, 39\ 646, 27\ 941, 111\ 504, 52\ 035, 8053).$$

There are actually ten resulting vectors (five after removing complementation symmetry). There is also a vector satisfying the constraints above except that $\delta_{K_2} = 0$; this in fact is precisely the vector derived in [21] as necessary for a 64-vertex superproportional graph. The above vector forces us to have 976 edges, which is the closest possible to $\binom{64}{2}/2 = 1008$ without actually being 1008.

Now, using a modification of an algorithm due to [21], we find such a graph H . The adjacency matrix is provided in the Appendix. Code in Java for both constructing and verifying the construction is provided in the arXiv listing of the paper. Finally, since this H has the number of induced copies X_H satisfy the hypotheses of Theorem 6.1, we see that the second half of Theorem 1.3 is proven.

7 | LOCAL LIMIT THEOREM FOR k -TERM ARITHMETIC PROGRESSIONS

Fix $k \geq 3$, which we will treat as constant throughout. Then fix $\lambda \in (0, 1/2)$ and choose $n \geq 1$ with $\gcd(n, (k-1)!) = 1$. Then choose an integer m with $p = m/n \in (\lambda, 1 - \lambda)$. We will show that

$$\text{kAP}(\mathbf{x}) = \sum_{a \in \mathbb{Z}/n\mathbb{Z}} \sum_{d \in [n/2]} \prod_{i=0}^{k-1} x_{a+id}$$

satisfies a local central limit theorem when we uniformly sample $\mathbf{x} = (x_i)_{i \in \mathbb{Z}/n\mathbb{Z}}$ among $\{0, 1\}$ vectors with $\sum x_i = m$. Write $y_i = (x_i - p)/\sqrt{p(1-p)}$ in order to expand into a p -biased basis, and let $y_T = \prod_{i \in T} y_i$. We obtain the expression

$$\text{kAP}'(\mathbf{y}) = \sum_{\ell=3}^k \sum_{a \in \mathbb{Z}/n\mathbb{Z}} \sum_{d \in [n/2]} \sum_{S \in \binom{[k]}{\ell}} p^{k-\frac{|S|}{2}} (1-p)^{\frac{|S|}{2}} \prod_{i \in S} y_{a+id}$$

after removing the linear and quadratic terms, which are deterministic given m . Let σ be the standard deviation of $\text{kAP}'(\mathbf{y})$ if \mathbf{y} were drawn independently (instead of with fixed sum); we will often switch between the independent model and the constrained model, and will note these shifts as they come. Note that $\mathbb{E}[\text{kAP}'(\mathbf{y})] = 0$ in the independent model, and also note $\sigma = \Theta_{k,p}(n)$. We prove that

$$|\mathbb{E}[e^{it \text{kAP}(\mathbf{y})/\sigma}] - e^{-t^2/2}|$$

is small for all $t \in [-\pi\sigma, \pi\sigma]$ in the constrained model. This will later be used to characterize the distribution of $\text{kAP}(\mathbf{x})$ in the independent model, although both results are of interest.

7.1 | Bounds for $|t| \leq n^\epsilon$

To handle these cases, we see that $\text{kAP}'(\mathbf{y})$ is Gaussian in the independent model, and then transfer to the conditioned model.

Define for k -order linear forms the norm

$$\|A\|_{\text{op}} = \sup_{\|v_i\|_2=1} |A(v_1, v_2, \dots, v_k)|.$$

Next define the k th derivative (tensor) operators for $f \in C^k(\mathbb{R}^n)$ as

$$\langle D^k f(x), (u_1, \dots, u_k) \rangle = \sum_{i_1, i_2, \dots, i_k \in [n]} \frac{\partial^k f}{\partial x_{i_1} \dots \partial x_{i_k}}(u_1)_{i_1} \dots (u_k)_{i_k}$$

for vectors $u_1, \dots, u_k \in \mathbb{R}^n$, and let

$$M_r(g) = \sup_{x \in \mathbb{R}^n} \|D^r g(x)\|_{\text{op}}.$$

Finally, define

$$\overline{\text{kAP}}^\ell(\mathbf{y}) = \frac{1}{\sigma_\ell} \sum_{a \in \mathbb{Z}/n\mathbb{Z}} \sum_{d \in [n/2]} \sum_{S \in \binom{[k]}{\ell}} \prod_{i \in S} y_{a+id},$$

where σ_ℓ is chosen, so that $\text{Var}[\overline{\text{kAP}}^\ell(\mathbf{y})] = 1$ where $y_i = (x_i - p)/\sqrt{p(1-p)}$ if $x_i \sim \text{Ber}(p)$. The key technical result of [6] is the following quantitative central limit theorem.

Theorem 7.1. *Let $\gcd(n, (k-1)!) = 1$ and let y_i be defined as before and z'_i be standard normal random variables. Then for any C^3 function $\psi : \mathbb{R}^{k-1} \rightarrow \mathbb{R}$ we have*

$$\left| \mathbb{E}[\psi(z'_1, z'_3, \dots, z'_k)] - \psi(\overline{\text{kAP}}^1(\mathbf{y}), \overline{\text{kAP}}^3(\mathbf{y}), \dots, \overline{\text{kAP}}^k(\mathbf{y})) \right| \lesssim_{k,p} \frac{M_3(\psi) + M_2(\psi)}{n^{1/2}}.$$

Using this, we prove the following lemma.

Lemma 7.2. *For all $\varepsilon > 0$ we have*

$$\left| \mathbb{E}_x \left[e^{it \text{kAP}'(\mathbf{y})/\sigma} \mid \sum y_i = 0 \right] - e^{-t^2/2} \right| \lesssim_{k,\lambda,\varepsilon} t/n^{1/4-\varepsilon} + t^3/n^{1/2} + \exp(-\Omega_{k,\lambda,\varepsilon}((\log n)^2)).$$

Remark. The λ dependence merely reflects that p is bounded away from 0 and 1 in terms of λ , which controls the constants in all inequalities.

Proof. The proof is similar to the proof of Lemma 3.3, except that we have a slightly worse bound in the independent model to start with. The key idea is to bootstrap from Theorem 7.1. We perform the following procedure.

- (1) Sample each element y_i to be $\sqrt{1-p}/\sqrt{p}$ with probability p and $-\sqrt{p}/\sqrt{1-p}$ with probability $1-p$.
- (2) Adjust a random subset of the $\sqrt{1-p}/\sqrt{p}$ to $-\sqrt{p}/\sqrt{1-p}$ or vice versa, so that the sum is as desired. Specifically, determine which direction the adjustment needs to be and uniformly sample the correct amount in the correct direction.

Let y_i be the initial sample and y'_i be the sample of elements adjusted. Let S_0 be a random variable indicating the set of changed random variables and S_1 which direction the change occurred in. We will show that

$$\mathbb{P}[|kAP'(\mathbf{y}) - kAP'(\mathbf{y}')| \geq n^{3/4+\varepsilon}]$$

is super-polynomially small for any $\varepsilon > 0$. First we find $\mathbb{P}[|S_0| \geq n^{1/2} \log n] = \exp(-\Omega_\lambda((\log n)^2))$. Next, note that given S_0, S_1 we can write $kAP'(\mathbf{y}) - kAP'(\mathbf{y}')$ as a bounded degree polynomial in y_j for $j \notin S_0$ since we know exactly what changed and how. More specifically, write it as $\sum_{T \subseteq S_0^c} a_T y_T$. Note a_T are random variables that are functions of S_0, S_1 . Every term y_T in the expansion must correspond to a subset of a k -AP (that is, a term in the original kAP' polynomial) that hits S_0 (so that its coefficient may change). Furthermore, each coefficient in kAP' is bounded in terms of λ, k . Thus,

$$\sum_{T \subseteq S_0^c} |a_T| \lesssim_{k,\lambda} |S_0| n \lesssim_{k,\lambda} n^{3/2} \log n,$$

the latter inequality occurring with high probability over the randomness of S_0, S_1 . Now we claim with high probability that

$$\sum_{T \subseteq S_0^c} a_T^2 \lesssim_{k,\lambda} n^{3/2} (\log n)^2.$$

To do this, we make the following three observations, assuming that $|S_0|$ is small (since it occurs with high probability). Before we dive in, we note that given $|S_0|, S_1$, the set S_0 is distributed uniformly. Also, the contributing terms to a_T are all subsets S of S_0 such that $S \cup T$ is a subset of a k -AP of size at least 3.

- (1) For all $|T| \geq 2$ note that $|a_T| \lesssim_{k,\lambda} 1$ as given two elements in a k -AP there are only $\Theta_k(1)$ ways to extend it. Thus, the L^1 bound above suffices to establish the estimate for these coefficients.
- (2) For $|T| = 1$, say $T = \{b\}$, recall that a_T is a sum over subsets S of S_0 such that $S \cup \{b\}$ is a subset of a k -AP of size at least 3. Given $|S_0|$, we will show this count is small with high probability over the randomness of S_0 . Note that this random variable is monotonic. Additionally, if we consider picking elements of S_0 independently with probability $(\log n)^2 n^{-1/2}$ instead, we see with probability greater than $1/3$, say, that $|S_0|$ is bigger than it needs to be. These facts combined show that it suffices to bound the count of valid S in this new process with high probability.

Now partition the set of possible S into $\Theta_k(1)$ pieces of size $\Theta_k(n)$ such that in each piece the elements do not intersect. (This can be accomplished explicitly, or by noting that the intersection graph has bounded degree, and coloring this graph then rebalancing.) By Chernoff with probability at least $1 - \exp(-(\log n)^4)$ the coefficient is at most $O_k((\log n)^4)$. Square it and multiply by n after a union bound.

- (3) Finally, for the constant coefficient this is bounded (up to factors depending on k and λ) by the number of such subsets of $|T| \geq 3$ in S_0 , so that they are a subset of a k -AP. Now consider the martingale where each element of S_0 is revealed on at a time. Since each element can be in at most $\lesssim_k |S_0|$ such sets we have concentration of order $|S_0|^{3/2}$ by Azuma–Hoeffding and thus this coefficient is bounded by $n^{3/4} \log n$ with high probability as the expectation is $\lesssim n^{3/4}$,

which gives bounds of the claimed quality. (Note that one can by a more subtle argument bound this coefficient by $n^{1/2}(\log n)^C$ but this easier bound is sufficient.)

In summary, we have shown that the above bound on these coefficients occurs with probability $1 - \exp(-\Omega_{k,\lambda}((\log n)^2))$ over the randomness of S_0, S_1 .

Now suppose we are in one of these cases (and also suppose $|S_0| \leq n^{1/2} \log n$, since we can), and fix the values S_0, S_1 . We have that $kAP'(\mathbf{y}) - kAP'(\mathbf{y}')$ is a polynomial of bounded degree in \mathbf{y} , which are now drawn independently with a fixed sum based on S_0 , which is at most $O_\lambda(n^{1/2} \log n)$. Shift back to the Boolean model, so that we have some polynomial in $x \in \{0, 1\}^{S_0}$ where we condition on a sum of size $pn + O(n^{1/2} \log n)$. Consider an independent model of selecting the functions x with the same sum. Its probability q is bounded away from 0,1 in terms of λ . By hypercontractivity (Theorem 2.1), we see that the probability of $|kAP'(\mathbf{y}) - kAP'(\mathbf{y}')| \geq n^{3/4+\varepsilon}$ in the independent model of selecting S_0^c is $\exp(-\Omega_{k,\lambda}(n^{\varepsilon'}))$ where ε' depends only on ε, k . Therefore, in the conditioned sum model of S_0^c , which occurs with probability $\Omega(1/n)$ in the independent model, we still have that this event occurs with small probability (otherwise the independent model would have a bigger probability).

Now we can add back in the cases where S_0, S_1 are not sufficiently nice to merit the above bounds hence the above deduction. Overall, we find (over all the randomness) that

$$\mathbb{P}[|kAP'(\mathbf{y}) - kAP'(\mathbf{y}')| \geq n^{3/4+\varepsilon}] = \exp(-\Omega_{k,\lambda,\varepsilon}((\log n)^2)).$$

Finally, note that $kAP'(\mathbf{y}) = \sum_{\ell=2}^k \sigma_\ell \overline{kAP'}^\ell(\mathbf{y})$ where $\sigma_\ell = \Theta_{k,\ell}(n)$. Thus,

$$\begin{aligned} & |\mathbb{E}[e^{it kAP'(\mathbf{y}')/\sigma} - e^{-t^2/2}]| \\ & \leq \left| \mathbb{E} \left[e^{it \sum_{\ell=3}^k \sigma_\ell \overline{kAP'}^\ell(\mathbf{y}) / \sqrt{\sum_{\ell=3}^k \sigma_\ell^2}} - e^{it \sum_{\ell=3}^k \sigma_\ell \overline{kAP'}^\ell(\mathbf{y}') / \sqrt{\sum_{\ell=3}^k \sigma_\ell^2}} \right] \right| \\ & \quad + \left| \mathbb{E} \left[e^{it \sum_{\ell=3}^k \sigma_\ell \overline{kAP'}^\ell(\mathbf{y}) / \sqrt{\sum_{\ell=3}^k \sigma_\ell^2}} - e^{-t^2/2} \right] \right| \\ & \lesssim_{k,\lambda,\varepsilon} \exp(-\Omega_{k,\lambda,\varepsilon}((\log n)^2)) + t |n^{3/4+\varepsilon}/n| + t^3/n^{1/2} \end{aligned}$$

where in the final inequality we have used the coupling inequality between the two distributions, as well as the fact that $M_3(g)$ for $g(\mathbf{a}) = e^{it(\mathbf{r} \cdot \mathbf{a})}$, $\mathbf{r} \in \mathbb{R}^k$, is $O_{k,\mathbf{r}}(t^3)$. \square

Note that this bound allows us to handle any $|t| \leq n^{1/8-\varepsilon}$ but we will only use the bound up to $|t| \leq n^\varepsilon$ to obtain better bounds.

7.2 | Bounds for $n^\varepsilon \leq |t| \leq n^{1-\varepsilon}$

For this section, relying on decoupling techniques, on a first pass the reader is advised to ignore various technical maneuvers required to deal with the fact that our random variables are constrained to live on a slice of the hypercube.

Fix a real number $\beta \in (0, 1)$ and set $S = \lfloor n^\beta \rfloor$ (we will take care to ensure β is bounded away from the endpoints). Then take $A_j = \lfloor jn/(10k^2) + n/2 \rfloor + 2 \lfloor jS, (j+1)S \rfloor$ for $1 \leq j \leq k-1$. Note

that these are essentially intervals of length n^β that are spaced a constant fraction. It follows easily that any rainbow k -APs containing an element in A_1, A_2, \dots, A_{k-1} (and an element in their complement) are forced to be genuine k -APs in \mathbb{Z} ; this follows trivially noting that all the sets are sufficiently close to the center of $\mathbb{Z}/n\mathbb{Z}$ and the common difference between them is $n/(10k) + \Theta(n^\beta)$. Set $T = (\mathbb{Z}/n\mathbb{Z}) \setminus \bigcup_{i=1}^{k-1} A_i$, and note $|T| = \Theta_k(n)$.

Now consider the following random process: select a uniform random subset of $\mathbb{Z}/n\mathbb{Z}$ of size m , then resample the elements not in T conditional on the outcome within T . Let X be the indicator vector of the subset of T chosen, and Y_j^b for $b \in \{0, 1\}$ be the subset of A_i chosen, as well as the resample. As introduced in the decoupling section, let $\mathbf{Y} = (Y_1^0, Y_1^1, \dots, Y_{k-1}^0, Y_{k-1}^1)$. We can alternatively view this process as sampling Z_i , the number of edges chosen in each A_i if a subset is chosen with m elements uniformly, and then sampling X and two independent copies of Y_i^0, Y_i^1 (conditional on Z_i).

Let x_i for $i \in T$ be 1 or 0 depending on if $i \in X$, and let x_i^b for $i \in A_j$ and $b \in \{0, 1\}$ be 1 or 0 depending on whether $i \in Y_j^b$. Now note that, recalling from Subsection 2.4 that non-rainbow functions are in the kernel of α ,

$$\alpha(\text{kAP})(X, \mathbf{Y}) = \sum_{i \in T} x_i \sum_{\substack{\text{Rainbow } k\text{-AP} \\ \mathcal{A} \text{ including } i}} \prod_{j \in \mathcal{A}} (x_j^1 - x_j^0)$$

hence

$$\alpha(\text{kAP})(X, \mathbf{Y}) = \sum_{i \in T} a_i x_i,$$

where

$$a_i = \sum_{\substack{\text{Rainbow } k\text{-AP} \\ \mathcal{A} \text{ including } i}} \prod_{j \in \mathcal{A}} (x_j^1 - x_j^0).$$

We aim to prove for $n^\varepsilon \leq |t| \leq n^{1-\varepsilon}$ that

$$|\varphi_{\text{kAP}'/\sigma}(t)| \lesssim_\varepsilon \exp(-\Omega_{k,\lambda}(n^{\varepsilon'})),$$

where ε' depends only on k, ε . To begin, by Lemma 2.10 we have

$$|\varphi_{\text{kAP}'/\sigma}(t)|^{2^k} \leq \mathbb{E}_{\mathbf{Y}} |\mathbb{E}_X e^{it\alpha(\text{kAP})(X, \mathbf{Y})/\sigma}|,$$

where the change from kAP' to kAP occurs merely because kAP' , kAP are the same up to a constant in the conditioned model.

Hence, it will suffice to show with high probability over the randomness of \mathbf{Y} that the inner expectation is small. Since given \mathbf{Y} , X is chosen uniformly with some fixed sum, we will be able to apply Lemma 2.7. To do so, we need to know that $\text{Var}[a_i]$ is large. We also need $|a_i|$ to be not too large for all $i \in T$.

To prove this occurs with high probability, it will be more convenient to pretend that \mathbf{Y} is sampled independently with probability p . To transfer from this to the true distribution, we use the following argument (which was also used for the graph statistic results in Subsection 3.4).

Define a *suitable* outcome to be if $|Z_i - p|A_i| \leq \sqrt{|A_i|} \log |A_i|$, say. By Azuma–Hoeffding and union-bounding over the fixed number of sets considered, there is an overwhelming probability that all Z_i are suitable. In particular, probability of failure is $\exp(-\Omega_\lambda((\log n)^2))$.

If we sample the elements of A_i with probability p independently (sampling twice) then we attain any suitable number of elements in all B_i with probability at least $\exp(-\Omega_\lambda((\log n)^2))$. Thus, if in this independent model, an event has probability at most $\exp(-\Omega_\lambda((\log n)^3))$, then even in the conditioned size model within suitable outcomes it occurs with this probability, perhaps weakening the constants in the exponent (we used a similar trick in the small $|t|$ regime as well.) Then we must add back in the unsuitable cases, which account for a probability of at most $\exp(-\Omega_\lambda((\log n)^2))$ as noted above.

We now proceed to prove the desired control (with high probability) on $\text{Var}[a_i]$ in the independent model of selecting the Y_j^b , which as shown above will transfer to the desired model. We do so by proving a number of bounds on these coefficients. Note that the a_i are polynomials in the y_j^b , which are now being selected independently with probability p . It is worth noting that a_i is a nonzero polynomial only for $\Theta_k(n^\beta)$ values $i \in T$, as the rainbow k -APs can only include within T elements from two regions of prescribed width. To be more precise we prove the following lemma.

Lemma 7.3. *Let X, \mathbf{Y}, a_i, T be as above. Let C be a suitably large constant. Then have the following concentration bounds in the model where each element is sampled with probability p .*

(1) *We have that*

$$\mathbb{P} \left[\sup_{i \in T} |a_i| \geq n^{\beta/2} (\log n)^C \right] \leq \exp(-\Omega_\lambda((\log n)^3)).$$

(2) *We have that*

$$\mathbb{P} \left[\left| \sum_{i \in T} a_i \right| \leq n^\beta (\log n)^C \right] \leq \exp(-\Omega_\lambda((\log n)^3)).$$

(3) *We have that*

$$\mathbb{E} \left[\sum_{i \in T} a_i^2 \right] = \Theta_\lambda(n^{2\beta})$$

and

$$\text{Var} \left[\sum_{i \in T} a_i^2 \right] = O_\lambda(n^{3\beta/2} (\log n)^{2C}).$$

7.2.1 | Proof of Lemma 7.3(1)

We show $|a_i| \lesssim_k n^{\beta/2} (\log n)^C$ with high probability. Note that a_i is composed of $O_k(n^\beta)$ terms with constant coefficients. Therefore, the sum of squares of its coefficients is $O_k(n^\beta)$, and hypercontractivity (Theorem 2.1) immediately gives the desired result for C chosen large enough depending on

k (which bounds the degree of the polynomial considered). Recall that in fact $a_i = 0$ except for $\Theta_k(n^\beta)$ values in admissible ‘target’ regions to the left and right of the A_i .

7.2.2 | Proof of Lemma 7.3(2)

We show $\sum_{i \in T} a_i \lesssim_k n^\beta (\log n)^C$ with high probability. Note that a_i is composed of $O_k(n^\beta)$ terms with constant coefficients, and each term extends to be included in at most 2 polynomials a_i . Therefore, the sum of squares of coefficients of the total polynomial is $O_k(n^{2\beta})$ because of our bound on nonzero polynomials. By hypercontractivity (Theorem 2.1) the result follows, with C chosen large enough depending on k (which bounds the degree of the polynomials considered) to obtain a good enough concentration.

7.2.3 | Proof of Lemma 7.3(3)

We show $\sum_{i \in T} a_i^2$ concentrates on a value of size $\Theta_{k,\lambda}(n^{2\beta})$. Note that $\mathbb{E}[a_i^2] = \Theta_{k,\lambda}(n^\beta)$ for most $i \in T$ with nonzero a_i since the only nonzero terms come from choosing the same rainbow k -AP twice, which yields $\Theta_k(n^\beta)$ possibilities for i near the middle of the two nonzero target regions. Near the fringes, there could be less, but this does not affect the order of magnitude given.

Now it suffices to show that the standard deviation of $\sum_{i \in T} a_i^2$ (over the randomness of the independently chosen x_j^b) is of smaller order by a power of n ; hypercontractivity (Theorem 2.1) will then give the desired concentration. The desired variance is

$$\sum_{i,j \in T} (\mathbb{E}[a_i^2 a_j^2] - \mathbb{E}[a_i^2] \mathbb{E}[a_j^2]).$$

For terms $i = j$, there are $O_k(n^\beta)$ of them with nonzero a_i . By the L^∞ bound above, with high probability $|a_i| \lesssim_k n^{\beta/2} (\log n)^C$, hence the expectation terms are bounded by $n^{2\beta} (\log n)^{4C}$ for a total of $O_{k,\lambda}(n^{3\beta} (\log n)^{4C})$, which is acceptable.

For terms $i \neq j$, expand into further covariance terms coming from considering two rainbow k -APs $\mathcal{A}_1, \mathcal{A}_2$ including i and $\mathcal{B}_1, \mathcal{B}_2$ including j :

$$\mathbb{E}[a_i^2 a_j^2] - \mathbb{E}[a_i^2] \mathbb{E}[a_j^2] = \sum_{\substack{\mathcal{A}_1, \mathcal{A}_2 \text{ incl. } i \\ \mathcal{B}_1, \mathcal{B}_2 \text{ incl. } j}}^* \prod_{b \in \{1,2\}} \left[\prod_{j \in \mathcal{A}_b} (x_j^1 - x_j^0) \prod_{j \in \mathcal{B}_b} (x_j^1 - x_j^0) \right],$$

where \sum^* denotes a sum over all the k -APs being rainbow.

The covariance term coming from these vanishes unless every element in the union of the four progressions is covered at least twice. If $\mathcal{A}_1 = \mathcal{A}_2$ we see that either there are $O_k(1)$ choices for $\mathcal{B}_1, \mathcal{B}_2$ (so they both hit \mathcal{A}_1) otherwise they must be equal and disjoint from \mathcal{A}_1 . In the latter case, we have independence so the term is zero. In the former case, we have a term of size $O_{k,\lambda}(n^\beta)$ as there are $O_k(n^\beta)$ choices for \mathcal{A}_1 . Summing over all i, j we obtain a contribution of $O_{k,\lambda}(n^{3\beta})$, which is acceptable. A similar analysis holds if $\mathcal{B}_1 = \mathcal{B}_2$.

If both pairs are unequal, we see that $\mathcal{A}_1, \mathcal{A}_2$ are disjoint and the same for the others. Hence, $\mathcal{A}_1 \cup \mathcal{A}_2$ spans $2k - 2$ disjoint vertices and the same for $\mathcal{B}_1 \cup \mathcal{B}_2$, hence these unions must be

identical else the term is zero. After choosing $i, \mathcal{A}_1, \mathcal{A}_2$ we see there are $O_k(1)$ choices for $\mathcal{B}_1, \mathcal{B}_2$ and j . This gives a contribution of $O_{k,\lambda}(n^{3\beta})$ again.

Overall, we have shown that the variance is $O_{k,\lambda}(n^{3\beta}(\log n)^{4C})$ hence the standard deviation is $O_{k,\lambda}(n^{3\beta/2}(\log n)^{2C})$, which is indeed much smaller than $n^{2\beta}$, so we are done.

This (finally) concludes the proof of Lemma 3.5. Now we use these bounds to conclude our argument in the intermediate range of $|t|$.

7.2.4 | Putting it together

Lemma 7.4. *For $\varepsilon > 0$ and $|t| \in [n^{2\varepsilon}, \sigma n^{-4\varepsilon}]$ we have*

$$\left| \mathbb{E}_X \left[e^{it \text{kAP}'(\mathbf{y})/\sigma} \mid \sum y_i = 0 \right] - e^{-t^2/2} \right| \lesssim \exp(-\Omega_{k,\lambda}((\log n)^2)) + \exp(-\Omega_{k,\lambda}(n^\varepsilon)).$$

Proof. Note that $e^{-t^2/2}$ is sufficiently small in this range that it can be ignored. Let X, \mathbf{Y} be as at the beginning of Subsection 7.2. By Lemma 2.10, we have

$$\left| \mathbb{E}_X \left[e^{it \text{kAP}'(\mathbf{y})/\sigma} \mid \sum y_i = 0 \right] \right|^{2^{k-1}} \leq \left| \mathbb{E}_X e^{it\alpha(\text{kAP}')(\mathbf{X}, \mathbf{Y})/\sigma} \right|.$$

Now, with high probability over \mathbf{Y} (in the conditioned model) we have the bounds given in Lemma 7.3. In particular, the probability of failure is $\exp(-\Omega_{k,\lambda}((\log n)^2))$. In particular, the average of a_i^2 is $n^{2\beta-1}$ in magnitude (since most of these values are zeros when β is small, this could be less than 1) while the average of a_i is $n^{\beta-1}(\log n)^C$ in magnitude. The square of the latter is much smaller than the former, hence we see that $\text{Var}[a_i]$ is of order $\Theta_{k,\lambda}(n^{2\beta-1})$. Also, $|a_i| \leq n^{\beta/2}(\log n)^C$ for all $i \in T$. This allows us to apply Lemma 2.7 to deduce in these cases that

$$\left| \mathbb{E}_X e^{it\alpha(\text{kAP}')(\mathbf{X}, \mathbf{Y})/\sigma} \right| \lesssim \exp(-\Omega_{k,\lambda}((t^2 n / \sigma^2) \text{Var}[a_i])) \lesssim \exp(-\Omega_{k,\lambda}(t^2 n^{2\beta-2}))$$

as long as $|t/\sigma| n^{\beta/2}(\log n)^C \lesssim_\lambda 1$. Therefore, the bound is good as long as $t \in [n^{1-\beta+\varepsilon}, n^{1-\beta/2-\varepsilon}]$, say. Now varying β between 5ε and $1 - \varepsilon$ (taking care to keep it bounded away from the endpoints) we see this covers the range $[n^{2\varepsilon}, n^{1-4\varepsilon}]$, which is good enough for our purposes. \square

7.3 | Bounds for $n^{1-\varepsilon} \leq |t| \leq \pi\sigma$

For this section, we develop a series of sets upon which the decoupling estimates will be formed. They stem from a tensor product construction which we outline below. The key difficulty is in ensuring that the decoupled expression has many coefficients that are ± 1 , so that we can apply Lemma 2.7 up to the values $t = \pm\pi\sigma$.

Lemma 7.5. *Suppose $k \geq 4$. Let $A_i = \{k + 4i, k + 4i + 2\}$ for $1 \leq i \leq k - 1$. Then the only k -term arithmetic progressions with 1 element in each of the A_i are $\{k + 4i\}$ for $0 \leq i \leq k - 1$, $\{k + 4i\}$ for $1 \leq i \leq k$, $\{k + 4i + 2\}$ for $0 \leq i \leq k - 1$, and $\{k + 4i + 2\}$ for $1 \leq i \leq k$. Note that arithmetic progressions here means an arithmetic progression in \mathbb{Z}*

Proof. Since the arithmetic progression in \mathbb{Z} contains elements in A_1 and A_2 we must have that the common difference is in the set 2,4,6. Using this the result follows as extending the progression from A_1, A_2 to A_3 forces the common difference to be 4. \square

For $k = 3$ let $A_1 = \{16, 18\}$ and $A_2 = \{22, 32\}$. Then the 3-APs which contain exactly one element in A_1 and A_2 are $\{0, 16, 32\}, \{10, 16, 22\}, \{16, 19, 22\}, \{16, 22, 28\}, \{18, 22, 26\}, \{4, 18, 32\}, \{14, 18, 22\}, \{16, 24, 32\}, \{18, 20, 22\}, \{18, 25, 32\}, \{18, 32, 46\}$, and $\{16, 32, 48\}$. Finally, for $k = 3$ define

$$A_0 = \{0, 4, 10, 14, 19, 24, 20, 25, 26, 28, 46, 48\}$$

and for $k \geq 4$ define

$$A_0 = \{k, k + 2, 5k, 5k + 2\},$$

which are precisely the elements to which these k -APs extend. Note that all the extensions are to distinct integers.

Lemma 7.6. *Let n be sufficiently large in terms of k . Embed B_1, \dots, B_{k-1} into $\mathbb{Z}/\ell\mathbb{Z}$ where $\ell = 100k^3$ where $B_i = 10k^2 + A_i$, that is, each element is shifted by a constant. Then the only k -terms APs with exactly 1 element in each B_i are as in the previous lemma.*

Proof. For $k = 3$ the proposition is trivial. For $k \geq 4$ note that two consecutive terms of the k -AP lie in a pair of B_i, B_j and thus the common difference is bounded by $5k$ and therefore the AP cannot wrap around the edges. The result then follows by the previous lemma. \square

Finally, define C_i^u for $0 \leq i \leq k - 1$ to be the set of u -digit numbers in base $\ell = 100k^3$ with digits only in B_i . Choose ℓ , so that $\ell^u = \Theta_k(n^{1/2})$ and thus $|C_0^u| = \Theta_k(n^{\delta_k})$ for some constant $\delta_k > 0$. Now embed C_i^u into $\mathbb{Z}/n\mathbb{Z}$ as the sets D_i^u by adding $\lfloor n/2 \rfloor$ to all the elements. The key fact which follows from the previous lemmas is that the all k -APs with 1 element in D_i^u for $1 \leq i \leq k - 1$ complete into D_0^u . Indeed, we see that the common difference is $\Theta_k(n^{1/2})$ hence again there is no wraparound as the numbers are all near $n/2$.

We now use our decoupling lemma. Let X, Y_j^0 for $1 \leq j \leq k - 1$ be a fixed-sum sample of $m = pn$ elements from $\mathbb{Z}/n\mathbb{Z}$ where Y_j^0 is the subset of D_j^u for $j \geq 1$ chosen and X is the subset of $(\mathbb{Z}/n\mathbb{Z})/\cup_{j=1}^{k-1} D_j^u$ chosen. Then resample Y_j^1 for $1 \leq j \leq k - 1$ having the same sum as Y_j^0 . Note that

$$\begin{aligned} \alpha(\text{kAP}')(X, \mathbf{Y})(x) &= \sum_{i \in D_0^u} x_i \prod_{\substack{j \in \text{Unique rainbow} \\ k\text{-AP including } i}} (y_j^1 - y_j^0) \\ &= \sum_{i \in D_0^u} a_i x_i, \end{aligned}$$

using that each i is in a rainbow k -AP by construction only if it is in D_0^u , in which case it is in a unique such progression.

Lemma 7.7. *Let $k\text{AP}$, $k\text{AP}'$, σ be as above. Then for $|t| \leq \pi\sigma$,*

$$|\varphi_{k\text{AP}'/\sigma}(t)| \leq \exp(-\Omega_{k,\lambda}(n^{\delta_k})) + \exp(-\Omega_{k,\lambda}(t^2 n^{\delta_k}/\sigma^2)). \quad (7.1)$$

Proof. Sample X, \mathbf{Y} as above. The key claim is that with high probability over the randomness of \mathbf{Y} , we have a positive fraction of the coefficients a_i for $i \in D_0^u$ are 0, 1, and -1 . (Note that these are the only possible coefficients as each rainbow AP includes a unique element in D_0^u .)

We first consider the independent model of sampling. In it, we see there is a $\Theta_{k,\lambda}(1)$ probability of obtaining each of $\{0, \pm 1\}$ as a coefficient for each $i \in D_0^u$. By Azuma–Hoeffding, this translates to a probability of $1 - \exp(-\Omega_{k,\lambda}(n^{\delta_k}))$ that each value occurs in a $\Theta_{k,\lambda}(1)$ fraction of the a_i for $i \in D_0^u$. Now, if instead we sample $X, Y_1^0, \dots, Y_{k-1}^0$ with constrained sum, and then resample Y_1^1, \dots, Y_{k-1}^1 , similar bounds hold by repeatedly applying Azuma–Hoeffding: first, the number of elements X, \mathbf{Y} are concentrated near a p fraction, and then the coefficients a_i are concentrated near some positive fraction. (See Lemmas 4.1 and 4.3 for similar arguments in the graph statistic setting.)

Now we apply Lemma 2.10. We again have

$$|\varphi_{k\text{AP}'/\sigma}(t)|^{2^{k-1}} \leq \mathbb{E}_{\mathbf{Y}} |\mathbb{E}_X e^{it\alpha(k\text{AP})(X, \mathbf{Y})/\sigma}|,$$

and now let E be the subset of D_0^u with coefficient in $\{0, 1\}$, with corresponding vector X' (this only depends on the randomness of \mathbf{Y}). Then we have

$$|\varphi_{k\text{AP}'/\sigma}(t)|^{2^{k-1}} \leq \mathbb{E}_{\mathbf{Y}, X \setminus X'} \left| \mathbb{E}_{X'} e^{(it/\sigma) \sum_{j \in E} a_j x_j} \right|.$$

By the above considerations, with high probability over the randomness of \mathbf{Y} we have a positive proportion of E being 0, 1, and $|E| = \Theta_{k,\lambda}(n^{\delta_k})$. Then $\text{Var}[a_j]$ over $j \in E$ is $\Theta_{k,\lambda}(1)$, hence we deduce by Lemma 2.7 that

$$|\varphi_{k\text{AP}'/\sigma}(t)|^{2^{k-1}} \leq \exp(-\Omega_{k,\lambda}(n^{\delta_k})) + \exp(-\Omega_{k,\lambda}(t^2 n^{\delta_k}/\sigma^2)).$$

This also only applies if $|t/\sigma| \cdot (1 - 0) \leq \pi$, which precisely hits the top of the range. \square

Therefore, for t in the given range, we deduce the desired quality of bounds.

7.4 | Deriving the final result

We are ready to prove Theorem 1.4, and then transfer the result to the independent setting.

Proof of Theorem 1.4. Recall from earlier that

$$k\text{AP}(\mathbf{y}) = \sum_{\ell=0}^k \sum_{a \in \mathbb{Z}/n\mathbb{Z}} \sum_{d \in [n/2]} \sum_{S \in \binom{[k]}{\ell}} p^{k-\frac{\ell}{2}} (1-p)^{\frac{\ell}{2}} \prod_{i \in S} y_{a+id},$$

which differs from $kAP'(\mathbf{y})$ differ by a deterministic constant given the sum of y , call it Y , and thus our various decoupling estimates apply. To be more explicit, let

$$\begin{aligned}\mu = kAP(\mathbf{y}) - kAP'(\mathbf{y}) &= p^k \binom{n}{2} + p^{k-\frac{1}{2}}(1-p)^{\frac{1}{2}} k \frac{n-1}{2} Y \\ &\quad + \frac{p^{k-1}(1-p)}{2} \binom{k}{2} \left(Y^2 + \frac{2p-1}{\sqrt{p(1-p)}} Y - n \right) \\ &= p^k \binom{n}{2} - np^{k-1}(1-p) \binom{k}{2} / 2,\end{aligned}$$

the last expression coming from the facts

$$\sum_{i \neq j} y_i y_j = Y^2 - \sum_i y_i^2, \quad y_i^2 + \frac{2p-1}{\sqrt{p(1-p)}} y_i - 1 = 0.$$

Let $Z' = (kAP(\mathbf{y}) - \mu)/\sigma = kAP'(\mathbf{y})/\sigma$ and set $\varphi_n(t) = \mathbb{E}[e^{itZ'}]$ and $\varphi(t) = \mathbb{E}[e^{itZ}]$ where $Z \sim \mathcal{N}(0, 1)$. Now note that

$$\begin{aligned}& \left(\int_{-\pi\sigma}^{\pi\sigma} |\varphi(t) - \varphi_n(t)| dt \right) / \sigma \\ &= \left(\int_{|t| \leq n^\varepsilon} |\varphi(t) - \varphi_n(t)| dt + \int_{n^\varepsilon \leq |t| \leq \sigma \cdot n^{-\varepsilon}} |\varphi(t) - \varphi_n(t)| dt \right. \\ &\quad \left. + \int_{\sigma \cdot n^{-\varepsilon} \leq |t| \leq \pi \cdot \sigma} |\varphi(t) - \varphi_n(t)| dt \right) / \sigma \\ &\lesssim_\lambda \left(\int_{|t| \leq n^\varepsilon} |t| / n^{1/4-\varepsilon} dt + \int_{n^\varepsilon \leq |t| \leq \sigma \cdot n^{-\varepsilon}} \exp(-\Omega_{k,\lambda}(n^{\varepsilon'})) dt \right. \\ &\quad \left. + \int_{\sigma \cdot n^{-\varepsilon} \leq |t| \leq \pi \cdot \sigma} \exp(-\Omega_{k,\lambda}(n^{\delta_k})) + \exp(-\Omega_{k,\lambda}(t^2 n^{\delta_k} / \sigma^2)) dt \right) / \sigma \\ &\lesssim_\lambda 1/(\sigma \cdot n^{1/4-3\varepsilon}).\end{aligned}$$

The bounds applied were derived in the previous subsections. Given this we are almost able to derive the necessary result; however once again we have the issue that μ, σ are not exactly the true mean or standard deviation μ_k, σ_k of kAP .

Using techniques completely analogous to Lemma 3.4, we can use the coupling in the proof of Lemma 7.2 to see that $\sigma_k = \sigma(1 + O_{\lambda,\varepsilon}(n^{\varepsilon-1/4}))$, and we also have $\mu_k = \mu(1 + O_{\lambda,\varepsilon}(n^{-2}))$ via explicit calculation (similar to in Lemma 3.4). This finishes. \square

A similar analysis to the transfer given in Section 5 allows us to obtain a local limit theorem for k -APs in the independent model. However, the local behavior is not Gaussian but rather comes from the superimposition of an infinite ensemble of Gaussians. For the following theorem note

that we have, in the independent model,

$$(\text{Var}[\text{kAP}'(\mathbf{y}) | \text{Ber}(p)])^{1/2} = c_{k,p} n(1 + O_{k,p}(1/n))$$

and

$$(\text{Var}[\text{kAP}(\mathbf{y}) | \text{Ber}(p)])^{1/2} = p^{k-1/2}(1-p)^{1/2} k n^{3/2} / 2(1 + O_{k,p}(1/n))$$

for some constant $c_{k,p}$ depending only on k, p . It is worth noting that $c_{k,p}$ can be seen to be continuous on $p \in [\lambda, 1 - \lambda]$. Finally, define $C_{k,p}$ via

$$\frac{\sigma_k}{\sigma \sqrt{n} \sqrt{p(1-p)}} = \frac{1}{C_{k,p}} (1 + O_{k,p}(1/n)).$$

Theorem 7.8. Fix $k \geq 3$ and let $p \in (\lambda, 1 - \lambda)$. Choose $n \geq 1$ with $\gcd(n, (k-1)!) = 1$ and sample a random set with indicator vector \mathbf{x} , with each element drawn independently with probability p . Furthermore, let $\mu_k = \mathbb{E}[\text{kAP}(\mathbf{x})]$ and $\sigma_k = \text{Var}[\text{kAP}(\mathbf{x})]$. Finally, define $Z_k = (\text{kAP}(\mathbf{x}) - \mu_k) / \sigma_k$ and $\mathcal{N}(z) = e^{-z^2/2} / \sqrt{2\pi}$. Then we have for any $\varepsilon > 0$ that

$$\sup_{z \in (\mathbb{Z} - \mu_k) / \sigma_k} \left| \sigma_k \mathbb{P}[Z_k = z] - \mathcal{N}(z) \sum_{m \in \mathbb{Z}} \frac{1}{C_{k,p}} \mathcal{N}\left(\frac{pn + \sqrt{p(1-p)}z\sqrt{n} + \frac{(1-p)(k-1)(1-z^2)}{2} - m}{C_{k,p}} \right) \right| \\ \lesssim_{\lambda, \varepsilon} n^{\varepsilon-1/4},$$

where $C_{k,p}$ is defined as above.

Proof. Let kAP denote the number of k -term arithmetic progressions. In what follows we suppress k dependence in asymptotic notation. Let $\sigma_k^2 = \text{Var}[\text{kAP} | \text{Ber}(p)]$ and $\mu_k = \mathbb{E}[\text{kAP} | \text{Ber}(p)]$ define the standard deviation and mean in the independent model. We now recall from the proof of Theorem 1.4 that if $Y = \sum y_i$ then

$$\text{kAP}(\mathbf{y}) = p^k \binom{n}{2} + p^{k-\frac{1}{2}}(1-p)^{\frac{1}{2}} k \frac{n-1}{2} Y + p^{k-1}(1-p) \binom{k}{2} \frac{Y^2 + \frac{2p-1}{\sqrt{p(1-p)}} Y - n}{2} + \text{kAP}'(\mathbf{y}).$$

For the sake of simplicity, let

$$f(Y) = p^k \binom{n}{2} + p^{k-\frac{1}{2}}(1-p)^{\frac{1}{2}} k \frac{n-1}{2} Y + p^{k-1}(1-p) \binom{k}{2} \frac{Y^2 + \frac{2p-1}{\sqrt{p(1-p)}} Y - n}{2}.$$

First note that continuity of $c_{k,p}$ implies that

$$\text{Var}[\text{kAP}'(\mathbf{x}) | \text{Ber}(p)] = (1 + O_\lambda(\log n / n^{1/2})) \text{Var}[\text{kAP}'(\mathbf{x}) | \text{Ber}(p')]$$

if $p' = (1 + O_\lambda(\log n/n^{1/2}))p$. Given this and the deductions at the end of the proof of Theorem 1.4 that

$$\text{Var}[\text{kAP}(\mathbf{x}) | \sum x_i = m] = \text{Var}[\text{kAP}'(\mathbf{x}) | \sum x_i = m] = (1 + O_{\lambda,\varepsilon}(n^{\varepsilon-1/4})) \text{Var}[\text{kAP}' | \text{Ber}(q)]$$

(here $q = m/n \in (\lambda, 1 - \lambda)$), it follows for any $m, m' \in [pn - n^{1/2} \log n, pn + n^{1/2} \log n]$ that

$$\begin{aligned} \text{Var}[\text{kAP}(\mathbf{x}) | \sum x_i = m] &= (1 + O_{\lambda,\varepsilon}(n^{\varepsilon-1/4})) \text{Var}[\text{kAP}(\mathbf{x}) | \sum x_i = m'] \\ &= (1 + O_{\lambda,\varepsilon}(n^{\varepsilon-1/4})) \text{Var}[\text{kAP}'(\mathbf{x}) | \text{Ber}(p)]. \end{aligned}$$

From now we denote $\sigma^2 = \text{Var}[\text{kAP}' | \text{Ber}(p)]$, so $\sigma = \Theta_\lambda(n)$. We now explicitly use that the expectation of kAP varies with the function $f(y)$ to deduce our local limit theorem. In particular, consider $m \in [pn - n^{1/2} \log n, pn + n^{1/2} \log n]$ and let $y = (m - pn)/\sqrt{p(1-p)}$. Note $y = O_\lambda(n^{1/2} \log n)$. Then by linearity of expectation we have that

$$\begin{aligned} \mathbb{E}[\text{kAP}(\mathbf{x}) | \sum x_i = m] &= \binom{n}{2} \prod_{i=0}^{k-1} \left(\frac{m-i}{n-i} \right) \\ &= \left(\frac{m}{n} \right)^k \binom{n}{2} - \frac{n \left(\frac{m}{n} \right)^{k-1} \left(1 - \frac{m}{n} \right) \binom{k}{2}}{2} + O_\lambda(1) \\ &= f(y) + O_\lambda(n^{1/2} \log n). \end{aligned}$$

We are now in position to explicitly calculate the distribution of kAP under the independent model. Let $\sigma_m^2 = \text{Var}[\text{kAP}(\mathbf{x}) | \sum x_i = m]$ and $\mu_m = \mathbb{E}[\text{kAP}(\mathbf{x}) | \sum x_i = m]$. Now note that

$$\begin{aligned} \mathbb{P}[\text{kAP}(\mathbf{x}) = x] &= \sum_{m \in \mathbb{Z}} \mathbb{P}[\text{kAP}(\mathbf{x}) = x | \sum x_i = m] \mathbb{P}[\sum x_i = m] \\ &= \sum_{\substack{m \in [pn - n^{1/2} \log n, \\ pn + n^{1/2} \log n]}} \mathbb{P}[\text{kAP}(\mathbf{x}) = x | \sum x_i = m] \mathbb{P}[\sum x_i = m] + \exp(-\Omega_\lambda((\log n)^2)), \end{aligned}$$

where we have used Chernoff to bound the probability that number of elements deviates too far from the mean. For the sake of clarity we will implicitly assume that x is within $\sigma_k(\log n)^C$ of the mean; for x outside this range and C sufficiently large the probability of attaining x is super-polynomially small by hypercontractivity (Theorem 2.1) so the desired statement is trivial. This assumption will be used implicitly later on. Now let \mathcal{M}_x denote the set of m such that

$$|x - f(y)| \leq \sigma(\log n)^C$$

and

$$m \in [pn - n^{1/2} \log n, pn + n^{1/2} \log n]$$

for a suitably large C . (As before, $y = (m - pn)/\sqrt{p(1-p)}$.) Now suppose that $m \in [pn - n^{1/2} \log n, pn + n^{1/2} \log n] \setminus \mathcal{M}_x$. Then

$$\begin{aligned} \mathbb{P}[\text{kAP}(\mathbf{x}) = x | \sum x_i = m] &\leq \mathbb{P}[|\text{kAP}'(\mathbf{x})| \geq \sigma(\log n)^C / 2] / \mathbb{P}\left[\sum x_i = m\right] \\ &\lesssim \exp(-\Omega_\lambda((\log n)^2)), \end{aligned}$$

using that $\mathbb{P}[\sum x_e = m] \gtrsim \exp(-O_\lambda((\log n)^2))$ and then choosing C sufficiently large, so that the bound coming from hypercontractivity (Theorem 2.1) on the numerator is sufficiently strong. The key point is that since $f'(y)$ is a linear function with slope $p^{k-1/2}(1-p)^{1/2}k(n-1)/2 + O_\lambda(n^{1/2} \log n)$ for $|y| \lesssim_\lambda n^{1/2} \log n$, we deduce $|\mathcal{M}_x| = \Theta_\lambda((\log n)^C)$. Thus, we have that

$$\begin{aligned} \mathbb{P}[\text{kAP}(x) = x] &= \sum_{\substack{m \in [pn - n^{1/2} \log n, \\ pn + n^{1/2} \log n]}} \mathbb{P}\left[\text{kAP}(x) = x | \sum x_i = m\right] \mathbb{P}\left[\sum x_i = m\right] \\ &\quad + \exp(-\Omega_\lambda((\log n)^2)) \\ &= \sum_{m \in \mathcal{M}_x} \mathbb{P}\left[\text{kAP}(x) = x | \sum x_i = m\right] \mathbb{P}\left[\sum x_i = m\right] + \exp(-\Omega_\lambda((\log n)^2)). \end{aligned}$$

Now using Theorem 1.4 and that σ_m is approximately equal to σ , the last summation equals

$$\begin{aligned} &\sum_{m \in \mathcal{M}_x} \left(\frac{1}{\sigma_m} \mathcal{N}\left(\frac{x - \mu_m}{\sigma_m}\right) + O_\lambda\left(\frac{n^{\varepsilon-1/4}}{\sigma}\right) \right) \mathbb{P}\left[\sum x_i = m\right] \\ &= \sum_{m \in \mathcal{M}_x} \frac{1}{\sigma_m} \mathcal{N}\left(\frac{x - \mu_m}{\sigma_m}\right) \mathbb{P}\left[\sum x_e = m\right] + O_\lambda\left(\frac{|\mathcal{M}_x|}{n^{\frac{3-\varepsilon}{4}} \sigma}\right), \end{aligned}$$

where we use that probability of having a given number of elements is $O_\lambda(1/n^{1/2})$. Now note that σ_k is order $n^{1/2}$ larger than σ . Therefore, the error term can be seen to be $O_\lambda(n^{\varepsilon-1/4} \sigma_H^{-1})$, which is the correct magnitude. Now $\sigma_m = (1 + O_\lambda(n^{\varepsilon-1/4}))\sigma$ and $\mu_m = f(y) + O_\lambda((\log n)^2 n^{-1/2} \sigma)$ for all $m \in [pn - n^{1/2} \log n, pn + n^{1/2} \log n]$ by the remarks which began the proof. It follows that

$$\begin{aligned} &\sum_{m \in \mathcal{M}_x} \frac{1}{\sigma_m} \mathcal{N}\left(\frac{x - \mu_m}{\sigma_m}\right) \mathbb{P}\left[\sum x_e = m\right] \\ &= \sum_{m \in \mathcal{M}_x} \frac{1}{\sigma} \mathcal{N}\left(\frac{x - f(y)}{\sigma}\right) \mathbb{P}\left[\sum x_e = m\right] + O_\lambda\left(\frac{n^{\varepsilon-1/4}}{\sigma_k}\right). \end{aligned}$$

At this point the rest is elementary, but nontrivial, calculation. Let y^* be the solution to $f(y^*) = x$ and let $m^* = y^* \sqrt{p(1-p)} + pn$ be the corresponding m . Note that $|m - m^*| \lesssim_\lambda (\log n)^C$ since f has slope $\Theta_\lambda(\sigma)$ on \mathcal{M}_x . This is enough to conclude that $\mathbb{P}[\sum x_e = m]$ is essentially constant

over $m \in \mathcal{M}_x$, close enough to replace the above with

$$\mathbb{P}\left[\sum x_e = \lfloor m^* \rfloor\right] \sum_{m \in \mathcal{M}_x} \frac{1}{\sigma} \mathcal{N}\left(\frac{f(y^*) - f(y)}{\sigma}\right)$$

without increasing the error term. Now, using that $f(y)$ has derivative $p^{k-1/2}(1-p)^{1/2}k(n-1)/2 + O_\lambda(n^{1/2} \log n)$ for $|y| \lesssim_\lambda n^{1/2} \log n$, we can (up to acceptable errors) rewrite the above as

$$\mathbb{P}\left[\sum x_e = \lfloor m^* \rfloor\right] \sum_{m \in \mathcal{M}_x} \frac{1}{\sigma} \mathcal{N}\left(\frac{(y^* - y)(p^{k-1/2}(1-p)^{1/2}k(n-1)/2)}{\sigma}\right).$$

Using the value of σ_k and completing the above sum we find that it is close to

$$\mathbb{P}\left[\sum x_e = \lfloor m^* \rfloor\right] \sum_{y \in (\mathbb{Z} - pn)/\sqrt{p(1-p)}} \frac{1}{\sigma} \mathcal{N}\left(\frac{(y^* - y)\sigma_k}{\sqrt{n} \cdot \sigma}\right).$$

Note that $m^* - m = (y^* - y)\sqrt{p(1-p)}$. Now we compute y^* up to a $o(1)$ additive accuracy. Letting $z = (x - \mu_k)/\sigma_k$, and using $|z| \leq (\log n)^C$ we find that

$$\begin{aligned} y^* &= z\sqrt{n} + \frac{(1-p)^{1/2}(k-1)}{2p^{1/2}} - \frac{(1-p)^{1/2}(k-1)z^2}{2p^{1/2}} + O_{\lambda,\varepsilon}(n^{\varepsilon-1/2}). \\ &= z\sqrt{n} + O_\lambda((\log n)^{2C}). \end{aligned}$$

Substituting in this expression we find that the above, up to tolerable losses, is

$$\frac{1}{\sqrt{p(1-p)n}} \mathcal{N}(z) \sum_{y \in (\mathbb{Z} - pn)/\sqrt{p(1-p)}} \frac{1}{\sigma} \mathcal{N}\left(\frac{(y^* - y)\sigma_k}{\sqrt{n} \cdot \sigma}\right),$$

which up to appropriate errors is

$$\frac{1}{\sigma_k} \mathcal{N}(z) \sum_{m \in \mathbb{Z}} \frac{1}{C_{k,p}} \mathcal{N}\left(\frac{pn + \sqrt{p(1-p)}z\sqrt{n} + \frac{(1-p)(k-1)(1-z^2)}{2} - m}{C_{k,p}}\right).$$

The result follows. □

This allows us to answer a question of the authors and Berkowitz [6, Question 16] regarding the maximum ratio between pointwise probabilities near the mean. Indeed, Theorem 7.8 precisely pins down these probabilities to what was expected given the heuristics in [6]. The answer ultimately is the (predicted) ratio of two infinite sums as given above; explicitly, if

$$g(x) = \sum_{m \in \mathbb{Z}} \frac{1}{C_{k,p}} \mathcal{N}\left(\frac{x - m}{C_{k,p}}\right),$$

the maximum ratio of pointwise probabilities near the mean is $\sup g(x)/\inf g(x)$.

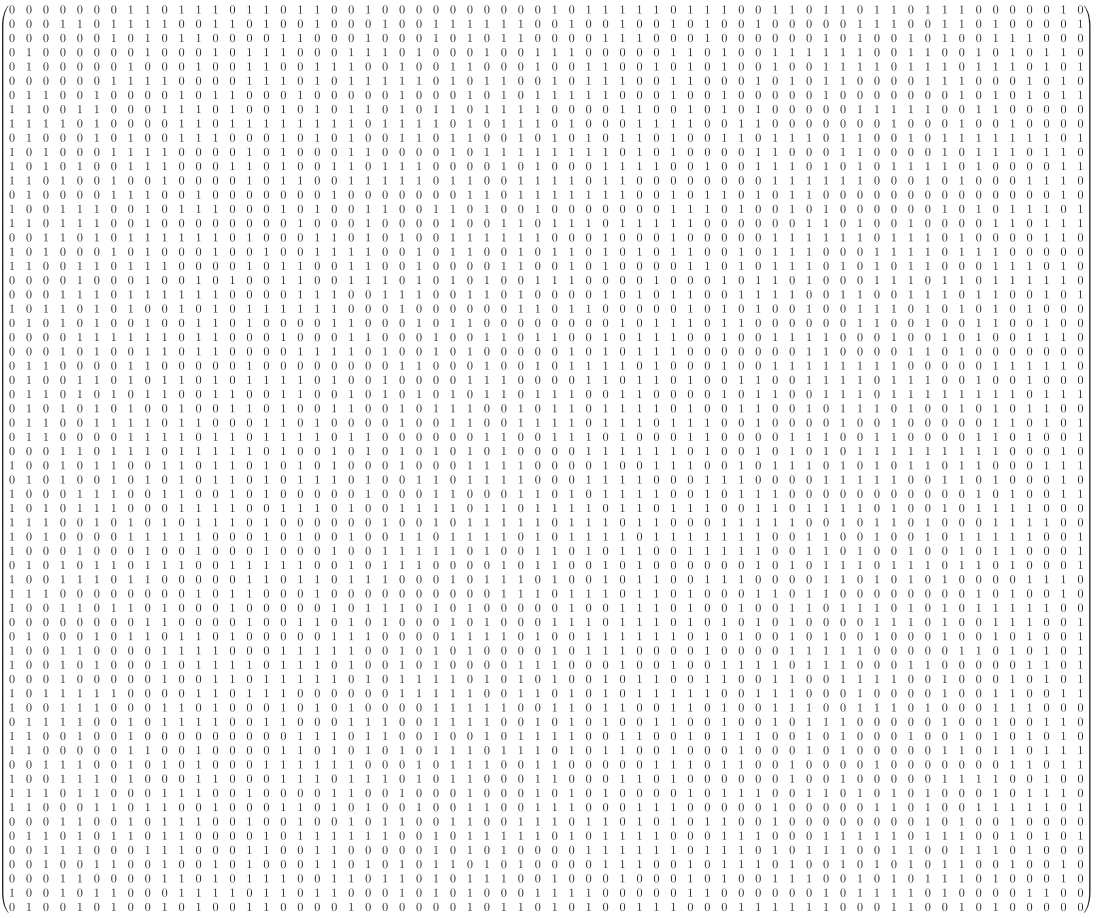


FIGURE A.1 Adjacency matrix of the 64 vertex counterexample

This example highlights the power of deducing a local limit theorem from a ‘fixed size’ model, especially in a case such as this where the independent model does not satisfy a local central limit theorem as demonstrated in [6]. Indeed, we end up with the ‘central limit behavior’ at the scale of σ_k , along with a multiplier that depends on the $\sigma_k n^{-1/2}$ scale that oscillates according to a theta series.

This technique also immediately gives the precise asymptotic for the maximal point in the distribution of the number of k -term arithmetic progressions, which answers a question in [11]. From the above, we see that the answer is

$$\sup_{a \in \mathbb{Z}} \mathbb{P}[\text{kAP}(\mathbf{x}) = a \mid \text{Ber}(p)] = \frac{\sup g(x)}{\sigma_k} (1 + O_{\lambda, \varepsilon}(n^{\varepsilon-1/4})).$$

APPENDIX: ADJACENCY MATRIX OF THE CONSTRUCTION

Figure A.1 shows the adjacency matrix of the counterexample mentioned in Section 6. It is also included separately in the arXiv listing of this paper.

A visualization of this graph is shown in Figure A.2.

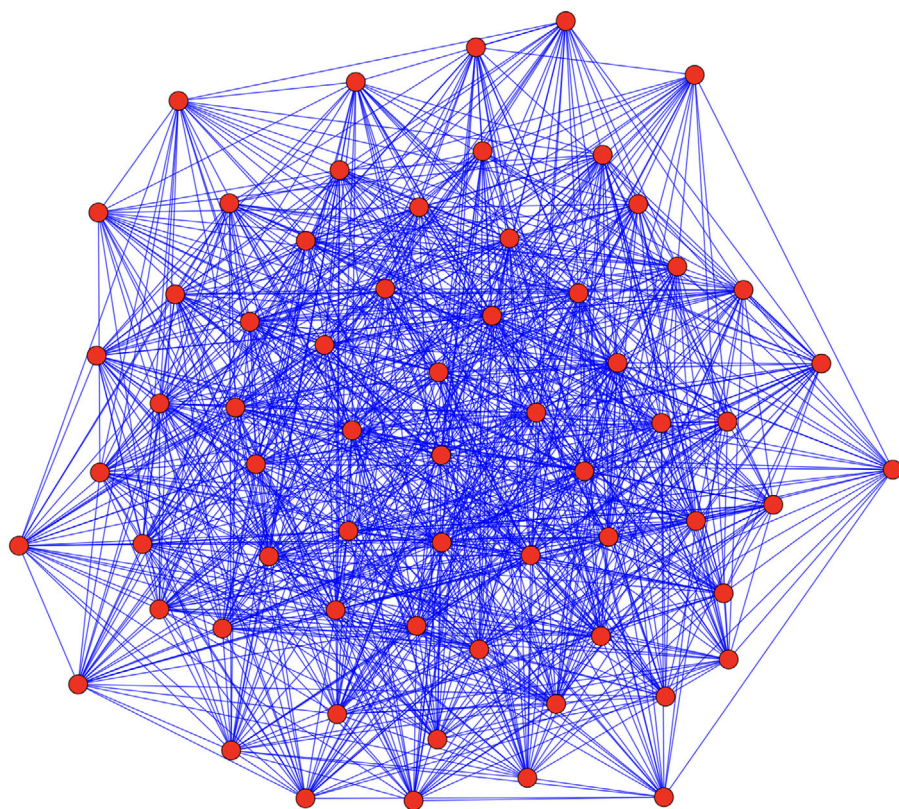


FIGURE A.2 Graph drawing of the 64 vertex counterexample

ACKNOWLEDGEMENTS

We thank Yufei Zhao for suggesting the problem and extensive comments on the manuscript, and thank Ross Berkowitz for useful discussions about the subgraph count problem. We thank Vishesh Jain for mentioning the trick of using hypercontractivity on the hypercube to deduce bounds on a slice. We thank Daniel Zhu for various typographical fixes. Finally we thank the referee for extraordinarily careful proofreading and various edits which substantially improved the presentation.

JOURNAL INFORMATION

The *Journal of the London Mathematical Society* is wholly owned and managed by the London Mathematical Society, a not-for-profit Charity registered with the UK Charity Commission. All surplus income from its publishing programme is used to support mathematicians and mathematics research in the form of research grants, conference grants, prizes, initiatives for early career researchers and the promotion of mathematics.

REFERENCES

1. A. D. Barbour, M. Karoński, and A. Ruciński, *A central limit theorem for decomposable random variables with applications to random graphs*, J. Combin. Theory Ser. B **47** (1989), 125–145.
2. M. Behrisch, A. Coja-Oghlan, and M. Kang, *The order of the giant component of random hypergraphs*, Random Structures Algorithms **36** (2010), 149–184.

3. E. A. Bender, E. R. Canfield, and B. D. McKay, *Asymptotic properties of labeled connected graphs*, Random Structures Algorithms **3** (1992), 183–202.
4. R. Berkowitz, *A local limit theorem for cliques in $G(n, p)$* , arXiv:1811.03527.
5. R. Berkowitz, *A quantitative local limit theorem for triangles in random graphs*, arXiv:1610.01281.
6. R. Berkowitz, A. Sah, and M. Sawhney, *Number of arithmetic progressions in dense random subsets of $\mathbb{Z}/n\mathbb{Z}$* , Isr. J. Math. **244** (2021), 589–620.
7. B. B. Bhattacharya, S. Ganguly, X. Shao, and Y. Zhao, *Upper tail large deviations for arithmetic progressions in a random set*, Int. Math. Res. Not. IMRN (2020), no. 1, 167–213.
8. B. DeMarco and A. Redlich, *Graph decomposition and parity*, J. Graph Theory **82** (2016), 374–386.
9. P. Erdős and A. Rényi, *On the evolution of random graphs*, Bull. Inst. Internat. Statist. **38** (1961), 343–347.
10. C. G. Esseen, *On the Kolmogorov-Rogozin inequality for the concentration function*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **5** (1966), 210–216.
11. J. Fox, M. Kwan, and L. Saueremann, *Anticoncentration for subgraph counts in random graphs*, Ann. Probab. **49** (2021), no. 3, 1515–1553.
12. J. Gilmer and S. Kopparty, *A local central limit theorem for triangles in a random graph*, Random Structures Algorithms **48** (2016), 732–750.
13. B. V. Gnedenko, *On a local limit theorem of the theory of probability*, Uspekhi Mat. Nauk **3** (1948), 187–194.
14. M. Harel, F. Mousset, and W. Samotij, *Upper tails via high moments and entropic stability*, arXiv:1904.08212.
15. H.-K. Hwang, *Limit theorems for mergesort*, Random Structures Algorithms **8** (1996), 319–336.
16. V. Jain, *Approximate spielman-teng theorems for the least singular value of random combinatorial matrices*, Isr. J. Math. **242** (2021), 461–500.
17. S. Janson, *Orthogonal decompositions and functional limit theorems for random graph statistics*, Mem. Amer. Math. Soc. **111** (1994), vi+78.
18. S. Janson, *A graph Fourier transform and proportional graphs*, Proceedings of the Sixth International Seminar on Random Graphs and Probabilistic Methods in Combinatorics and Computer Science, “Random Graphs ’93” (Poznań, 1993), vol. 6, 1995, pp. 341–351.
19. S. Janson, T. Łuczak, and A. Ruciński, *Random graphs*, Wiley-interscience series in discrete mathematics and optimization, Wiley-Interscience, New York, 2000.
20. S. Janson and L. Warnke, *The lower tail: Poisson approximation revisited*, Random Structures Algorithms **48** (2016), 219–246.
21. J. Kärman, *An example of a superproportional graph*, Proceedings of the Fifth International Seminar on Random Graphs and Probabilistic Methods in Combinatorics and Computer Science (Poznań, 1991), vol. 5, 1994, pp. 95–98.
22. P. G. Kolaitis and S. Kopparty, *Random graphs and the parity quantifier*, J. ACM **60** (2013), 1–34.
23. M. Loeb, J. Matoušek, and O. Pangrác, *Triangles in random graphs*, Discrete Math. **289** (2004), 181–185.
24. R. Meka, O. Nguyen, and V. Vu, *Anti-concentration for polynomials of independent random variables*, Theory Comput. **12** (2016), 1–17.
25. K. Nowicki and J. C. Wierman, *Subgraph counts in random graphs using incomplete U-statistics methods*, Proceedings of the First Japan Conference on Graph Theory and Applications (Hakone, 1986), vol. 72, 1988, pp. 299–310.
26. R. O’Donnell, *Analysis of Boolean functions*, Cambridge University Press, New York, 2014.
27. N. Ross, *Fundamentals of Stein’s method*, Probab. Surv. **8** (2011), 210–293.
28. A. Ruciński, *When are small subgraphs of a random graph normally distributed?*, Probab. Theory Related Fields **78** (1988), 1–10.
29. T. Tao and V. Vu, *Additive combinatorics*, Cambridge Studies in Advanced Mathematics, vol. 105, Cambridge University Press, Cambridge, 2006.
30. L. Warnke, *Upper tails for arithmetic progressions in random subsets*, Israel J. Math. **221** (2017), 317–365.