LOCAL LIMIT THEOREM FOR JOINT SUBGRAPH COUNTS

ASHWIN SAH[1], MEHTAAB SAWHNEY[2], AND DANIEL G. ZHU[3]

ABSTRACT. Extending a previous result of the first two authors, we prove a local limit theorem for the joint distribution of subgraph counts in the Erdős-Rényi random graph $G(n, p)$. This limit can be described as a nonlinear transformation of a multivariate normal distribution, where the components of the multivariate normal correspond to the graph factors of Janson. As an application, we show a number of results concerning the existence and enumeration of proportional graphs and related concepts, answering various questions of Janson and collaborators in the affirmative.

1. INTRODUCTION

For a fixed graph H and probability $p \in (0, 1)$, the count X_H of subgraphs isomorphic to H in the Erdős-Rényi random graph $G(n, p)$ has been one of the most frequently studied random variables in the theory of random graphs. While a central limit theorem for X_H, establishing that it, properly normalized, converges to a normal distribution in the limit $n \to \infty$, was established by Nowicki in 1985 [13] (see also [14, 17]), a series of recent works has focused on the local aspects of the distribution of X_H, with the aim of controlling the point probabilities $\mathbb{P}[X_H = x]$ (see [3] and references therein). Chief among these results are a series of local limit theorems [4, 2, 1, 18], culminating in the following local limit theorem of the first two authors for all connected subgraph counts:

Theorem 1.1 ([18, Theorem 1.1]). *Let H be a connected graph with at least two vertices and let $p \in (\lambda, 1 - \lambda)$. Then, for every $\varepsilon > 0$ and $x \in \mathbb{Z}$, we have*

$$\left| \sigma \mathbb{P}[X_H = x] - \mathcal{N}\left(\frac{x - \mathbb{E}[X_H]}{\sigma} \right) \right| \lesssim_{H, \lambda, \varepsilon} n^{-1/2 + \varepsilon},$$

where $\sigma^2 = \mathrm{Var}[X_H]$ and $\mathcal{N}(z) = (2\pi)^{-1/2} e^{-z^2/2}$ is the standard normal probability distribution function.

The main objective of this paper is to develop a local limit theorem for joint distributions of connected subgraph counts, controlling probabilities of the form $\mathbb{P}[X_H = x_H \; \forall H \in \mathcal{H}]$ for certain sets of connected graphs $\mathcal{H}$. Such a result would be applicable far beyond connected subgraph counts, as every graph statistic counting structures of bounded size,[1] including subgraph counts of disconnected graphs and induced subgraph counts, can be written as a function of n and a finite number of connected subgraph counts (see Section 2 of this paper for more details).

[1]DEPARTMENT OF MATHEMATICS, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE, MA 02139, USA
[2]DEPARTMENT OF MATHEMATICS, COLUMBIA UNIVERSITY, NEW YORK, NY 10027, USA
[3]DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY, PRINCETON, NJ 08544, USA
E-mail addresses: asah@mit.edu, m.sawhney@columbia.edu, zhd@princeton.edu.

[1]Formally, such a graph statistic must be vertex-symmetric and satisfy the condition that, for every n, it can be written as a polynomial in the indicator functions of the edges of the graph with degree bounded independently of n.

From a purely distributional standpoint, the behavior of $(X_H)_{H \in \mathcal{H}}$ was settled (albeit somewhat unsatisfyingly) by Nowicki [13], who observed that for large n, almost all the variation in X_H is controlled by the edge count of the graph, thus implying that in the $n \to \infty$ limit, any finite collection of subgraph counts converges in distribution to a collection of perfectly correlated normal random variables. In order to study graph statistics on $G(n, p)$ for which the above description is insufficiently precise, Janson [5, 7] developed the notion of *graph factors* γ_H, analogues of subgraph counts obtained through an orthogonalization process. Specifically, letting x_e be the indicator variable of an edge e and $\chi_e = (x_e - p)/\sqrt{p(1-p)}$ be a normalized version of x_e satisfying $\mathbb{E}[\chi_e] = 0$ and $\mathbb{E}[\chi_e^2] = 1$, Janson defined

$$\gamma_H = \sum_{H' \cong H} \prod_{e \in E(H')} \chi_e,$$

where the sum is over all subgraphs H' of K_n isomorphic to H. It is straightforward to show that if H and H' have no isolated vertices, then $\mathbb{E}[\gamma_H \gamma_{H'}]$ is 0 if $H \not\cong H'$ and otherwise equal to the number of subgraphs of K_n isomorphic to $H \cong H'$. Moreover, we will show in Section 2 of this paper that if we restrict ourselves to connected H, then the γ_H are algebraically independent, and that the X_H can be canonically be written as polynomials in the γ_H, and vice versa.

Janson proved that for any finite collection $\mathcal{H}$ of (nonisomorphic) connected graphs with at least two vertices,[2] the distribution of $(\gamma_H)_{H \in \mathcal{H}}$, after appropriate scaling, converges to a standard multivariate normal distribution, with no correlations between different graph factors. Transforming back to subgraph counts yields a model of the joint distribution of subgraph counts consisting of a polynomial transform of a multivariate normal distribution.

The main result of this paper states that this model is asymptotically correct even at the level of point probabilities.

Theorem 1.2. *Let $p \in (\lambda, 1 - \lambda)$ and let $\mathcal{H}$ be a downwards closed set of nonisomorphic connected graphs with at least two vertices (in the sense of Definition 2.7). Then, for every $\varepsilon > 0$ and permissible $(y_H)_{H \in \mathcal{H}}$ (in the sense of Definition 2.11) we have*

$$\left| \mathbb{P}[\gamma_H = y_H \; \forall H \in \mathcal{H}] \prod_{H \in \mathcal{H}} \left((p(1-p))^{e(H)/2} \sigma_H \right) - \prod_{H \in \mathcal{H}} \mathcal{N}(y_H/\sigma_H) \right| \lesssim_{\mathcal{H}, \lambda, \varepsilon} n^{-1/2+\varepsilon},$$

where $\sigma_H^2 = \mathrm{Var}[\gamma_H]$ and $\mathcal{N}(z) = (2\pi)^{-1/2} e^{-z^2/2}$ is the standard normal probability distribution function.

Two terms in Theorem 1.2, which will be defined formally in Section 2, merit discussion. First of all, the "downwards closed" condition is needed to ensure that there is a bijective mapping between $(\gamma_H)_{H \in \mathcal{H}}$ and $(X_H)_{H \in \mathcal{H}}$, and encompasses choices for $\mathcal{H}$ such as the set of connected graphs H with $2 \leq v(H) \leq a$ and $e(H) \leq b$, for arbitrary positive integers $a \geq 2$ and $b \geq 1$. Second, a tuple $(y_H)_{H \in \mathcal{H}}$ is permissible if and only if it corresponds to values for $(X_H)_{H \in \mathcal{H}}$ that are integers; obviously, if $(y_H)_{H \in \mathcal{H}}$ is not permissible, we have $\mathbb{P}[\gamma_H = y_H \; \forall H \in \mathcal{H}] = 0$. The set of permissible $(y_H)_{H \in \mathcal{H}}$ forms a "skew lattice"[3] with density $(p(1-p))^{\sum_{H \in \mathcal{H}} e(H)/2}$, explaining the appearance of that term in Theorem 1.2.

Combined with appropriate tail bounds on the γ_H, Theorem 1.2 paves the way for a complete description of the local distribution of any graph statistic that can be written in terms of a finite number of connected subgraph counts, including disconnected subgraph counts and induced

---

[2]If H is a single vertex, γ_H is always n and thus is uninteresting in this context.

[3]For the purposes of this introduction, a *skew lattice* is, after possibly permuting coordinates, the image of $\mathbb{Z}^d$ under a "upper triangular" map of the form $(x_1, \ldots, x_d) \mapsto (a_1 x_1 + \psi_1, a_2 x_2 + \psi_2(x_1), \ldots, a_d x_d + \psi_d(x_1, \ldots, x_{d-1}))$ for nonzero $a_1, \ldots, a_d$, constant ψ_1, and arbitrary functions $\psi_2, \ldots, \psi_d$. The density of such a skew lattice is $1/|a_1 a_2 \cdots a_d|$.

subgraph counts which may not necessarily have an associated local central limit theorem (see [18, Theorem 1.3]). Moreover, Theorem 1.2 locally controls the joint distribution of any collection of such graph statistics; in particular, since edge count is a subgraph count, it allows for local limit theorems to be proven in the $G(n, m)$ model as well.

Another application of Theorem 1.2 is to show the existence of graphs with exactly specified subgraph counts, as it implies that given any integer tuple $(x_H)_{H \in \mathcal{H}}$ depending on n, the probability that $X_H = x_H$ for all H is positive for sufficiently large n, provided that corresponding values of $(\gamma_H/\sigma_H)_{H \in \mathcal{H}}$ are bounded. In Section 6, we apply this to the theory of proportional graphs, developed to describe the different asymptotic behaviors of the induced subgraph count Y_H of a graph H in the $G(n, p)$ and $G(n, m)$ models. Answering a question of Janson [7], we show that for almost all[4] rational $p \in (0, 1)$ (the characterization of which we explicitly state), there exist infinitely many H such that the variable Y_H in the $G(n, \lfloor p\binom{n}{2} \rfloor)$ model satisfies $\mathrm{Var}[Y_H] \asymp n^{2v(H)-6}$ and converges to a nonnormal distribution. Interestingly, the integrality constraints translate to nontrivial number theory; for instance, if we set $p = \frac{1}{2}$, every such H must have more than 10^{390} vertices!

Proof overview.

Fourier analysis, Stein's method, and decoupling. The proof of Theorem 1.2 is Fourier-analytic in nature and follows the same basic outline as [18] (which in turn generalized techniques from [1]). After applying Fourier inversion, we aim to show a multivariate characteristic function $\varphi_X^{\mathcal{F}}(\boldsymbol{t})$ (whose precise definition we will postpone) involving the X_H is close to $\varphi_Z^{\mathcal{F}}(\boldsymbol{t})$, the characteristic function of a transformed multivariate normal. To accomplish this, we use two classes of techniques depending on the size of $\boldsymbol{t}$. At small frequencies, we use a multivariate generalization of Stein's method of exchangeable pairs due to Meckes [12] to show that the γ_H are well-approximated by independent normal random variables. At larger $\boldsymbol{t}$, where $\varphi_Z^{\mathcal{F}}(\boldsymbol{t})$ becomes negligibly small, we use a decoupling trick of Berkowitz [1] to show that $\varphi_X^{\mathcal{F}}(\boldsymbol{t})$ is similarly small. Numerous different setups are needed to handle various cases concerning the sizes of the components of $\boldsymbol{t}$, but as in [18], there is a fundamental distinction between $\boldsymbol{t}$ of intermediate-size and $\boldsymbol{t}$ that are extremely large.

Integral factor systems. A major step within the proof is to identify the appropriate lattice on which to apply Fourier inversion. If one uses $(X_H)_{H \in \mathcal{H}}$, the aforementioned correlations between the different X_H make it so that the characteristic function is no longer concentrated sufficiently near the origin, rendering our proof method useless. On the other hand, the space of permissible $(\gamma_H)_{H \in \mathcal{H}}$, where these correlations have been removed, is not a lattice at all! To resolve this issue, we introduce what we call an *integral factor system* $\mathcal{F} = (F_H)_{H \in \mathcal{H}}$, which behaves as a hybrid of (X_H) and (γ_H). Like (γ_H), most of the correlations between the X_H have been removed, making the characteristic function well-behaved. Moreover, the set of (F_H) corresponding to integral (X_H) is precisely $\mathbb{Z}^{\mathcal{H}}$, which allows Fourier inversion to be used. For further discussion motivating integral factor systems, see Section 2.2. We believe the underlying techniques will prove useful in many situations where one seeks to establish a multivariate local limit theorem.

Polynomial bases for graph statistics. To facilitate the conversion between the X_H, F_H, and γ_H, we introduce a number of closely related algebraic objects that keep track of the relations between these statistics, which can be viewed as either a discrete analogue of the flag algebras of Razborov [16], or a graph-theoretic analogue of the ring of symmetric functions. Just as there are many standard bases for symmetric functions (such as m_λ, p_λ, e_λ, and s_λ), we find several different bases (such as X_H and γ_H) of these objects, which are all used at different points in the proof.

---

[4]though not cofinitely many

Outline. In Section 2, we discuss the algebraic relations between the X_H and γ_H and construct integral factor systems. In Section 3 we set up the main Fourier inversion, while we control the characteristic function in Section 4 for small $\boldsymbol{t}$ and in Section 5 for large $\boldsymbol{t}$. In Section 6 we discuss applications to proportional graphs and related concepts.

Notation and conventions. All graphs considered in this paper are finite and simple. Given a graph G, we let $V(G)$ and $E(G)$ denote the vertices and edges of G, respectively, and let $v(G)$, $e(G)$, and $\operatorname{aut} G$ denote the number of vertices, edges, and automorphisms of G. We let $\sqcup$ denote the disjoint union of graphs, kG denote the disjoint union of k copies of G, and $\bar{G}$ denote the complement of G. We let K_n, $K_{m,n}$, and P_n denote the complete, complete bipartite, and path graph on n vertices, $m + n$ vertices, and n edges, respectively. We consider the empty graph K_0 to be disconnected and denote it using the symbol $\varnothing$. We also let $\bullet$ be a shorthand for K_1. Finally, given a graph H and partition $\mathcal{P}$ of its vertices such that no vertices in the same part are connected by an edge, we define $H/\mathcal{P}$ to be a graph on $\mathcal{P}$, with two partitions connected if there is at least one edge of H between their individual vertex sets.

We take $f \lesssim g$ to mean $f = O(g)$, with subscripts on either the $\lesssim$ or the O denoting dependence in the explicit constants. We let $f \asymp g$ mean $f = \Theta(g)$, with subscripts treated similarly. We define $[n] = \{1, 2, \ldots, n\}$, which we at times identify with the vertices of K_n and $G(n, p)$. When considering some $0 < p < 1$, we will always have $p \in (\lambda, 1 - \lambda)$. Whenever considering a set of graphs $\mathcal{H}$, the letter ℓ will always denote the maximum number of vertices of a graph in $\mathcal{H}$. Note that an asymptotic with constant depending on ℓ is equivalent to an asymptotic with constant depending on $\mathcal{H}$, as for each value of ℓ there are only finitely many possibilities for $\mathcal{H}$.

2. Subgraph Count Arithmetic

2.1. **Polynomial relations between subgraph count statistics.** In this subsection, we will treat X_H and γ_H as functions $\{\text{graphs}\} \to \mathbb{R}$. We will also define the following other statistics:

Definition 2.1. For a graph H let $\tilde{X}_H = \operatorname{aut} H \cdot X_H$ be the number of injective homomorphisms from H into a given graph. If $H_1, \ldots, H_m$ are the connected components of H, then let $X_H^* = \prod_{i \in [m]} X_{H_i}$ and $\tilde{X}_H^* = \prod_{i \in [m]} \tilde{X}_{H_i}$. Define $\tilde{\gamma}_H$, γ_H^*, and $\tilde{\gamma}_H^*$ similarly.

Remark 2.2. It is not in general true that $\tilde{X}_H^* = \operatorname{aut} H \cdot X_H^*$, since $\operatorname{aut}(H_1 \sqcup H_2)$ is not necessarily equal to $\operatorname{aut} H_1 \cdot \operatorname{aut} H_2$.

Definition 2.3. Given two graphs H_1 and H_2, write $H_1 \preceq H_2$ if H_1 can be obtained from H_2 via a combination of edge deletion, vertex deletion, and merging disconnected vertices. Write $H_1 \prec H_2$ if $H_1 \preceq H_2$ but $H_1 \neq H_2$.

This is clearly a partial order on graphs.

Proposition 2.4. *Let μ_H and ν_H denote any two of*

$$X_H, \quad X_H^*, \quad (p(1-p))^{e(H)/2}\gamma_H, \quad \text{or } (p(1-p))^{e(H)/2}\gamma_H^*.$$

Then, for any H, one can write ν_H as a linear combination $\sum_{H' \preceq H} a_{H'}\mu_{H'}$, where $a_H \asymp_H 1$, $a_{H'} \lesssim_H 1$ for all $H' \preceq H$, and $a_H = 1$ if H is connected. (Note that these bounds are independent of p.)

Proof. It is not difficult to see that the condition in Proposition 2.4 between μ_H and ν_H is an equivalence relation. Therefore it suffices to check the condition three times for suitable μ_H and ν_H.

First we consider $\mu_H = X_H$ and $\nu_H = X_H^*$. If H is connected, we have $X_H^* = X_H$, so we are done. Otherwise, letting $H_1, \ldots, H_m$ denote the connected components of H, we claim that

$$\tilde{X}_H^* = \sum_{\mathcal{P}} \tilde{X}_{H/\mathcal{P}},$$

where $\mathcal{P}$ runs over all set partitions of $V(H)$ where all elements of $V(H_i)$ are in different parts for all i. Indeed, the left-hand side, evaluated on a given G, counts the number of graph homomorphisms $H \to G$ that are injective on each H_i. For each such homomorphism, there is a unique such set partition describing which vertices of H map to the same vertex of G, and the number of homomorphisms corresponding to a given set partition $\mathcal{P}$ can be easily shown to be exactly $\tilde{X}_{H/\mathcal{P}}$. Note that $H/\mathcal{P} \preceq H$ for all $\mathcal{P}$, and equality holds if and only if all the components of $\mathcal{P}$ are singletons. Converting from the $\tilde{X}_H$ and $\tilde{X}_H^*$ to X_H and X_H^*, we get the desired linear relation where all coefficients are independent of p (and hence certainly $O_H(1)$) and the coefficient of H is positive.

Now we consider $\mu_H = (p(1-p))^{e(H)/2}\gamma_H$ and $\nu_H = X_H$. Since $x_e = \sqrt{p(1-p)}\chi_e + p$, we conclude that

$$\tilde{X}_H = \sum_{H' \subseteq H} p^{e(H)-e(H')}(p(1-p))^{e(H')/2}\tilde{\gamma}_{H'}$$

and hence

$$(2.1) \qquad X_H = \sum_{H' \subseteq H} \frac{\operatorname{aut} H'}{\operatorname{aut} H} p^{e(H)-e(H')}(p(1-p))^{e(H')/2}\gamma_{H'}.$$

Finally, we consider $\mu_H = X_H^*$ and $\nu_H = (p(1-p))^{e(H)/2}\gamma_H^*$. Here, note that by combining the previous two cases we may construct identities of the form $(p(1-p))^{e(H)/2}\gamma_H = \sum_{H' \preceq H} a_{H'} X_{H'}^*$. Multiplying these identities for various connected H yields the desired. (Here we have used the fact that if $H_1' \preceq H_1$ and $H_2' \preceq H_2$, then $H_1' \sqcup H_2' \preceq H_1 \sqcup H_2$.) $\qquad\square$

We now construct the following object to encapsulate all these relations:

Definition 2.5. Let $\mathcal{R}$ be the $\mathbb{R}$-vector space generated by the X_H.

Corollary 2.6. *The sets $\{X_H\}$, $\{X_H^*\}$, $\{\gamma_H\}$, and $\{\gamma_H^*\}$ are all bases of $\mathcal{R}$. Moreover, as rings,*

$$\mathcal{R} = \mathbb{R}[X_H \colon H \text{ connected}] = \mathbb{R}[\gamma_H \colon H \text{ connected}].$$

Proof. The functions X_H are linearly independent, since X_H is 1 on the graph H but 0 on any graph $H' \prec H$. Thus $\{X_H\}$ is a basis of $\mathcal{R}$. The fact that $\{X_H^*\}$, $\{\gamma_H\}$, and $\{\gamma_H^*\}$ are bases as well follows from Proposition 2.4. The fact that $\mathcal{R} = \mathbb{R}[X_H : H \text{ connected}]$ (resp. $\mathbb{R}[\gamma_H : H \text{ connected}]$) is an alternate way of saying that the X_H^* (resp. γ_H^*) form a basis of $\mathcal{R}$. $\qquad\square$

In particular, the fact that $\mathcal{R} = \mathbb{R}[X_H \colon H \text{ connected}]$ implies that $\mathcal{R}$ is an $\mathbb{R}$-algebra.

At this point, we may now define the terms used in Theorem 1.2.

Definition 2.7. A finite set of non-$\bullet$ connected graphs $\mathcal{H}$ is *downwards closed* if for all $H \in \mathcal{H}$ and $H' \preceq H$ with H' connected and not $\bullet$, we have $H' \in \mathcal{H}$. In this case, it is true that if $H' \preceq H \in \mathcal{H}$, all components of H' are in $\mathcal{H} \cup \{\bullet\}$; thus Proposition 2.4 implies that

$$\mathbb{R}[X_H \colon H \in \mathcal{H} \cup \{\bullet\}] = \mathbb{R}[\gamma_H \colon H \in \mathcal{H} \cup \{\bullet\}].$$

Call this algebra $\mathcal{R}_{\mathcal{H}}$.

Remark 2.8. There does not appear to be a nice basis of $\mathcal{R}_{\mathcal{H}}$ in terms of the unstarred X_H or γ_H.

Now, in $\mathcal{R}$, we have $X_\bullet = \gamma_\bullet$, and both evaluate to the number of vertices of a given graph. Thus, if we are working with graphs with a fixed number of vertices n, it makes sense to consider the following object.

Definition 2.9. For an integer n, let $\mathcal{R}_n = \mathcal{R}/(\gamma_\bullet - n)$ and for downwards closed $\mathcal{H}$ let $\mathcal{R}_{\mathcal{H},n} = \mathcal{R}_\mathcal{H}/(\gamma_\bullet - n)$. Note that the intersection of the ideal $(\gamma_\bullet - n) \subseteq \mathcal{R}$ with $\mathcal{R}_\mathcal{H}$ is the ideal generated $(\gamma_\bullet - n)$ in $\mathcal{R}_\mathcal{H}$, so we may treat $\mathcal{R}_{\mathcal{H},n}$ as a subalgebra of $\mathcal{R}_n$.

Remark 2.10. Since $\mathcal{R}$ was defined as a subalgebra of the algebra of functions $\{\text{graphs}\} \to \mathbb{R}$, a relation between statistics in $\mathcal{R}$ can be verified by checking it for all graphs. However, this is not the case for $\mathcal{R}_n$; there are elements that evaluate to zero on all n-vertex graphs (such as X_H for H with more than n vertices) but are not actually zero.

We may now define the notion of permissibility as being values of $(\gamma_H)_{H \in \mathcal{H}}$ "corresponding" to integral subgraph counts, in the following manner.

Definition 2.11. An *evaluation* is an algebra homomorphism $\Phi \colon R_{\mathcal{H},n} \to \mathbb{R}$. Call an evaluation Φ *integral* if $\Phi(X_H) \in \mathbb{Z}$ for all $H \in \mathcal{H}$. For a given integer n and downwards closed $\mathcal{H}$, a tuple $(y_H)_{H \in \mathcal{H}}$ is *permissible* if there exists an integral evaluation Φ on $\mathcal{R}_{\mathcal{H},n}$ such that $\Phi(\gamma_H) = y_H$ for all $H \in \mathcal{H}$.

Finally, as we will frequently work with graphs with no isolated vertices, we abbreviate "graph with no isolated vertices" to "*NIV* graph".

2.2. **Motivational remarks on integral factor systems.** In order to prove Theorem 1.2, we will use Fourier inversion, deriving it from the result that a characteristic function based on the X_H is close to a characteristic function based on a multivariate normal. However, the appropriate lattice to use is far from obvious. To illuminate the types of issues that can arise, we begin with a motivational note.

As a toy example, let (X_1, X_2) be integer-valued random variables (depending on n) and suppose we wish to show that (X_1, X_2) is locally close in distribution to $(nZ_1, n^2 Z_2 + f(Z_1))$, where Z_1 and Z_2 are independent standard normal random variables and f is an arbitrary, but reasonably well-behaved, function (also depending on n). Since the map $(z_1, z_2) \mapsto (nz_1, n^2 z_2 + f(z_1))$ has Jacobian determinant n^3 for differentiable f, we formally wish to show that

$$\mathbb{P}[(X_1, X_2) = (x_1, x_2)] = \frac{\mathcal{N}(\frac{x_1}{n})\mathcal{N}(\frac{x_2 - f(x_1/n)}{n^2}) + o(1)}{n^3}.$$

A naïve attempt at Fourier inversion would see us define

$$\varphi_X(t_1, t_2) = \mathbb{E}[e^{i(t_1 X_1 + t_2 X_2)}] \quad \text{and} \quad \varphi_Z(t_1, t_2) = \mathbb{E}[e^{i(t_1 \cdot n Z_1 + t_2(n^2 Z_2 + f(Z_1)))}],$$

so that

$$\mathbb{P}[(X_1, X_2) = (x_1, x_2)] = \frac{1}{(2\pi)^2} \int_{[-\pi,\pi]^2} e^{-i(t_1 x_1 + t_2 x_2)} \varphi_X(\boldsymbol{t}) \, d\boldsymbol{t}$$

and

$$\frac{\mathcal{N}(\frac{x_1}{n})\mathcal{N}(\frac{x_2 - f(x_1/n)}{n^2})}{n^3} = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} e^{-i(t_1 x_1 + t_2 x_2)} \varphi_Z(\boldsymbol{t}) \, d\boldsymbol{t}.$$

Thus, we will be done if we can show that $\varphi_X(\boldsymbol{t})$ and $\varphi_Z(\boldsymbol{t})$ are sufficiently close.

To investigate the problems that can occur, we will describe $\varphi_Z(\boldsymbol{t})$ for various choices of f. First of all, in the case where $f(z) = 0$, we simply have $\varphi_Z(\boldsymbol{t}) = e^{-((nt_1)^2 + (n^2 t_2)^2)/2}$, which is concentrated in an region of area $\asymp n^{-3}$ near the origin; this situation is shown in Figure 1(a). If one could show

FIGURE 1. Conceptual drawings of $\varphi_Z(\boldsymbol{t})$ on $[-\pi, \pi]^2$ for (a) $f(z) = 0$, (b) $f(z) = \alpha n^3 z$, and (c) $f(z) = \alpha n^3 z^2$, for some constant $\alpha \asymp 1$. Black areas indicate where $\varphi_Z(\boldsymbol{t})$ is of constant order, gray areas indicate where $\varphi_Z(\boldsymbol{t})$ is polynomially small, and white areas indicate where $\varphi_Z(\boldsymbol{t})$ is superpolynomially small.

$|\varphi_X(\boldsymbol{t}) - \varphi_Z(\boldsymbol{t})| = o(1)$ near the origin and $\varphi_X(\boldsymbol{t}) = n^{-\omega(1)}$ elsewhere in $[-\pi, \pi]^2$, the local limit theorem would then be proved.

A complication occurs, however, if we have $f(z) = \alpha n^3 z$, for some constant $\alpha \asymp 1$. In this case,

$$\varphi_Z(\boldsymbol{t}) = e^{-((nt_1 + \alpha n^3 t_2)^2 + (n^2 t_2)^2)/2},$$

a sheared version of the previous case; this situation is shown in Figure 1(b). While the characteristic function is still well-concentrated in a region of area n^{-3}, the region in question is roughly a slanted rectangle of size $\Theta(1) \times \Theta(n^{-3})$, which notably escapes the square $[-\pi, \pi]^2$. In this case, our goal would be to show that $\varphi_X(\boldsymbol{t})$ is concentrated on the image of this box under the "mod 2π" reduction map $\mathbb{R}^2 \to [-\pi, \pi)^2$, and we would have to carefully construct a correspondence between values of $\boldsymbol{t}$ and $\boldsymbol{t}'$ with $\boldsymbol{t} \equiv \boldsymbol{t}' \pmod{2\pi}$ where we can prove $\varphi_X(\boldsymbol{t}) = \varphi_Z(\boldsymbol{t}') + o(1)$.

While this is already somewhat of a headache, a more serious problem arises if we introduce significant nonlinearity into f; for example, by letting $f(z) = \alpha n^3 z^2$ for some $\alpha \asymp 1$. In this case,

$$\varphi_Z(\boldsymbol{t}) = e^{-(n^2 t_2)^2/2} \cdot \mathbb{E}[e^{i(nt_1 \cdot Z_1 + \alpha n^3 t_2 \cdot Z_1^2)}] = e^{-(n^2 t_2)^2/2} \frac{e^{-n^2 t_1^2/(2 - 4i\alpha n^3 t_2)}}{\sqrt{1 - 2i\alpha n^3 t_2}},$$

which has magnitude

$$\exp\left(-\frac{1}{2}\left(n^4 t_2^2 + \frac{n^2 t_1^2}{1 + 4\alpha^2 n^6 t_2^2}\right)\right) \Big/ \left(1 + 4\alpha^2 n^6 t_2^2\right)^{1/4}.$$

This function is of constant order on a $\Theta(n^{-1}) \times \Theta(n^{-3})$ rectangle centered at the origin, but unlike the previous cases, it does not immediately rapidly decay, and there is now a new intermediate regime in a box of size $\Theta(1) \times \Theta(n^{-2})$ where the characteristic function is of size $\asymp n^{-1/2}$. This situation is shown in Figure 1(c). For our purposes, this is incredibly problematic, as to prove the local limit theorem one would now need bounds of the form $|\varphi_X(\boldsymbol{t}) - \varphi_Z(\boldsymbol{t})| = o(n^{-1})$ in the intermediate regime due to the magnitude of its area, which are implausibly strong.

To motivate the solution to these problems, we make the observation that, fundamentally, all our issues arose because $f(z)$ became much larger than n^2; indeed, if instead we had $f(z) = \alpha n^2 z$ or $f(z) = \alpha n^2 z^2$, one can show that although $\varphi_Z(\boldsymbol{t})$ would undergo some distortion, the basic picture would remain the same as the $f(z) = 0$ case, with large values in a $\Theta(n^{-1}) \times \Theta(n^{-2})$ rectangle near the origin and rapid decay away from this rectangle. To force this to happen, we transform (X_1, X_2) to a different set of integer-valued random variables (F_1, F_2), given by $F_1 = X_1$ and $F_2 = X_2 - g(X_1)$, where g sends integers to integers. It is clear that the local limit theorem we wish

to prove is equivalent to proving that (F_1, F_2) is locally close in distribution to $(nZ_1, n^2 Z_2 + f^*(Z_1))$, where $f^*(z) = f(z) - g(nz)$. In our examples, we may choose $g(x) = \lfloor n^2 \alpha \rfloor x$ and $g(x) = \lfloor n\alpha \rfloor x^2$, respectively, so that $f^*(z) = n\{n^2\alpha\}z$ and $f^*(z) = n^2\{n\alpha\}z^2$, which are sufficiently small for our purposes.

We remark that the fact that such nice choices of g were available cannot be taken for granted; if we instead had to prove that (X_1, X_2) was close to $(n^2 Z_1, n^3 Z_2 + \alpha n^4 Z_1^2)$, then we would need $g(n^2 z) = \alpha(n^2 z)^2 + O(n^3)$, and no integer polynomial g satisfies this condition for noninteger α. Any choice of g in this case would likely be difficult to analyze, and it is doubtful that our methods would easily transfer. Thankfully, in the case of subgraph counts it is always the case that a polynomial transform will suffice.

2.3. **Integral factor systems.** We now define an integral factor system.

Definition 2.12. Given downwards closed $\mathcal{H}$, an integer n, and some $\eta > 0$, a collection of statistics $\mathcal{F} = (F_H)_{H \in \mathcal{H}}$, where $F_H \in \mathcal{R}_{\mathcal{H},n}$ for all $H \in \mathcal{H}$, is an $(\mathcal{H}, n, \eta)$-*integral factor system* (abbreviated $(\mathcal{H}, n, \eta)$-*IFS* or *IFS* if the parameters are unimportant) if for all $H \in \mathcal{H}$, in $\mathcal{R}_n$ we can write

$$F_H = \sum_{\text{NIV } H' \preceq H} a_{H'} X_{H'}^* = \sum_{\text{NIV } H' \preceq H} b_{H'} \gamma_{H'}^* = \sum_{\text{NIV } H' \preceq H} c_{H'} \gamma_{H'},$$

where

(1) all $a_{H'}$ are integers and $a_H = 1$;
(2) $|b_{H'}| \leq \eta n^{(v(H)-v(H'))/2}$ for all H' and $b_H = (p(1-p))^{e(H)/2}$;
(3) $|c_{H'}| \leq \eta n^{(v(H)-v(H'))/2}$ for all H' and $c_H = (p(1-p))^{e(H)/2}$.

Lemma 2.13. *For every $\mathcal{H}$ and n, an $(\mathcal{H}, n, \eta)$-IFS exists with $\eta \lesssim_{\lambda,\ell} 1$.*

The proof of this lemma proceeds in two steps. We first prove it ignoring condition (3) of Definition 2.12. Then we prove that in fact, any F_H satisfying condition (2) must also satisfy (3), up to multiplying η by a constant depending on λ and ℓ.

Proof of Lemma 2.13 ignoring condition (3). We claim that in this case we can actually get $|b_{H'}| \leq 1$ always. To start, observe that if H consists of the disjoint union of an NIV graph H' and m isolated vertices, then $\gamma_H^* = n^m \gamma_{H'}^*$ in $\mathcal{R}_n$. Therefore, by applying Proposition 2.4 with $\mu_H = \gamma_H^*$ and $\nu_H = X_H$ and multiplying such relations, we find that for any NIV graph H, in $\mathcal{R}_n$ we can write X_H^* as a linear combination of $\gamma_{H'}^*$ for NIV $H' \preceq H$ such that the coefficient of γ_H^* is $(p(1-p))^{e(H)/2} < 1$.

We now describe a process to construct F_H. Start with the identity

$$X_H^* = \sum_{\text{NIV } H' \preceq H} s_{H'} \gamma_{H'}^*.$$

While there exists some H' with $|s_{H'}| > 1$, choose one that is maximal under $\preceq$, and add a suitable integer multiple of an identity of the form

$$X_{H'}^* = \sum_{\text{NIV } H'' \preceq H'} s_{H''}' \gamma_{H''}^*$$

to make $|s_{H'}| \leq 1$, which is possible since $s_{H'}'$ is positive but less than 1. Note that this process must terminate, since the only coefficients that can be pushed outside of $[-1, 1]$ are the $s_{H''}$ with $H'' \prec H'$. Also, we never have $H' = H$, since we had $|s_{H'}| \leq 1$ at the start; thus the leading terms on both sides remain X_H and $(p(1-p))^{e(H)/2}\gamma_H$, respectively. In the end, we find that an integer linear combination of $X_{H'}^*$ is equal to a linear combination of $\gamma_{H'}^*$, where all the coefficents on the right-hand side have magnitude bounded by 1. We set F_H to be this quantity. $\square$

To finish the proof, we first need a small lemma.

Lemma 2.14. *Let H be an NIV graph, and suppose $\gamma_H^* = \sum_{H' \preceq H} a_{H'} \gamma_{H'}$ in $\mathcal{R}$. Then, if $a_{H'} \neq 0$, then H' has at most $v(H) - v(H')$ isolated vertices.*

Proof. Let H'' be an NIV graph, and consider $\mathbb{E}_{G(n,p)}[\gamma_{H'} \gamma_{H''}]$ for arbitrary H' with at most ℓ vertices. This is zero unless H' is the disjoint union of H'' and some number m of isolated vertices, in which case it is precisely

$$\frac{v(H'')!}{\operatorname{aut} H''} \binom{n}{v(H'')} \binom{n - v(H'')}{m} \asymp n^{v(H'')+m}.$$

Thus, it follows that if $\mathbb{E}_{G(n,p)}[\gamma_H^* \gamma_{H''}] \lesssim n^{v(H'')+m}$, then we must have $a_{H'} = 0$ for any H' that is the disjoint union of H'' and more than m isolated vertices. We now claim that $\mathbb{E}_{G(n,p)}[\gamma_H^* \gamma_{H''}] \lesssim n^{(v(H'')+v(H))/2}$, which will finish since any H' equal to the disjoint union of H'' and m isolated vertices with $a_{H'} \neq 0$ must satisfy

$$m \leq \frac{v(H) - v(H'')}{2} \iff m \leq v(H) - v(H'') - m = v(H) - v(H').$$

Expanding out γ_H^*, it suffices to show that for NIV graphs $H_1, \ldots, H_m$, we have

$$\mathbb{E}_{G(n,p)}[\gamma_{H_1} \cdots \gamma_{H_m}] \lesssim n^{(v(H_1)+\cdots v(H_m))/2}.$$

To see this, note that if we expand out $\gamma_{H_1} \cdots \gamma_{H_m}$ in terms of the χ_e, we get a sum of many terms, indexed by copies of $H_1, \ldots, H_m$ in an n-vertex complete graph. Each such term is bounded in terms of p, and has zero expectation if some edge is covered only once. Since no H_i has an isolated vertex, no edge being covered once implies that no vertex is covered once, meaning that every term with nonzero expectation uses at most $\frac{1}{2}(v(H_1) + \cdots + v(H_m))$ vertices. Furthermore, there are $O_{H_i}(1)$ possible overlap patterns given a choice of these vertices. Thus there are $O_{H_i}(n^{(v(H_1)+\cdots v(H_m))/2})$ terms with nonzero expectation, concluding the proof. □

Given Lemma 2.14, the finish is relatively straightforward:

Proof that condition (2) implies (3) in Definition 2.12. By Proposition 2.4, we may write each $\gamma_{H'}^*$ as a sum $\sum_{H'' \preceq H'} s_{H''} \gamma_{H''}$ where $s_{H''} \lesssim_{\lambda,\ell} 1$. We may ignore terms with $s_{H''} = 0$. If H'' is the disjoint union of an NIV graph H^* and m isolated vertices, then we have

$$\gamma_{H''} = \binom{n - v(H^*)}{m} \gamma_{H^*},$$

which, by Lemma 2.14, is in fact $O_\ell(n^{\frac{1}{2}(v(H')-v(H^*))}) \gamma_{H^*}$. Combining terms using condition (2) yields the desired expression of $F_H = \sum_{\text{NIV } H' \preceq H} c_{H'} \gamma_{H'}$ with $c_{H'} \lesssim_{\lambda,\ell} n^{\frac{1}{2}(v(H)-v(H'))} \eta$. Moreover, since the coefficient of γ_H in γ_H^* is 1, we have $c_H = (p(1-p))^{e(H)/2}$. □

3. SETTING UP THE MAIN PROOF

In this section, we use Fourier inversion to reduce Theorem 1.2 to two inequalities on a characteristic function: one (Lemma 3.1) at low frequencies and one (Lemma 3.2) at high frequencies, which are proven in the two subsequent sections.

3.1. **Fourier inversion.** Using Lemma 2.13, we take an $(\mathcal{H}, n, \eta)$-IFS $\mathcal{F} = (F_H)_{H \in \mathcal{H}}$ with $\eta \lesssim_{\lambda, \ell} 1$. Observe that we may consider each F_H as both a random variable on $G(n, p)$ and a function of the $(\gamma_H)_{H \in \mathcal{H}}$. Thus, for $\boldsymbol{t} = (t_H)_{H \in \mathcal{H}}$, we may define

$$\varphi_X^{\mathcal{F}}(\boldsymbol{t}) = \mathbb{E}_{G(n,p)}\left[e^{i \sum_H t_H F_H}\right] \quad \text{and} \quad \varphi_Z^{\mathcal{F}}(\boldsymbol{t}) = \mathbb{E}\left[e^{i \sum_H t_H F_H(\sigma_{H'} Z_{H'} \colon H' \in \mathcal{H})}\right],$$

where the $Z_{H'}$ in the second expectation are independent standard normal random variables.

Consider an integral evaluation Φ. Since F_H over $G(n, p)$ is always integer-valued, by Fourier inversion we have

$$\mathbb{P}_{G(n,p)}[F_H = \Phi(F_H) \; \forall H \in \mathcal{H}] = \frac{1}{(2\pi)^{|\mathcal{H}|}} \int_{[-\pi,\pi]^{\mathcal{H}}} e^{-i \sum_H t_H \Phi(F_H)} \varphi_X(\boldsymbol{t}) \, d^{|\mathcal{H}|} \boldsymbol{t}.$$

Meanwhile, the probability density of $(F_H(\sigma_{H'} Z_{H'} \colon H' \in \mathcal{H}))_{H \in \mathcal{H}}$ at $(\Phi(F_H))_{H \in \mathcal{H}}$ is both

$$\frac{1}{(2\pi)^{|\mathcal{H}|}} \int_{\mathbb{R}^{\mathcal{H}}} e^{-i \sum_H t_H \Phi(F_H)} \varphi_Z(\boldsymbol{t}) \, d^{|\mathcal{H}|} \boldsymbol{t},$$

by Fourier inversion, and

$$\prod_{H \in \mathcal{H}} \frac{\mathcal{N}(\Phi(\gamma_H)/\sigma_H)}{(p(1-p))^{e(H)/2} \sigma_H},$$

since the Jacobian determinant of the transform $(\gamma_H)_{H \in \mathcal{H}} \mapsto (F_H)_{H \in \mathcal{H}}$ is $\prod_{H \in \mathcal{H}} (p(1-p))^{e(H)/2}$. As a result

$$\left| \mathbb{P}[\gamma_H = \Phi(\gamma_H) \; \forall H] \prod_{H \in \mathcal{H}} \left((p(1-p))^{e(H)/2} \sigma_H \right) - \prod_{H \in \mathcal{H}} \mathcal{N}(y_H/\sigma_H) \right|$$

$$= \prod_{H \in \mathcal{H}} \left((p(1-p))^{e(H)/2} \sigma_H \right) \frac{1}{(2\pi)^{|\mathcal{H}|}} \left| \int_{\mathbb{R}^{\mathcal{H}}} e^{-i \sum_H t_H \Phi(F_H)} (\varphi_X^{\mathcal{F}}(\boldsymbol{t}) \mathbf{1}_{\boldsymbol{t} \in [-\pi,\pi]^{\mathcal{H}}} - \varphi_Z^{\mathcal{F}}(\boldsymbol{t})) \, d^{|\mathcal{H}|} \boldsymbol{t} \right|$$

$$\lesssim_\ell \prod_{H \in \mathcal{H}} n^{v(H)/2} \int_{\mathbb{R}^{\mathcal{H}}} |\varphi_X^{\mathcal{F}}(\boldsymbol{t}) \mathbf{1}_{\boldsymbol{t} \in [-\pi,\pi]^{\mathcal{H}}} - \varphi_Z^{\mathcal{F}}(\boldsymbol{t})| \, d^{|\mathcal{H}|} \boldsymbol{t}.$$

To show that this expression is $O_{\lambda, \ell, \varepsilon}(n^{-1/2+\varepsilon})$ and hence deduce Theorem 1.2, it suffices to show the following three lemmas:

Lemma 3.1. *Let $\mathcal{H}$ be a downwards closed set of graphs, n be a positive integer, and $\mathcal{F}$ be an $(\mathcal{H}, n, \eta)$-IFS. If we let $\varepsilon > 0$ and $I_\varepsilon = \prod_{H \in \mathcal{H}} [-n^{-v(H)/2+\varepsilon}, n^{-v(H)/2+\varepsilon}]$, then for $\boldsymbol{t} \in I_\varepsilon$ we have*

$$|\varphi_X^{\mathcal{F}}(\boldsymbol{t}) - \varphi_Z^{\mathcal{F}}(\boldsymbol{t})| \lesssim_{\lambda, \ell, \varepsilon, \eta} n^{-1/2+4\varepsilon}.$$

Lemma 3.2. *With $\mathcal{H}, n, \mathcal{F}, \varepsilon, I_\varepsilon$ as above, we have that for $\boldsymbol{t} \in [-\pi, \pi]^{\mathcal{H}} \setminus I_\varepsilon$,*

$$|\varphi_X^{\mathcal{F}}(\boldsymbol{t})| = n^{-\omega_{\lambda, \ell, \varepsilon, \eta}(1)}.$$

Lemma 3.3. *With $\mathcal{H}, n, \mathcal{F}, \varepsilon, I_\varepsilon$ as above, we have*

$$\int_{\mathbb{R}^{\mathcal{H}} \setminus I_\varepsilon} |\varphi_Z^{\mathcal{F}}(\boldsymbol{t})| \, d^k \boldsymbol{t} = n^{-\omega_{\lambda, \ell, \varepsilon, \eta}(1)}.$$

3.2. **Local smoothness of transformed normals.** Lemma 3.3 has nothing to do with graph theory, so we dispense with it here. In this section, we normalize the Fourier transform using the convention $\hat{f}(\boldsymbol{t}) = \int_{\mathbb{R}^d} e^{i \boldsymbol{t} \cdot \boldsymbol{x}} f(\boldsymbol{x}) \, d^d\boldsymbol{x}$.

Definition 3.4. Let d be a positive integer and let $a > 0$ be a real number. A (d, a)-*quasishear* is a map $\Psi \colon \mathbb{R}^d \to \mathbb{R}^d$ given by

$$\Psi(x_1, \ldots, x_d) = (x_1 + \psi_1, x_2 + \psi_2(x_1), x_3 + \psi_3(x_1, x_2), \cdots, x_d + \psi_d(x_1, \ldots, x_{d-1})),$$

where $|\psi_1| \leq a$ and ψ_i, for $2 \leq i \leq d$, are polynomials of degree at most a and all coefficients bounded by a.

Proposition 3.5. *The inverse of a (d, a)-quasishear is a $(d, O_{d,a}(1))$-quasishear.*

Proof. Given a (d, a)-quasishear,

$$(x_1, \ldots, x_d) \mapsto (x_1 + \psi_1, x_2 + \psi_2(x_1), x_3 + \psi_3(x_1, x_2), \cdots, x_d + \psi_d(x_1, \ldots, x_{d-1}))$$

its inverse can be defined as $(x_1, \ldots, x_d) \mapsto (y_1, \ldots, y_d)$, where the y_i are defined recursively by letting $y_1 = x_1 - \psi_1$ and $y_i = x_i - \psi_i(y_1, \ldots, y_{i-1})$ for $2 \leq i \leq d$. Observe that each y_i can be written as a finite expression in terms of the x_i and ψ_i, so the degrees and coefficients must remain bounded. $\square$

Lemma 3.6. *Let $\mathcal{N} \colon \mathbb{R}^d \to \mathbb{R}$, given by $\mathcal{N}(\boldsymbol{x}) = (2\pi)^{-d/2} e^{-\|\boldsymbol{x}\|_2^2/2}$, be the standard multivariate normal probability density. Then for any (d, a)-quasishear Ψ,*

$$|\widehat{\mathcal{N} \circ \Psi}(\boldsymbol{t})| = \|\boldsymbol{t}\|_\infty^{-\omega_{d,a}(1)},$$

where the asymptotic is taken in the limit $\|\boldsymbol{t}\|_\infty \to \infty$.

Proof. Let r be a positive integer. Then for every index $j \in [d]$, we have

$$|t_j|^r |\widehat{\mathcal{N} \circ \Psi}(\boldsymbol{t})| = |(\partial_j^r (\mathcal{N} \circ \Psi))^{\wedge}(\boldsymbol{t})| \leq \|\partial_j^r (\mathcal{N} \circ \Psi)\|_1.$$

We now claim that $\partial_j^r (\mathcal{N} \circ \Psi)(\boldsymbol{x}) = (\mathcal{N} \circ \Psi)(\boldsymbol{x}) P(\boldsymbol{x})$ for some polynomial $P(\boldsymbol{x})$ of degree and coefficients that are bounded in terms of d, a, and r. This can be proven by induction on r. The $r = 0$ case is trivial, and we note that

$$\partial_j((\mathcal{N} \circ \Psi)(\boldsymbol{x}) P(\boldsymbol{x})) = (\mathcal{N} \circ \Psi)(\boldsymbol{x})(-(\Psi(\boldsymbol{x}) \cdot \partial_j \Psi(\boldsymbol{x})) P(\boldsymbol{x}) + \partial_j P(\boldsymbol{x})),$$

where we have used the fact that $\nabla \mathcal{N}(\boldsymbol{x}) = -\boldsymbol{x} \mathcal{N}(\boldsymbol{x})$. As a result, we have

$$\|\partial_j^r (\mathcal{N} \circ \Psi)\|_1 = \int_{\mathbb{R}^d} |P(\boldsymbol{x})| \mathcal{N}(\Psi(\boldsymbol{x})) \, d^d\boldsymbol{x} = \int_{\mathbb{R}^d} |P(\Psi^{-1}(\boldsymbol{x}))| \mathcal{N}(\boldsymbol{x}) \, d^d\boldsymbol{x}.$$

Proposition 3.5 implies that $P(\Psi^{-1}(\boldsymbol{x}))$ is also a polynomial with bounded degree and coefficients, so this entire integral is bounded. Since j was arbitrary, we conclude that $|\widehat{\mathcal{N} \circ \Psi}(\boldsymbol{x})| \lesssim_{d,a,r} \|\boldsymbol{t}\|_\infty^{-r}$. Since r was arbitrary, we are done. $\square$

Proof of Lemma 3.3. Let Ψ be the polynomial map sending

$$\left(\frac{\gamma_H}{\sigma_H}\right)_{H \in \mathcal{H}} \mapsto \left(\frac{F_H}{(p(1-p))^{e(H)/2} \sigma_H}\right)_{H \in \mathcal{H}}.$$

By condition (2) of Definition 2.12 and the fact that $\sigma_H \asymp_\ell n^{v(H)/2}$, we find that Ψ is an $(|\mathcal{H}|, O_{\lambda,\ell}(\eta))$-quasishear.

Now, essentially by definition, $\varphi_Z^{\mathcal{F}}(\boldsymbol{t}) = \widehat{\mathcal{N} \circ \Psi^{-1}}(\boldsymbol{t}')$, where $\boldsymbol{t}' = (t'_H)_{H \in \mathcal{H}}$ is given by

$$t'_H = (p(1-p))^{e(H)/2} \sigma_H t_H.$$

Thus, by Lemma 3.6, we conclude that there is a function g, depending on λ, ℓ, and η, that decays superpolynomially and such that

$$\varphi_Z^{\mathcal{F}}(\boldsymbol{t}) \leq g\left(\max_{H \in \mathcal{H}} |t_H| n^{v(H)/2}\right).$$

But then, changing to the integration variable $\tau = \max_{H \in \mathcal{H}} |t_H| n^{v(H)/2}$, we have

$$\int_{\mathbb{R}^{\mathcal{H}} \setminus I_\varepsilon} |\varphi_Z^{\mathcal{F}}(\boldsymbol{t})| \, d^k \boldsymbol{t} \lesssim_\ell \prod_{H \in \mathcal{H}} n^{-v(H)/2} \cdot \int_{n^\varepsilon}^\infty \tau^{|\mathcal{H}|-1} g(\tau) \, d\tau.$$

But this is $n^{-\omega_{\lambda,\ell,\varepsilon,\eta}(1)}$, so we are done. $\qquad\square$

3.3. **Estimates on Boolean functions.** The proofs of Lemmas 3.1 and 3.2 will need a few results from the analysis of Boolean functions, which we state here. First, we will frequently use hypercontractivity, in the following two forms:

Lemma 3.7 (Moment hypercontractivity [15, Thm. 10.21])**.** *Let $p \in (\lambda, 1 - \lambda)$ and let f be a polynomial of degree at most D evaluated on independent p-biased random bits. Then for $q > 2$,*

$$\mathbb{E}[|f|^q]^{1/q} \leq (\sqrt{q-1} \cdot \lambda^{1/q-1/2})^D \cdot \mathbb{E}[|f|^2]^{1/2}.$$

Lemma 3.8 (Concentration hypercontractivity [15, Thm. 10.24])**.** *If λ, D, and nonzero f are as above, then for any $t \geq (2e/\lambda)^{D/2}$, then*

$$\mathbb{P}[|f| \geq t\mathbb{E}[f^2]^{1/2}] \leq \lambda^D \exp\left(-\frac{D}{2e}\lambda t^{2/D}\right).$$

The only context in which we will use the latter result is through the following corollary.

Corollary 3.9. *If λ, D, and nonzero f are as above, then*

$$\mathbb{P}[|f| \geq n^\varepsilon \mathbb{E}[f^2]^{1/2}] = n^{-\omega_{D,\lambda,\varepsilon}(1)}.$$

Finally, we have the following simple estimate on linear functions.

Proposition 3.10. *Let $p \in (\lambda, 1 - \lambda)$ and let $\boldsymbol{x} = (x_i)_{i \in I}$ be a vector of independent p-biased random bits. If $a \in \mathbb{R}^I$ satisfies $|a_i| \leq 1.2\pi$ for all $i \in I$, then*

$$|\mathbb{E}[e^{i\boldsymbol{a}\cdot\boldsymbol{x}}]| \leq e^{-\Omega_\lambda(\|\boldsymbol{a}\|_2^2)}.$$

Proof. As the x_i are independent, it suffices to show this result for $|I| = 1$, i.e. $|(1-p) + pe^{ia}| \leq e^{-\Omega_\lambda(a^2)}$. This is true since

$$|(1-p) + pe^{ia}|^2 = (1-p)^2 + p^2 + 2p(1-p)\cos a$$

$$= 1 - 2p(1-p)(1 - \cos a) \leq 1 - \frac{\lambda(1-\lambda)a^2}{5} \leq e^{-\lambda(1-\lambda)a^2/5},$$

where we have used the fact that $p(1-p) \geq \lambda(1-\lambda)$ and $1 - \cos a \geq a^2/10$ for $|a| \leq 1.2\pi$. $\qquad\square$

Combining the previous results, we obtain the following result, which appears as [1, Theorem 3] and also occurs in [18, Section 3.4.6], albeit both times with significantly different language. We include a proof for completeness.

Lemma 3.11. *Let p, λ, I, x_i be as in Proposition 3.10. For $i \in I$ let $\chi_i = (x_i - p)/\sqrt{p(1-p)}$ and for $S \subseteq I$ let $\chi_S = \prod_{i \in S} \chi_i$. For all $i \in I$, let δ_i be a real number. Let $\mathcal{S}$ be a collection of subsets of I of size at least 2 and for all $S \in \mathcal{S}$ let δ_S be a real number. Suppose n is a positive integer and $\varepsilon > 0$ such that*

(1) $|I| \leq n^{1/\varepsilon}$;
(2) $|S| \leq 1/\varepsilon$ for all $S \in \mathcal{S}$;
(3) $|\delta_i| \leq 1.2\pi\sqrt{p(1-p)}$ for all $i \in I$;
(4) $\sum_{i \in I} \delta_i^2 \geq n^\varepsilon$;
(5) $\sum_{S \in \mathcal{S}} \delta_S^2 \leq n^{-\varepsilon}$.

Then

$$\left| \mathbb{E}\left[e^{i\left(\sum_{i \in I} \delta_i \chi_i + \sum_{S \in \mathcal{S}} \delta_S \chi_S\right)} \right] \right| = n^{-\omega_{\lambda,\varepsilon}(1)}.$$

Proof. Let $L = \sum_{i \in I} \delta_i \chi_i$ and $U = \sum_{S \in \mathcal{S}} \delta_S \chi_S$. Let r be a positive integer. By Taylor approximation, we have

$$\left| e^{iU} - \sum_{j=0}^{r-1} \frac{(iU)^j}{j!} \right| \leq \frac{2|U|^r}{r!},$$

so
(3.1)

$$\left| \mathbb{E}[e^{i(L+U)}] \right| = \left| \mathbb{E}\left[\sum_{j=0}^{r-1} \frac{e^{iL}(iU)^j}{j!} \right] + \mathbb{E}\left[e^{iL}\left(e^{iU} - \sum_{j=0}^{r-1} \frac{(iU)^j}{j!} \right) \right] \right| \leq \sum_{M \in \mathcal{M}} \left| \mathbb{E}\left[a_M M e^{iL} \right] \right| + \mathbb{E}\left[\frac{2|U|^r}{r!} \right],$$

where $\mathcal{M}$ is the set of monomials (in the χ_i) occurring in $\sum_{j=0}^{r-1} \frac{(iU)^j}{j!}$ and

$$\sum_{M \in \mathcal{M}} a_M M = \sum_{j=0}^{r-1} \frac{(iU)^j}{j!}.$$

Observe that by condition (2), each $M \in \mathcal{M}$ involves at most r/ε variables. For a given M, if we let $I' \subseteq I$ index the variables not in M, by conditions (3) and (4) we find that $\sum_{i \in I'} \delta_i^2 = \Omega_{\varepsilon,r}(n^\varepsilon)$. Applying Proposition 3.10 (and condition (3)), we find that

$$\left| \mathbb{E}\left[e^{\sum_{i \in I'} \delta_i \chi_i} \right] \right| = n^{-\omega_{\lambda,\varepsilon,r}(1)},$$

so as a result, since M is bounded in terms of λ, ε, r, we have

$$\left| \mathbb{E}\left[a_M M e^{iL} \right] \right| \leq |a_M| n^{-\omega_{\lambda,\varepsilon,r}(1)}.$$

To crudely bound $\sum_{M \in \mathcal{M}} |a_M|$, we note that by condition (5) we have $|\delta_S| \leq 1$ for all $S \in \mathcal{S}$. Also, by conditions (1) and (2) we find that $|\mathcal{S}| \leq n^{1/\varepsilon^2}$. Therefore, if we imagine expanding out $\sum_{j=0}^{r-1} \frac{(iU)^j}{j!}$, we have at most $\sum_{j=0}^{r-1} n^{j/\varepsilon^2} \leq r n^{r/\varepsilon^2}$ terms to collect, so $\sum_{M \in \mathcal{M}} |a_M| \leq r n^{r/\varepsilon^2}$. As a result,

(3.2)
$$\sum_{M \in \mathcal{M}} \left| \mathbb{E}\left[a_M M e^{iL} \right] \right| \leq r n^{r/\varepsilon^2} n^{-\omega_{\lambda,\varepsilon,r}(1)} = n^{-\omega_{\lambda,\varepsilon,r}(1)}.$$

Finally, we observe that by Lemma 3.7, we have

(3.3)
$$\mathbb{E}\left[\frac{2|U|^r}{r!} \right] \lesssim_{\lambda,\varepsilon,r} \mathbb{E}[U^2]^{r/2} \leq n^{-r\varepsilon/2}.$$

Combining (3.1), (3.2), and (3.3), we find that for all r,

$$\left| \mathbb{E}\left[e^{i(L+U)} \right] \right| = n^{-\omega_{\lambda,\varepsilon,r}(1)} + O_{\lambda,\varepsilon,r}(n^{-r\varepsilon/2}).$$

Since r was arbitrary, we are done. $\qquad\square$

4. Distributional Result

Lemma 3.1 is proven through a multivariate version of Stein's method of exchangeable pairs, which we now state.

For real matrices define the *Hilbert-Schmidt norm* $\|A\|_{\mathrm{HS}} = \sqrt{\mathrm{tr}(A^T A)}$ and the *operator norm* $\|A\|_{\mathrm{op}} = \max_{|v|=1} \|Av\|$. Also, for $f \in C^k(\mathbb{R}^d)$ define

$$M_k(f) = \sup_{\substack{x, u_1, \ldots, u_k \in \mathbb{R}^d \\ |u_i|=1}} |\partial_{u_1} \cdots \partial_{u_k} f(x)|.$$

Moreover, call two random variables Y and Y' *exchangeable* if (Y, Y') and (Y', Y) have the same distribution.

Theorem 4.1 ([12]). *Let (Y, Y') be an exchangeable pair of random vectors in $\mathbb{R}^d$. Suppose that there is an invertible matrix Λ such that $\mathbb{E}[Y' - Y \mid Y] = -\Lambda Y$ and a random matrix E such that $\mathbb{E}[E \mid Y] = \mathbb{E}[(Y' - Y)(Y' - Y)^T - 2\Lambda \mid Y]$. Then for $g \in C^3(\mathbb{R}^d)$,*

$$\left| \mathbb{E}[g(Y)] - \mathbb{E}[g(Z)] \right| \leq \|\Lambda^{-1}\|_{\mathrm{op}} \left(\frac{\sqrt{d}}{4} M_2(g) \mathbb{E}[\|E\|_{\mathrm{HS}}] + \frac{1}{9} M_3(g) \mathbb{E}[|Y' - Y|^3] \right),$$

where Z is a standard multivariate normal on $\mathbb{R}^d$.

The following statement is our main distributional result.

Theorem 4.2. *Let $p \in (\lambda, 1 - \lambda)$ and $\mathcal{H} = \{H_1, \ldots, H_k\}$ be set of nonisomorphic non-• connected graphs. Letting $\sigma_H^2 = \mathrm{Var}[\gamma_H]$ and $Z_1, \ldots, Z_k$ be a collection of independent standard normals, then for any $\psi \in C^3(\mathbb{R}^k)$ we have*

$$\left| \mathbb{E}[\psi(\gamma_{H_1}/\sigma_{H_1}, \ldots, \gamma_{H_k}/\sigma_{H_k})] - \mathbb{E}[\psi(Z_1, \ldots, Z_k)] \right| \lesssim_{\mathcal{H}, \lambda} \frac{M_2(\psi) + M_3(\psi)}{n^{1/2}}.$$

Proof. Aiming to apply Theorem 4.1, our setup is as follows: let $X = (x_e)_{e \in \binom{[n]}{2}}$ be the indicator variables of the edges of $G(n, p)$. Sample I from $\binom{[n]}{2}$ uniformly at random, and let $X' = (x'_e)_{e \in \binom{[n]}{2}}$ be X but with x_I resampled. Then, let

$$Y = \left(\frac{\gamma_{H_1}(X)}{\sigma_{H_1}}, \ldots, \frac{\gamma_{H_k}(X)}{\sigma_{H_k}} \right) \quad \text{and} \quad Y' = \left(\frac{\gamma_{H_1}(X')}{\sigma_{H_1}}, \ldots, \frac{\gamma_{H_k}(X')}{\sigma_{H_k}} \right).$$

Since X and X' are exchangeable, so are Y and Y'. Also, conditioning on X, every contribution to $\gamma_{H_i}(X)/\sigma_{H_i}$ has a $e(H_i)/\binom{n}{2}$ chance of having one of its edges resampled, which leads to a zero contribution in $\mathbb{E}[\gamma_{H_i}(X')/\sigma_{H_i} \mid X, I]$. Thus, $\mathbb{E}[X' - X \mid X] = -\Lambda X$ where $\Lambda = \mathrm{diag}(e(H_i)/\binom{n}{2})_{i \in [k]}$. Finally, we define

$$E = \mathbb{E}[(Y' - Y)(Y' - Y)^T - 2\Lambda \mid X].$$

It thus remains to bound $\mathbb{E}[|Y' - Y|^3], \mathbb{E}[\|E\|_{\mathrm{HS}}] \lesssim_{\mathcal{H}, \lambda} n^{-5/2}$.

First, we observe

$$\mathbb{E}[|Y' - Y|^3] \leq \mathbb{E}[|Y' - Y|^4]^{3/4} \leq \left(k \sum_{i \in [k]} \frac{\mathbb{E}[(\gamma_{H_i}(X') - \gamma_{H_i}(X))^4]}{\sigma_{H_i}^4} \right)^{3/4}.$$

By symmetry, we may fix an I in the above expectation. Then, $\gamma_{H_i}(X') - \gamma_{H_i}(X)$ consists of $O_{\mathcal{H}}(n^{v(H_i)-2})$ terms and thus has variance $O_{\mathcal{H}}(n^{v(H_i)-2})$. By Lemma 3.7, it follows that $\mathbb{E}[(\gamma_{H_i}(X') - \gamma_{H_i}(X))^4] \lesssim_{\mathcal{H}, \lambda} n^{2v(H_i)-4}$, and the bound follows since $\sigma_{H_i} \asymp_{\mathcal{H}} n^{v(H_i)/2}$.

We now deal with E. Defining χ_e, χ'_I similarly to the introduction and abbreviating $\prod_{e \in S} \chi_e$ as χ_S, we have

$$
\mathbb{E}[(Y'_i - Y_i)(Y'_j - Y_j) \mid X, I]
$$

$$
= \frac{1}{\sigma_{H_i} \sigma_{H_j}} \sum_{\substack{H'_1 \cong H_i,\, H'_2 \cong H_j \\ I \in E(H'_1), E(H'_2)}} \mathbb{E}\Big[(\chi_{E(H'_1) \setminus I} \chi'_I - \chi_{E(H'_1)})(\chi_{E(H'_2) \setminus I} \chi'_I - \chi_{E(H'_2)}) \,\Big|\, X, I\Big]
$$

$$
= \frac{1}{\sigma_{H_i} \sigma_{H_j}} \sum_{\substack{H'_1 \cong H_i,\, H'_2 \cong H_j \\ I \in E(H'_1), E(H'_2)}} (\chi_{E(H'_1) \setminus I} \chi_{E(H'_2) \setminus I} + \chi_{E(H'_1)} \chi_{E(H'_2)}),
$$

which implies that

$$
E_{ij} = \frac{1}{\sigma_{H_i} \sigma_{H_j} \binom{n}{2}} \sum_{\substack{H'_1 \cong H_i,\, H'_2 \cong H_j \\ e \in E(H'_1), E(H'_2)}} (\chi_{E(H'_1) \setminus e} \chi_{E(H'_2) \setminus e} + \chi_{E(H'_1)} \chi_{E(H'_2)}) - \frac{2e(H_i)}{\binom{n}{2}} \delta_{ij},
$$

where δ_{ij} is the Kronecker delta. At this point, it is easy to see that $\mathbb{E}[E_{ij}] = 0$.[5] For future convenience, let E'_{ij} be the sum in the above expression.

We may now bound

$$
\mathbb{E}[\|E\|_{\mathrm{HS}}] \le \mathbb{E}[\|E\|_{\mathrm{HS}}^2]^{1/2} = \Big(\sum_{i,j} \mathbb{E}[E_{ij}^2]\Big)^{1/2} = \Bigg(\sum_{i,j} \frac{\mathrm{Var}[E'_{ij}]}{\sigma_{H_i}^2 \sigma_{H_j}^2 \binom{n}{2}^2}\Bigg)^{1/2},
$$

which means that we need to show that $\mathrm{Var}[E'_{ij}] \lesssim_{\mathcal{H}, \lambda} n^{v(H_i) + v(H_j) - 1}$. To do this, we expand out

$$
\mathrm{Var}[E'_{ij}] = \sum_{\substack{H'_1, H'_3 \cong H_i \\ H'_2, H'_4 \cong H_j}} \sum_{\substack{e_1 \in E(H'_1), E(H'_2) \\ e_2 \in E(H'_3), E(H'_4)}} \mathrm{Cov}[\chi_{E(H'_1) \setminus e_1} \chi_{E(H'_2) \setminus e_1} + \chi_{E(H'_1)} \chi_{E(H'_2)},
$$
$$
\chi_{E(H'_3) \setminus e_2} \chi_{E(H'_4) \setminus e_2} + \chi_{E(H'_3)} \chi_{E(H'_4)}]
$$

and claim that all but $O_{\mathcal{H}}(n^{v(H_i) + v(H_j) - 1})$ terms are zero, which finishes as each covariance is $O_{\mathcal{H}, \lambda}(1)$.

If an edge appears exactly once in $E(H'_1), E(H'_2), E(H'_3), E(H'_4)$, the respective covariance vanishes. Thus, since H_i, H_j have no isolated vertices, each vertex must appear in $V(H'_1), V(H'_2), V(H'_3), V(H'_4)$ at least twice. If a vertex appears more than twice, then the number of total vertices is at most $v(H_i) + v(H_j) - 1$, which corresponds to $O_{\mathcal{H}}(n^{v(H_i) + v(H_j) - 1})$ terms. Thus we may assume that every vertex is covered exactly twice, and every edge at least twice.

$V(H'_1)$ must intersect $V(H'_2)$, as $E(H'_1) \cap E(H'_2) \ne \varnothing$. If $V(H'_1)$ also intersects $V(H'_3)$ or $V(H'_4)$, then since each vertex is covered exactly twice we may partition $V(H'_1)$ into multiple parts based on which other $V(H'_i)$ it is in. Then by the connectedness of H_i there must exist an edge in $E(H'_1)$ joining two of $V(H'_2), V(H'_3), V(H'_4)$, which cannot be covered by any other edge set and thus leads to a zero covariance. Applying similar logic to $V(H'_2)$ implies that $V(H'_1) = V(H'_2)$, which in turn implies that $V(H'_3) = V(H'_4)$. However, in this case, the two variables in the covariance are independent, so we are done. $\qquad\square$

Proof of Lemma 3.1. Let $\psi(z_H \colon H \in \mathcal{H}) = e^{i \sum_H t_H F_H(\sigma_{H'} z_{H'} \,:\, H' \in \mathcal{H})}$ and let $\kappa \colon [0, \infty) \to [0, 1]$ be a smooth decreasing cutoff function that is 1 on $[0, 1]$ and 0 on $[2, \infty)$. Furthermore, let $\psi'(\boldsymbol{z}) = \psi(\boldsymbol{z}) \kappa(\|\boldsymbol{z}\|_2 / n^{\varepsilon'})$, where $\varepsilon' > 0$ is to be determined later.

---

[5]This also follows straightforwardly from the definition of E, exchangeability, and the fact that $\mathbb{E}[YY^T]$ is the identity.

Let $Y = (\gamma_H/\sigma_H)_{H \in \mathcal{H}}$ and Z be an $\mathcal{H}$-indexed vector of independent standard normals. Then the quantities $\varphi_X^{\mathcal{F}}(\boldsymbol{t})$ and $\varphi_Z^{\mathcal{H}}(\boldsymbol{t})$ are precisely $\mathbb{E}[\psi(Y)]$ and $\mathbb{E}[\psi(Z)]$. By Corollary 3.9 and direct computation, respectively, we have $\|Y\|_2, \|Z\|_2 \leq n^{\varepsilon'}$ with probability $1 - n^{-\omega_{\lambda,\ell,\varepsilon'}(1)}$, so since $|\psi'(\boldsymbol{z}) - \psi(\boldsymbol{z})| \leq 1$ always we have

$$|\varphi_X^{\mathcal{F}}(\boldsymbol{t}) - \varphi_Z^{\mathcal{F}}(\boldsymbol{t})| = |\mathbb{E}[\psi(Y) - \psi(Z)]| \leq |\mathbb{E}[\psi'(Y) - \psi'(Z)]| + n^{-\omega_{\lambda,\ell,\varepsilon'}(1)}.$$

After applying Theorem 4.2 to ψ', it suffices to show that $M_2(\psi'), M_3(\psi') \lesssim_{\ell,\varepsilon,\eta} n^{4\varepsilon}$. To see this, observe that condition (2) of Definition 2.12 implies that $F_H(\sigma_H z_H \colon H \in \mathcal{H})/n^{v(H)/2}$ is a polynomial in $\boldsymbol{z}$ of bounded degree and coefficients bounded by $O_\ell(\eta)$. Thus, $\psi(\boldsymbol{z}) = e^{if(\boldsymbol{z})}$ for some polynomial f of bounded degree and coefficients bounded by $O_\ell(\eta \max_{H \in \mathcal{H}} |t_H| n^{v(H)/2}) \leq O_\ell(\eta n^{\varepsilon})$. Since taking k partial derivatives of e^{if} causes at most k factors of f and its derivatives to appear in front of the exponential, we conclude that $|\partial_{u_1} \cdots \partial_{u_k} e^{if}| \lesssim_{\ell,k} (1 + \eta n^{\varepsilon})^k (1 + \|\boldsymbol{z}\|_2)^{O_{\ell,k}(1)}$ for unit vectors $u_1, \ldots, u_k$.

Since ψ' is identically zero for $\|\boldsymbol{z}\|_2 \geq 2n^{\varepsilon'}$ its derivatives are certainly zero. Since κ and its derivatives are bounded at other $\boldsymbol{z}$, we conclude that $M_k(\psi') \lesssim_{\ell,k} (1 + \eta n^{\varepsilon})^k n^{O_{\ell,k}(\varepsilon')}$. By choosing ε' to be sufficiently small, we therefore can get $M_2(\psi'), M_3(\psi') \lesssim_{\ell,\varepsilon,\eta} n^{4\varepsilon}$, as desired. $\square$

5. Decoupling Methods

In this section, we will prove Lemma 3.2, which will complete the proof of Theorem 1.2. Conceptually, this is quite simple: in Section 5.1 we state a decoupling inequality, which we use to transform $\varphi_X^{\mathcal{F}}(\boldsymbol{t})$ into an expression that can be bounded using Lemma 3.11. However, the precise details in this procedure are quite technical, and the bulk of the work consists of showing that regardless of the value of $\boldsymbol{t}$, there is always exists a decoupling setup that will be effective. To accomplish this, in Section 5.2 we develop a list of conditions (Lemma 5.2) required for decoupling to work in the case of random graphs. In Sections 5.3 and 5.4 we identify two families of decoupling setups that satisfy the conditions in Lemma 5.2, which are subsequently applied in Section 5.5.

5.1. **The decoupling inequality.** Let k be a nonnegative integer (which we will call the *decoupling parameter*). Following [1], for a function $f \colon \Omega_0 \times \prod_{i=1}^k \Omega_k \to \mathbb{R}$, where $\Omega_0, \Omega_1, \ldots, \Omega_k$ are arbitrary sets, define $\alpha(f) \colon \Omega_0 \times \prod_{i=1}^k \Omega_k^2 \to \mathbb{R}$ given by

$$\alpha(f)(x, y_1^0, y_1^1, \ldots, y_k^0, y_k^1) = \sum_{(i_1, \ldots, i_k) \in \{0,1\}^k} (-1)^{i_1 + \cdots + i_k} f(x, y_1^{i_1}, \ldots, y_k^{i_k}).$$

We then have the following result, which we prove for completeness.

Lemma 5.1 ([1])**.** *Suppose $X, Y_1, Y_2, \ldots, Y_k$ are random variables on $\Omega_0, \Omega_1, \Omega_2, \ldots, \Omega_k$, respectively. If for $i \in [k]$ we let Y_i' be an independent copy of Y_i, then*

$$\left| \mathbb{E}_{X,Y}[e^{if(X,Y)}] \right|^{2^k} \leq \mathbb{E}_{\mathbf{Y}}\left[\left| \mathbb{E}_X[e^{i\alpha(f)(X,\mathbf{Y})}] \right| \right],$$

where we abbreviate $Y = (Y_1, \ldots, Y_k)$ and $\mathbf{Y} = (Y_1, Y_1', \ldots, Y_k, Y_k')$.

Proof. We proceed with induction on k, with the $k = 0$ case being trivial. Now assume the result is true for $k - 1$ and let $\tilde{Y} = (Y_1, \ldots, Y_{k-1})$ and $\tilde{\mathbf{Y}} = (Y_1, Y_1', \ldots, Y_{k-1}, Y_{k-1}')$. By Cauchy-Schwarz we have

$$\left| \mathbb{E}_{X,Y}[e^{if(X,Y)}] \right|^{2^k} \leq \left| \mathbb{E}_{X,\tilde{Y}}\left[\left| \mathbb{E}_{Y_k}[e^{if(X,\tilde{Y},Y_k)}] \right|^2 \right] \right|^{2^{k-1}} = \left| \mathbb{E}_{X,\tilde{Y}}\left[\mathbb{E}_{Y_k,Y_k'}[e^{i(f(X,\tilde{Y},Y_k)-f(X,\tilde{Y},Y_k'))}] \right] \right|^{2^{k-1}}.$$

Define the function

$$\tilde{f}((x, y_k, y_k'), y_1, \ldots, y_{k-1}) = f(x, y_1, \ldots, y_{k-1}, y_k) - f(x, y_1, \ldots, y_{k-1}, y_k'),$$

and observe that $\alpha(\tilde{f})((X, Y_k, Y'_k), \tilde{\mathbf{Y}}) = \alpha(f)(X, \mathbf{Y})$. Applying the inductive hypothesis to $\tilde{f}$, the quantity is therefore bounded by

$$\mathbb{E}_{\tilde{\mathbf{Y}}}\left[\left|\mathbb{E}_{X, Y_k, Y'_k}[e^{i\alpha(f)(X, \mathbf{Y})}]\right|\right] \leq \mathbb{E}_{\mathbf{Y}}\left[\left|\mathbb{E}_X[e^{i\alpha(f)(X, \mathbf{Y})}]\right|\right],$$

as desired. $\qquad\square$

5.2. Decoupling on random graphs.
Let $\mathcal{H}$ be a set of (isomorphism classes of) NIV graphs on at most ℓ vertices. For $H \in \mathcal{H}$, let Δ_H be real constants and define the statistic $f = \sum_{H \in \mathcal{H}} \Delta_H \gamma_H$.

Take a nonnegative integer k and a partition of $\binom{[n]}{2}$ into sets $A, B_1, \ldots, B_k, C$. Sample all the edges with probability p, letting x_e and χ_e denote the indicator variable of an edge e and its normalization, respectively. Furthermore, resample all edges in $\bigcup_i B_i$, defining x'_e and χ'_e similarly. Let $X = (x_e)_{e \in A}$, $Y_i = (x_e)_{e \in B_i}$, $Y'_i = (x'_e)_{e \in B_i}$, and $W = (x_e)_{e \in C}$. As f is naturally a function of (X, W) and $Y_1, \ldots, Y_k$, we may apply Lemma 5.1 to get

$$(5.1) \qquad |\mathbb{E}_{G(n,p)}[e^{if}]|^{2^k} \leq \mathbb{E}_{\mathbf{Y}}\left[\left|\mathbb{E}_{X,W} e^{i\alpha(f)(X,W,\mathbf{Y})}\right|\right] \leq \mathbb{E}_{\mathbf{Y},W}\left[\left|\mathbb{E}_X e^{i\alpha(f)(X,W,\mathbf{Y})}\right|\right].$$

For a set $S \subseteq A$, let δ_S be the function of $W, \mathbf{Y}$ such that $\alpha(f)(X, W, \mathbf{Y}) = \sum_S \delta_S \chi_S$. (Abbreviate $\delta_e = \delta_{\{e\}}$.) For a set of edges T disjoint from A, define

$$\zeta_T = \prod_{e \in T \cap C} \chi_e \cdot \prod_{i \in [k]} \left(\prod_{e \in T \cap B_i} \chi_e - \prod_{e \in T \cap B_i} \chi'_e \right).$$

Call a set of edges, and by extension a subgraph, *rainbow* if it intersects B_i for all i. For a graph H and a subset $S \subseteq A$, let $N_{H,S}$ be the number of rainbow copies H' of H such that $E(H') \cap A = S$.

We make the following easy observations:

(1) Each ζ_T is a polynomial in the x_e of degree at most $|T|$.
(2) If T is not rainbow, $\zeta_T = 0$.
(3) For rainbow T and T', we have $\mathbb{E}[\zeta_T \zeta_{T'}] = 2^k \cdot \mathbf{1}_{T=T'}$.
(4) If T_1, T_2, T_3, T_4 are such that there exists an edge contained within exactly one T_i, then $\mathbb{E}[\zeta_{T_1} \zeta_{T_2} \zeta_{T_3} \zeta_{T_4}] = 0$.
(5) We have

$$\delta_S = \sum_{H \in \mathcal{H}} \Delta_H \sum_{\substack{H' \cong H \\ E(H') \cap A = S}} \zeta_{E(H') \setminus A}.$$

(6) As a consequence of (2), (3), and (5), we have $\mathbb{E}[\delta_S^2] = 2^k \sum_{H \in \mathcal{H}} \Delta_H^2 N_{H,S}$. (Note that this deduction requires all $H \in \mathcal{H}$ to have no isolated vertices, since otherwise two different graphs could have the same edge sets.)

We may now express our conditions for successful decoupling in terms of the following lemma:

Lemma 5.2. *Suppose there is a nonempty subset $\mathcal{H}' \subseteq \mathcal{H}$ and some $\varepsilon > 0$ such that*

(1) We have one of the following:
  (1a) $\Delta_H^2 N_{H,e} \leq n^{-\varepsilon}$ for all $H \in \mathcal{H}'$ and $e \in A$;
  (1b) For $e \in A$ we always have

$$\left| \sum_{H \in \mathcal{H}'} \Delta_H \sum_{\substack{H' \cong H \\ E(H') \cap A = \{e\}}} \zeta_{E(H') \setminus A} \right| \leq 1.1\pi\sqrt{p(1-p)}.$$

(2) $\sum_{H \in \mathcal{H}'} \sum_{e \in A} \Delta_H^2 N_{H,e} \geq n^{\varepsilon}$;

(3) $\sum_{e \in A} \Delta_H^2 N_{H,e} \leq n^{-\varepsilon}$ for all $H \in \mathcal{H} \setminus \mathcal{H}'$;

(4) $\sum_{|S| \geq 2} \Delta_H^2 N_{H,S} \leq n^{-\varepsilon}$ for all $H \in \mathcal{H}$;

(5) We have

$$\sum_{(H_1, H_2, H_3, H_4) \in Q} \Delta_{H_1} \Delta_{H_2} \Delta_{H_3} \Delta_{H_4} \leq n^{-\varepsilon} \left(\sum_{H \in \mathcal{H}'} \sum_{e \in A} \Delta_H^2 N_{H,e} \right)^2,$$

where Q is the set of quadruples (H_1, H_2, H_3, H_4) of subgraphs such that

(5a) H_1, H_2, H_3, H_4 are isomorphic to elements of $\mathcal{H}'$ and are rainbow (Δ_{H_i} denotes Δ_H where $H \in \mathcal{H}'$ is isomorphic to H_i);

(5b) $E(H_1) \cap A = E(H_2) \cap A = \{e\}$ for some $e \in A$;

(5c) $E(H_3) \cap A = E(H_4) \cap A = \{e'\}$ for some $e' \in A$;

(5d) $(E(H_1) \cup E(H_2)) \cap (E(H_3) \cup E(H_4))$ is nonempty;

(5e) No edge lies in exactly one of the $E(H_i)$.

Then $\mathbb{E}_{G(n,p)}[e^{if}] = n^{-\omega_{\lambda,\ell,\varepsilon}(1)}$.

Proof. Note that we must have $k < \binom{\ell}{2}$, since otherwise we would always have $N_{H,e} = 0$ and condition (2) would be impossible to satisfy. Thus, by (5.1), it suffices to show that

$$\mathbb{E}_{\mathbf{Y},W}\left[\left| \mathbb{E}_X e^{i\alpha(f)(X,W,\mathbf{Y})} \right| \right] = n^{-\omega_{\lambda,\ell,\varepsilon}(1)}.$$

Since $e^{i\alpha(f)(X,W,\mathbf{Y})}$ is bounded, we will show that, with the δ_e and δ_S defined as above, the conditions of Lemma 3.11 (with $I = A$) hold with probability $1 - n^{-\omega_{\lambda,\ell,\varepsilon}(1)}$ (call this *high probability* for the remainder of the proof) for some suitable $\varepsilon'(\lambda, \ell, \varepsilon)$ (it is clear that the $\delta_\varnothing$ term can be ignored). This will finish the proof. To prevent ambiguity, we will let ε' refer to the ε in Lemma 3.11 and denote the conditions of Lemma 3.11 with square brackets.

Conditions [1] and [2] are obvious provided that $\varepsilon' \leq \ell^{-2}$. Moreover, since there are only polynomially many S, by Corollary 3.9 we know that with high probability $\delta_S^2 \leq n^{\varepsilon/2}\mathbb{E}[\delta_S^2]$ for all S. In this case, by condition (4) we have

$$\sum_{|S| \geq 2} \delta_S^2 \leq 2^k n^{\varepsilon/2} \sum_{H \in \mathcal{H}} \sum_{|S| \geq 2} \Delta_H^2 N_{H,S} = O_\ell(n^{-\varepsilon/2}).$$

Thus condition [5] is satisfied provided that $\varepsilon' < \varepsilon/2$.

To deal with conditions [3] and [4], we decompose $\delta_e = \delta_e' + r_e$, where

$$\delta_e' = \sum_{\substack{H \in \mathcal{H}' \\ }} \Delta_H \sum_{\substack{H' \cong H \\ E(H') \cap A = \{e\}}} \zeta_{E(H') \setminus A} \quad \text{and} \quad r_e = \sum_{\substack{H \in \mathcal{H} \setminus \mathcal{H}' \\ }} \Delta_H \sum_{\substack{H' \cong H \\ E(H') \cap A = \{e\}}} \zeta_{E(H') \setminus A}.$$

By condition (3) and Corollary 3.9, we have $|r_e| \leq n^{-\varepsilon/4}$ for all e with high probability. If condition (1a) holds, then we also have $|\delta_e'| \leq n^{-\varepsilon/4}$ for all e with high probability, which yields condition [3]. Condition (1b) simply states that $|\delta_e'| \leq 1.1\pi\sqrt{p(1-p)}$ always, so combining this with our bound on r_e also yields condition [3].

It remains to show condition [4], which is the most involved part of the proof. First of all, by the triangle inequality on $\mathbb{R}^A$ we have

$$\sum_{e \in A} \delta_e^2 \geq \left(\sqrt{\sum_{e \in A} (\delta_e')^2} - \sqrt{\sum_{e \in A} r_e^2} \right)^2.$$

By Corollary 3.9, we have $r_e^2 \leq n^{\varepsilon/2}\mathbb{E}[r_e^2]$ for all e with high probability, and in this case by condition (3) it is true that $\sqrt{\sum_{e\in A}\mathbb{E}[r_e^2]} \lesssim_\ell n^{-\varepsilon/4}$. Thus it suffices to show that $\sum_{e\in A}(\delta_e')^2 \geq n^{\varepsilon'}$ for some $\varepsilon' > 0$ with high probability.

To show this, we apply Corollary 3.9 on $\sum_{e\in A}((\delta_e')^2 - \mathbb{E}[(\delta_e')^2])$; in light of condition (2), for this to work, it suffices to show that

$$
\text{(5.2)} \qquad \text{Var}\left[\sum_{e\in A}(\delta_e')^2\right] \leq n^{-\varepsilon/2}\left(\sum_{e\in A}\mathbb{E}[(\delta_e')^2]\right)^2
$$

for large n.

At this point, we expand

$$
\text{Var}\left[\sum_{e\in A}(\delta_e')^2\right] = \sum_{H_1,H_2,H_3,H_4} \Delta_{H_1}\Delta_{H_2}\Delta_{H_3}\Delta_{H_4}\,\text{Cov}[\zeta_{E(H_1)\setminus A}\zeta_{E(H_2)\setminus A}, \zeta_{E(H_3)\setminus A}\zeta_{E(H_4)\setminus A}]
$$

where we sum over H_1, H_2, H_3, H_4 satisfying conditions (5a) to (5c). If such a quadruple is not in Q, it either violates (5e), in which case

$$
\mathbb{E}[\zeta_{E(H_1)\setminus A}\zeta_{E(H_2)\setminus A}\zeta_{E(H_3)\setminus A}\zeta_{E(H_4)\setminus A}] = \mathbb{E}[\zeta_{E(H_1)\setminus A}\zeta_{E(H_2)\setminus A}]\mathbb{E}[\zeta_{E(H_3)\setminus A}\zeta_{E(H_4)\setminus A}] = 0,
$$

or it violates (5d), in which case $\zeta_{E(H_1)\setminus A}\zeta_{E(H_2)\setminus A}$ and $\zeta_{E(H_3)\setminus A}\zeta_{E(H_4)\setminus A}$ are independent and the entire term is 0. Since the ζ_T are bounded in terms of λ and ℓ, we conclude that

$$
\mathbb{E}\left[\left(\sum_{e\in A}((\delta_e')^2 - \mathbb{E}[(\delta_e')^2])\right)^2\right] \lesssim_{\lambda,\ell} \sum_{(H_1,H_2,H_3,H_4)\in Q} \Delta_{H_1}\Delta_{H_2}\Delta_{H_3}\Delta_{H_4}.
$$

So condition (5) is exactly what we need to conclude (5.2) for large n. This concludes the proof. $\quad\square$

5.3. Decoupling setup I: midscale uniformity.

Lemma 5.3. *Let $0 \leq \alpha, \beta \leq 1 - \varepsilon$ be real parameters and let k be a nonnegative integer. Then $|\mathbb{E}[e^{if}]| = n^{-\omega_{\lambda,\ell,\varepsilon}(1)}$, provided that*

(1) For all H with $k + 2$ vertices, $n^{k+2}\Delta_H^2 \leq n^{2+k\beta-\varepsilon}$;

(2) There is some connected H with $k + 2$ vertices such that $n^{2\alpha+k\beta+\varepsilon} \leq n^{k+2}\Delta_H^2$;

(3) For every H with greater than $k + 2$ vertices we have $n^{v(H)}\Delta_H^2 \leq n^{2\alpha+k\beta-\varepsilon}$.

Proof. We let k be the decoupling parameter. Construct $k + 1$ disjoint vertex sets $V_0, V_1, \ldots, V_k$ such that $|V_0| = \lfloor\frac{1}{k+1}n^{1-\alpha}\rfloor$ and $V_i = \lfloor\frac{1}{k+1}n^{1-\beta}\rfloor$ for $i \in [k]$. Let A be edges within V_0, B_i be edges with one endpoint in V_i and one in V_j for some $0 \leq j \leq i$, and C be all other edges. It suffices to show that the conditions of Lemma 5.2 hold for large n and with ε replaced with $\varepsilon/2$. Similar to before, we denote the conditions of Lemma 5.2 with square brackets to prevent ambiguity.

Observe that any rainbow graph must have at least one vertex in V_i for all $i \in [k]$. Thus

$$
\text{(5.3)} \qquad N_{H,S} \lesssim_\ell n^{v(H)-v(S)-k\beta},
$$

where $v(S)$ denotes the number of vertices adjacent to at least one edge in S. Furthermore, we claim that this is tight if H is connected with $k + 2$ vertices and S is an edge e. To see this, pick any edge e' in H and label the vertices in H not adjacent to e' with the numbers $1, \ldots, k$ such that every vertex is adjacent to a vertex of e' or a vertex with a smaller label. Then, sending e' to e and the vertex labeled i to an arbitrary vertex in V_i yields a rainbow copy of H; moreover, this can be done in $\Omega_{\ell,\varepsilon}(n^{k(1-\beta)})$ ways.

It is now simple to check conditions [1] to [4]. We let $\mathcal{H}'$ consist of graphs in $\mathcal{H}$ with $k + 2$ vertices. Then by condition (1) and (5.3), we find that for all $H \in \mathcal{H}'$ and $e \in A$ we have $\Delta_H^2 N_{H,e} \lesssim_{\ell,\varepsilon} n^{-\varepsilon}$,

implying condition [1a]. By condition (2), there is some $H \in \mathcal{H}'$ with $\Delta_H^2 N_{H,e} \gtrsim_{\ell,\varepsilon} n^{-2(1-\alpha)+\varepsilon}$ for all $e \in A$; summing over the $\Theta_{\ell,\varepsilon}(n^{2(1-\alpha)})$ elements of A yields condition [2].

For all $H \in \mathcal{H} \setminus \mathcal{H}'$, it is either the case that H has fewer than $k+2$ vertices, in which case $N_{H,S} = 0$ for all $S \subseteq A$, or that condition (3) applies, which implies that $\Delta_H^2 N_{H,S} \lesssim_{\ell,\varepsilon} n^{2\alpha-v(S)-\varepsilon}$. For each isomorphism class of the graph associated with S, there are $\Theta_{\ell,\varepsilon}(n^{(1-\alpha)v(S)})$ corresponding subsets of A, so summing over the $O_\ell(1)$ such isomorphism classes yields that $\sum_{|S| \geq 1} \Delta_H^2 N_{H,S} \lesssim_{\ell,\varepsilon} n^{-\varepsilon}$, proving both conditions [3] and [4] for $H \in \mathcal{H} \setminus \mathcal{H}'$. (Here we have used $v(S) \geq 2$.) Condition [4] also holds for $H \in \mathcal{H}'$ as well since for such H we have $N_{H,S} = 0$ whenever $v(S) \geq 3$, which is the case for all S with $|S| \geq 2$.

Finally, to show condition [5], we claim that $|Q| \lesssim_{\ell,\varepsilon} n^{4(1-\alpha)+2k(1-\beta)-\varepsilon}$, which will suffice as we will then have

$$\sum_{(H_1,H_2,H_3,H_4) \in Q} \Delta_{H_1} \Delta_{H_2} \Delta_{H_3} \Delta_{H_4} \lesssim_{\ell,\varepsilon} n^{4(1-\alpha)+2k(1-\beta)-\varepsilon} \max_{H \in \mathcal{H}'} \Delta_H^4,$$

while the tightness of our bound on $N_{H,e}$ for $H \in \mathcal{H}'$ gives

$$\left(\sum_{e \in A} \sum_{H \in \mathcal{H}'} \Delta_H^2 N_{H,e} \right)^2 \gtrsim_{\ell,\varepsilon} n^{4(1-\alpha)+2k(1-\beta)} \max_{H \in \mathcal{H}'} \Delta_H^4.$$

To see this, consider some $(H_1, H_2, H_3, H_4) \in Q$ and let $V = \bigcup_{i \in [4]} V(H_i)$. We know that each H_i must have two vertices in V_0, so $|V \cap V_0| \leq 4$. Also, each H_i must have exactly one vertex in each V_j (for $j \in [k]$), and moreover, if some vertex $v \in V_j$ is a vertex of exactly one H_i, then the edges in $E(H_i) \cap B_j$ cannot be in $H_{i'}$ for any $i' \neq i$. Therefore $|V \cap V_j| \leq 2$.

Furthermore, suppose for the sake of contradiction that equality holds in all of these bounds, i.e. that $|V \cap V_0| = 4$ and $|V \cap V_j| = 2$ for all $j \in [k]$. By condition [5d], the set $V^* = (V(H_1) \cup V(H_2)) \cap (V(H_3) \cup V(H_4))$ is nonempty, and let j be minimal such that $V^* \cap V_j$ is nonempty. We cannot have $j = 0$, as it would contradict conditions [5b] and [5c] and our assumption that $|V \cap V_0| = 4$. Thus, we may assume, after possibly swapping H_3 and H_4, that $V(H_1) \cap V_j = V(H_3) \cap V_j$ consists of one vertex that is not the element of $V(H_2) \cap V_j$. However, since H_1 is rainbow, it contains an edge in B_j, and the only way for that edge to be covered by another H_i is if it is in H_3. But this implies that H_1 and H_3 share a vertex in $V_{j'}$ for some smaller j', contradicting the minimality of j. We conclude that we cannot have $|V \cap V_0| = 4$ and $|V \cap V_j| = 2$ for all $j \in [k]$.

Consequently, the number of choices for V is bounded above by

$$|V_0|^3 \prod_{i \in [k]} |V_i|^2 + \sum_{i \in [k]} \left(|V_0|^4 |V_i| \prod_{j \in [k] \setminus \{i\}} |V_j|^2 \right) \lesssim_\ell n^{4(1-\alpha)+2k(1-\beta)}(n^{\alpha-1} + n^{\beta-1}) \lesssim n^{4(1-\alpha)+2k(1-\beta)-\varepsilon}.$$

Since V is of bounded size, there are $O_\ell(1)$ many elements of Q with a given V, so we have proved the desired bound. $\qquad\square$

5.4. Decoupling setup II: hyperlocal uniformity.

Lemma 5.4. *Let $H \in \mathcal{H}$ be connected and let $\varepsilon \leq \alpha \leq 1$ be real. We get $|\mathbb{E}[e^{if}]| = n^{-\omega_{\lambda,\ell,\varepsilon}(1)}$ as long as*

(1) $n^{-\alpha+\varepsilon} \leq \Delta_H^2 \leq 1.1\pi^2(p(1-p))^{e(H)}$;

(2) For any graph $H' \succ H$, we have $n^{v(H')} \Delta_{H'}^2 < n^{v(H)-\alpha-\varepsilon}$.

Proof. We let $k = e(H) - 1$ and arbitrarily label the edges of H as $e_0, \ldots, e_k$. Consider $\lfloor n^\alpha/v(H) \rfloor$ vertex-disjoint copies of H (call these *standard copies*) and let A be the set of all copies of e_0, while for all $i \in [k]$ let B_i be the set of all copies of e_i. Place all other edges in C. We will let $\mathcal{H}' = \{H\}$.

Call a rainbow graph *super-rainbow* if it has nontrivial intersection with A. Observe that every super-rainbow graph H' must satisfy $H' \succeq H$ (by "gluing together" the standard copies). Moreover, we claim that a super-rainbow graph H' containing edges from a standard copies of H has at least $v(H) + a - 1$ vertices. This can be proven by induction; the base case $a = 1$ is clear, and if $a \geq 2$ we may first assume without loss of generality that $E(H') \cap C = \varnothing$. Then, by the connectedness of H, there are two standard copies of H and two corresponding vertices in those copies that are both in $V(H')$. After moving all edges in one of these copies to the other, which necessarily decreases the vertex count by at least 1, we may apply the inductive hypothesis. One immediate consequence of this claim are that the standard copies are the only super-rainbow copies of H. A less-immediate consequence is that for any $H' \in \mathcal{H}$, the number of super-rainbow copies of H' is $O_\ell(n^{v(H')-v(H)+\alpha})$. To see this, suppose such a copy has edges in a standard copies, which are incident to $b \geq v(H) + a - 1$ vertices. Then, each copy can be produced by choosing edges in the standard copies, which can be done in $O_\ell(n^{\alpha a})$ ways, identifying those chosen edges with a subgraph of H', which can be done in $O_\ell(1)$ ways, and choosing the rest of the vertices of H', which can be done in $O_\ell(n^{v(H')-b})$ ways, for $O_\ell(n^{v(H')-b+a+(\alpha-1)a}) \leq O_\ell(n^{v(H')-v(H)+\alpha})$ ways in total. The bound follows from summing over the $O_\ell(1)$ possible values of a and b.

We are now ready to check the conditions of Lemma 5.2 with ε replaced with $\varepsilon/2$, which we again denote using square brackets. Observe that if H' is a standard copy of H, then $\zeta_{E(H')\setminus A}$ is a product of k terms that are bounded by $1/\sqrt{p(1-p)}$; thus since $N_{H,e} = 1$ for all $e \in A$ by condition (1) we find that we always have

$$
\left| \Delta_H \sum_{\substack{H' \cong H \\ E(H') \cap A = \{e\}}} \zeta_{E(H')\setminus A} \right| \leq \sqrt{1.1}\,\pi \frac{(p(1-p))^{e(H)/2}}{(p(1-p))^{k/2}} = \sqrt{1.1}\,\pi \sqrt{p(1-p)},
$$

showing condition [1b]. Moreover, condition (1) also yields that $\Delta_H^2 \geq n^{-\alpha+\varepsilon}$, which is enough to show condition [2]. Condition [4] also holds for H as $N_{H,S} = 0$ for $|S| \geq 2$.

If $H' \neq H$, then either $H' \not\succeq H$, in which case $N_{H,S} = 0$ for all $|S| \geq 1$ and conditions [3] and [4] are trivial, or condition (2) applies. In this case, by our above discussion we have that $\sum_{|S|\geq 1} N_{H,S} \lesssim_\ell n^{v(H')-v(H)+\alpha}$, so

$$
\sum_{|S|\geq 1} \Delta_H^2 N_{H,S} \lesssim_\ell n^{-\varepsilon}.
$$

This implies conditions [3] and [4].

Finally, to show condition [5], we claim that Q simply consists of quadruples of four equal standard copies of H. Indeed, conditions [5a] to [5c] imply that any $(H_1, H_2, H_3, H_4) \in Q$ must consist of four super-rainbow copies of H and thus must be standard copies, and conditions [5b] and [5c] imply that $H_1 = H_2$ and $H_3 = H_4$. Condition [5d] ensures that $H_1 = H_2 = H_3 = H_4$. Therefore,

$$
\sum_{(H_1,H_2,H_3,H_4)\in Q} \Delta_{H_1}\Delta_{H_2}\Delta_{H_3}\Delta_{H_4} \leq n^\alpha \Delta_H^4,
$$

while

$$
\left(\sum_{e\in A} \sum_{H\in \mathcal{H}'} \Delta_H^2 N_{H,e} \right)^2 \gtrsim_{\ell,\varepsilon} n^{2\alpha}\Delta_H^4,
$$

which is enough to show condition [5]. $\qquad\qquad\square$

5.5. **Proof of Lemma 3.2.** Let $\mathcal{H}_r$ be the set of r-vertex graphs in $\mathcal{H}$ and let $\mathcal{H}'$ be the set of (possibly disconnected) NIV graphs H' such that $H' \preceq H$ for some $H \in \mathcal{H}$. By condition (3) of Definition 2.12, we can write $F_H = \sum_{H' \in \mathcal{H}'} c_{H,H'} \gamma_{H'}$ where $c_{H,H} = (p(1-p))^{e(H)/2}$ and $|c_{H,H'}| \leq \eta n^{(v(H)-v(H'))/2}$; in particular, we have $\sum_{H \in \mathcal{H}} t_H F_H = \sum_{H' \in \mathcal{H}'} \Delta_{H'} \gamma_{H'}$, where we define $\Delta_{H'} = \sum_{H \in \mathcal{H}} c_{H,H'} t_H$. Recall that $c_{H,H'} \neq 0$ implies $H \succeq H'$.

For $2 \leq r \leq \ell$, let $L_r = \max_{H \in \mathcal{H}_r} n^{r/2} |t_H|$. By hypothesis, there is some r with $L_r \geq n^\varepsilon$; therefore, we may pick some m such that $L_m \geq n^{\varepsilon/\ell} \max(1, \max_{r>m} L_r)$. We now split into cases.

Suppose $L_m^2 \leq n^{m-\varepsilon}$. From the bounds on $c_{H,H'}$, the contribution to $n^{v(H')/2} \Delta_{H'}$ from $\sum_{H \in \mathcal{H}_r} t_H F_H$ is $O_{\ell,\eta}(L_r)$. It follows that for $H \in \mathcal{H}'$ with more than m vertices we have $n^{v(H)} \Delta_H^2 \lesssim_{\ell,\eta} n^{-2\varepsilon/\ell} L_m^2$. Moreover, for m-vertex $H \in \mathcal{H}'$, we have $n^m \Delta_H^2 \lesssim_{\ell,\eta} L_m^2$ and $n^{m/2}|\Delta_H - \Delta_H'| \lesssim_{\ell,\eta} n^{-\varepsilon/\ell} L_m$, where we define $\Delta_H' = \sum_{H' \in \mathcal{H}_m} c_{H',H} t_{H'}$.

Now, let β be such that $L_m^2 = n^{m\beta + \varepsilon/\ell}$; note that

$$0 \leq \beta \leq 1 - \frac{\varepsilon + \varepsilon/\ell}{m} \leq 1 - \frac{\varepsilon}{\ell}.$$

We then apply Lemma 5.3 with α and β equal to this choice of β, $k = m - 2$, and with ε replaced with $\varepsilon/(2\ell)$. Condition (3) follows from our previous discussion, whereas to show condition (1) we need to verify for some positive $\Omega_{\ell,\varepsilon}(1)$ that

$$m\beta + \varepsilon/\ell + \Omega_{\ell,\varepsilon}(1) \leq 2 + (m-2)\beta - \frac{\varepsilon}{2\ell} \iff 2\beta + \Omega_{\ell,\varepsilon}(1) \leq 2 - \frac{3\varepsilon}{2\ell},$$

which is true. Finally, to show condition (2) we need to prove that there exists some $H \in \mathcal{H}_m$ with $n^{m/2}|\Delta_H| \geq n^{m\beta/2 + \varepsilon/(4\ell)} = n^{-\varepsilon/(4\ell)} L_m$. Suppose not; then by our bound on $|\Delta_H - \Delta_H'|$ we must then have $n^{m/2}|\Delta_H'| \lesssim_{\ell,\eta} n^{-\varepsilon/(4\ell)} L_m$. Let $M = (c_{H',H})_{H,H' \in \mathcal{H}_m}$ be a matrix with rows and columns indexed by $\mathcal{H}_m$; by construction, $(\Delta_H')_{H \in \mathcal{H}_m} = M \cdot (t_H)_{H \in \mathcal{H}_m}$. Observe that under the $\preceq$ order, M is upper triangular, has entries bounded by η, and has diagonal entries that are $\Theta_{\lambda,\ell}(1)$. As a result, M^{-1} has entries that are $O_{\lambda,\ell,\eta}(1)$, which implies that $n^{m/2}|t_H| \lesssim_{\lambda,\ell,\eta} n^{-\varepsilon/(4\ell)} L_m$ for all $H \in \mathcal{H}_m$, contradicting the definition of L_m.

If $n^{m-\varepsilon} < L_m^2 \leq \pi^2 n^m$, further let $\mathcal{H}_{m,s}$ be the set of graphs in $\mathcal{H}$ with m vertices and s edges and define $R_s = \max_{H \in \mathcal{H}_{m,s}} n^{m/2}|t_H|$. Since $L_m \geq n^{\varepsilon/\ell} \max(1, \max_{r>m} L_r)$, we may pick some u such that $R_u \geq n^{\varepsilon/\ell^3} \max(L_m/n^{\varepsilon/\ell}, \max_{s>u} R_u)$. By similar logic to the first case we find that for all $H \in \mathcal{H}'$ with greater than m vertices or with m vertices and greater than u edges, we have $n^{v(H)} \Delta_H^2 \lesssim_{\ell,\eta} n^{-2\varepsilon/\ell^3} R_u^2$. Moreover, since $H \prec H'$ implies $v(H) < v(H')$ or $e(H) < e(H')$, there exists some $H \in \mathcal{H}_{m,u}$ with $n^{v(H)} \Delta_H^2 = (1 + O_{\ell,\eta}(n^{-\varepsilon/\ell^3}))(p(1-p))^u R_u^2$. Choose α such that $R_u^2 = n^{m-\alpha+\varepsilon/\ell^3}$, and note that

$$\frac{\varepsilon}{\ell^3} - \frac{\log \pi^2}{\log n} \leq \alpha \leq \varepsilon + \frac{2\varepsilon}{\ell} - \frac{\varepsilon}{\ell^3} \leq 2\varepsilon.$$

The conditions of Lemma 5.4 can straightforwardly be shown to hold for this choice of H and α and with $\varepsilon/(2\ell^3)$ replacing ε, as long as $\varepsilon \leq \frac{1}{2}$ and n is sufficiently large, both of which we may assume. This completes the proof.

6. Application to Proportional Graphs

Qualitatively speaking, Theorem 1.2 says that given a list of desired connected subgraph counts for an n-vertex graph that are close to their expectations in $G(n,p)$, the only obstruction to the existence of a graph with those subgraph counts for large n is integrality. In this section, we apply this idea to enumerate and show the existence of p-proportional graphs and related concepts. We begin with the following definition of Janson.

Definition 6.1 ([7, p. 64])**.** Given a nonempty set of graphs $\mathcal{H}$ and some $0 < p < 1$, a graph G is $(p, \mathcal{H})\hat{}$*-proportional* if $v(G) \geq \max\{v(H)\colon H \in \mathcal{H}\}$ and $\gamma_H(G) = 0$ for all $H \in \mathcal{H}$.

Letting $\mathcal{C}_k$ be the set of (isomorphism classes of) connected k-vertex graphs, we additionally call G *p-proportional* if G is $(p, \mathcal{C}_2 \cup \mathcal{C}_3)\hat{}$-proportional, *p-superproportional* if G is $(p, \mathcal{C}_2 \cup \mathcal{C}_3 \cup \mathcal{C}_4)\hat{}$-proportional, and *p-hyperproportional* if G is $(p, \mathcal{C}_3 \cup \mathcal{C}_4 \cup \mathcal{C}_5 \cup \{2K_2\})\hat{}$-proportional. These definitions are motivated by the fact that they appear as possible degenerate cases in the distribution of the number of induced copies of G in the $G(n, p)$ and $G(n, m)$ models (for more details see [7, Sec. 10]).

Remark 6.2. The terms "*p*-proportional" and "*p*-superproportional" are have appeared previously in the literature [8, 7], while the term "*p*-hyperproportional" is new. We caution the reader that although p-superproportionality is a stronger condition than p-proportionality, it is impossible for a graph to be $(p, \{K_2, P_2, 2K_2\})\hat{}$-proportional (see [6] or the discussion below), implying that p-hyperproportionality and p-proportionality are in fact disjoint conditions.

Since every p-proportional graph G must have $e(G) = p\binom{v(G)}{2}$, no p-proportional graph can exist with p irrational. Conversely, for any rational $p \in (0, 1)$ it was shown by Kärrman [11] and Janson and Spencer [9] that there exist infinitely many p-proportional graphs. Janson and Spencer [9] additionally showed that n-vertex p-proportional graphs exist for all sufficiently large n satisfying an integrality condition (the same one that we discuss below). Turning to p-superproportional graphs, in 1994 Kärrman [10] published an example of a 64-vertex $\frac{1}{2}$-superproportional graph found by computer search, and no other examples have appeared in the literature.

Using Theorem 1.2, we are able to resolve a conjecture of Janson and Kratochvíl [8] that a local central limit theorem exists for proportional graphs. Specifically, given any downwards collection of NIV connected graphs $\mathcal{H}$, the probability that $G \sim G(n, p)$ is $(p, \mathcal{H})\hat{}$-proportional is

$$(6.1) \qquad \frac{\prod_{H \in \mathcal{H}}(\text{aut } H)^{1/2}}{(2\pi)^{|\mathcal{H}|/2} n^{\sum_{H \in \mathcal{H}} v(H)/2} (p(1-p))^{\sum_{H \in \mathcal{H}} e(H)/2}} \cdot (1 + O_{p,\varepsilon}(n^{-1/2+\varepsilon})),$$

provided that the all-zero tuple is permissible. To that end, we make the following definition.

Definition 6.3. A positive integer $n \geq 3$ is *p-proportional-compatible* (*p-PC* for short) if the evaluation on $\mathcal{R}_{\mathcal{C}_2 \cup \mathcal{C}_3, n}$ sending $\gamma_H \mapsto 0$ for all $H \in \mathcal{C}_2 \cup \mathcal{C}_3$ is integral. A positive integer $n \geq 4$ is *p-superproportional-compatible* (*p-SPC*) if the evaluation on $\mathcal{R}_{\mathcal{C}_2 \cup \mathcal{C}_3 \cup \mathcal{C}_4, n}$ sending $\gamma_H \mapsto 0$ for all $H \in \mathcal{C}_2 \cup \mathcal{C}_3 \cup \mathcal{C}_4$ is integral.

With p-hyperproportional graphs, the situation is slightly more complicated. Note that since we always have

$$\chi_e^2 + \frac{2p-1}{\sqrt{p(1-p)}}\chi_e - 1 = 0,$$

we conclude that

$$(6.2) \qquad \gamma_{K_2}^2 = 2\gamma_{2K_2} + 2\gamma_{P_2} - \frac{2p-1}{\sqrt{p(1-p)}}\gamma_{K_2} + \binom{n}{2}.$$

Therefore, G is p-hyperproportional if and only if G is $(p, \mathcal{C}_3 \cup \mathcal{C}_4 \cup \mathcal{C}_5)\hat{}$-proportional and

$$\gamma_{K_2}^2 + \frac{2p-1}{\sqrt{p(1-p)}}\gamma_{K_2} - \binom{n}{2} = 0 \iff \gamma_{K_2} = \frac{\frac{1}{2}-p}{\sqrt{p(1-p)}} \pm \sqrt{\binom{n}{2} + \frac{(\frac{1}{2}-p)^2}{p(1-p)}}.$$

Thus there are two types of p-hyperproportional graphs: those where the above equality holds with a $+$ sign and those where it holds with a $-$ sign. Call such graphs $(p, +)$*-hyperproportional*

and $(p, -)$-*hyperproportional*, respectively. Furthermore, define $(p, +)$- and $(p, -)$-*hyperproportional-compatible* $((p, \pm)$-*HPC*) positive integers $n \geq 5$ in a way analogous to Definition 6.3. Now, since in either case we have $|\gamma_{K_2}| \lesssim_p n \asymp \sigma_{K_2}$, Theorem 1.2 immediately implies the following.

Corollary 6.4. *For sufficiently large $(p, +)$-HPC (resp. $(p, -)$-HPC) n depending on p, there exists a $(p, +)$- (resp. $(p, -)$-) hyperproportional graph with n vertices.*

Moreover, it is straightforward to enumerate such graphs in a manner analogous to (6.1).

We conclude this section with results concerning the properties of p-PC, p-SPC, and $(p, \pm)$-HPC numbers, with their proofs, consisting of elementary number-theoretic computations, deferred to Appendix A. We begin with full characterizations of p-PC and p-SPC numbers, for which there exist infinitely many for every rational p.

Proposition 6.5. *Suppose $p = a/b$ where a and b are relatively prime positive integers. Then a positive integer $n \geq 3$ is p-PC if and only if*

(1) If $2 \mid b$, $\nu_2(n) \geq 3\nu_2(b)$ or $\nu_2(n - 1) \geq 3\nu_2(b) + 1$;
(2) If $3 \mid b$, $\max\{\nu_3(n), \nu_3(n - 1)\} \geq 3\nu_3(b) + 1$;
(3) For every prime $q > 3$, $\max\{\nu_q(n), \nu_q(n - 1)\} \geq 3\nu_q(b)$.

Proposition 6.6. *Suppose $p = a/b$ where a and b are relatively prime positive integers. Then a positive integer $n \geq 4$ is p-SPC if and only if*

(1) If $\nu_2(b) = 1$, $\nu_2(n) \geq 6$ or $\nu_2(n - 1) \geq 7$;
(2) If $\nu_2(b) = 2$, $\nu_2(n) \geq 13$ or $\nu_2(n - 1) \geq 11$;
(3) If $\nu_2(b) \geq 3$, $\nu_2(n) \geq 6\nu_2(b) + 1$ or $\nu_2(n - 1) \geq 6\nu_2(b)$;
(4) If $3 \mid b$, $\nu_3(n) \geq 6\nu_3(b)$ or $\nu_3(n - 1) \geq 6\nu_3(b) + 1$;
(5) For every prime $q > 3$, $\max\{\nu_q(n), \nu_q(n - 1)\} \geq 6\nu_q(b)$.

While it would be theoretically possible to write out a similar characterization of $(p, \pm)$-HPC numbers, there are so many cases that we believe that it is not worth stating. Instead of a full characterization, we observe the following special cases.

Proposition 6.7. *If n is $(p, \pm)$-HPC, then p must be rational. Moreover, given a rational $p \in (0, 1)$, there are infinitely many $(p, +)$-HPC numbers if and only if there are infinitely many $(p, -)$-HPC numbers, which occurs precisely when $p \in \{\frac{1}{3}, \frac{2}{3}\}$ or $p(1 - p)$ is not $2c^2$ for some rational c.*

As one might guess from the square roots involved in the definition of $(p, \pm)$-HPC numbers, $(p, \pm)$-HPC numbers correspond to solutions to generalized Pell equations subject to certain congruence relations, meaning that they are significantly rarer than p-PC or p-SPC numbers. As an example, we can characterize $(\frac{1}{2}, \pm)$-HPC numbers as follows.

Proposition 6.8. *A positive integer n is $(\frac{1}{2}, +)$-HPC if and only if it is $(\frac{1}{2}, -)$-HPC. Moreover this happens if and only if*

$$n = \left(\frac{(1 + \sqrt{2})^a + (1 + \sqrt{2})^{-a}}{2} \right)^2$$

for some positive integer $a \geq 2$ with $a \equiv 0, \pm 1, \pm 511 \pmod{1024}$. In particular, the smallest $(\frac{1}{2}, \pm)$-HPC number is

$$n = 3932696430232916987572576858855973259938483838348650079426054715877 \\ 66460908036341390115717616446659111649953158565894578440401902748 \\ 69003248953398849989229741078376175959761206581014547997844305527 \\ 64913183984205357972509264578282270494361671428382960796345633800 \\ 32684372594215577664684891651968024387144274928613262936942368361 \\ 6897572759524717640627101771763136024166847479649023407565312022$$

$$\approx 3.93 \times 10^{390}.$$

On the other hand, exactly when $p \in \{\frac{1}{3}, \frac{2}{3}\}$, the Pell equation becomes degenerate and the number of solutions increases drastically. In particular, we have the following.

Proposition 6.9. *A positive integer $n \geq 5$ is $(\frac{1}{3}, +)$-HPC if and only if it is $(\frac{2}{3}, -)$-HPC, which occurs if and only if $3^9 \mid n$ or $3^{10} \mid n - 2$. A positive integer $n \geq 5$ is $(\frac{1}{3}, -)$-HPC if and only if it is $(\frac{2}{3}, +)$-HPC, which occurs if and only if $3^{10} \mid n - 1$ or $3^{10} \mid n - 2$.*

Although we do not have a rigorous proof, it appears likely that the only HPC integers less than, say, 10^{100} are those mentioned above. While working with dense $3^9 = 19683$-vertex graphs is possible with computers, actually finding a 19683-vertex $\frac{1}{3}$-hyperproportional graph appears to be out of reach barring some algorithmic improvement.

APPENDIX A. ON p-PC, p-HPC, AND $(p, \pm)$-HPC NUMBERS

We begin by stating two relations in $\mathcal{R}_n$. First of all, we recall (2.1):

$$(A.1) \qquad X_H = \sum_{H' \subseteq H} p^{e(H) - e(H')} (p(1 - p))^{e(H')/2} \frac{\operatorname{aut} H'}{\operatorname{aut} H} \gamma_{H'}.$$

Moreover, if H is NIV and $m \geq 0$, we have

$$(A.2) \qquad \gamma_{H \sqcup m \bullet} = \binom{n - v(H)}{m} \gamma_H.$$

A.1. **Proof of Proposition 6.5.** Every graph with at most 3 vertices either has no edges or is the disjoint union of an element of $\mathcal{C}_2 \cup \mathcal{C}_3$ and some isolated vertices. Therefore, if Φ sends $\gamma_H \mapsto 0$ for all $H \in \mathcal{C}_2 \cup \mathcal{C}_3$, we find that for all $H \in \mathcal{C}_2 \cup \mathcal{C}_3$, all but one term in (A.1) is zero. Thus

$$\Phi(X_H) = p^{e(H)} \frac{v(H)!}{\operatorname{aut} H} \binom{n}{v(H)}.$$

Since $\frac{v(H)!}{\operatorname{aut} H} \binom{n}{v(H)}$ is an integer, we conclude that n is p-PC if and only if $b \mid \binom{n}{2}$, $b^2 \mid 3\binom{n}{3}$, and $b^3 \mid \binom{n}{3}$. Note that the third condition obviously implies the second. Thus n is p-PC if and only if

$$\nu_q(n) + \nu_q(n - 1) \geq \nu_q(b) + \nu_q(2) \quad \text{and} \quad \nu_q(n) + \nu_q(n - 1) + \nu_q(n - 2) \geq 3\nu_q(b) + \nu_q(6)$$

for every prime q that divides b.

If $q = 2$, the first condition implies that $n \equiv 0, 1 \pmod{4}$. If $n \equiv 0 \pmod{4}$, the conditions are satisfied if and only if we additionally have $\nu_2(n) \geq 3\nu_2(b)$. If $n \equiv 1 \pmod{4}$, the conditions are equivalent to $\nu_2(n - 1) \geq 3\nu_2(b) + 1$. This yields condition (1).

If $q \geq 3$, and n is not 0 or 1 mod q, the first condition yields a contradiction. Otherwise, if $n \equiv r \in \{0, 1\}$ modulo q, then the conditions are equivalent to $\nu_q(n - r) \geq 3\nu_q(b) + \nu_q(3)$, yielding conditions (2) and (3).

A.2. **Proof of Proposition 6.6.** Let Φ be the evaluation of $\mathcal{R}_{\mathcal{C}_2 \cup \mathcal{C}_3 \cup \mathcal{C}_4, n}$ sending $\gamma_H \mapsto 0$ for all $H \in \mathcal{C}_2 \cup \mathcal{C}_3 \cup \mathcal{C}_4$. By (A.2), we find that for all H with at most 4 vertices, we have $\Phi(\gamma_H) = 0$, unless H is edgeless or $H = 2K_2$. In the latter case, by (6.2) we find that

$$\Phi(\gamma_{2K_2}) = -\frac{1}{2} \binom{n}{2}.$$

Therefore, by (A.1), we find that for all $H \in \mathcal{C}_2 \cup \mathcal{C}_3 \cup \mathcal{C}_4$,

$$\Phi(X_H) = p^{e(H)} \frac{v(H)!}{\operatorname{aut} H} \binom{n}{v(H)} - p^{e(H) - 1} (1 - p) \frac{4 X_{2K_2}(H)}{\operatorname{aut} H} \binom{n}{2}.$$

Therefore, n is p-SPC if and only if the following nine quantities are integers:

$$\Phi(X_{K_2}) = p\binom{n}{2} \quad \Phi(X_{P_2}) = 3p^2\binom{n}{3} \quad \Phi(X_{K_3}) = p^3\binom{n}{3} \quad \Phi(X_{K_{1,3}}) = 4p^3\binom{n}{4}$$

$$\Phi(X_{P_3}) = 12p^3\binom{n}{4} - 2p^2(1-p)\binom{n}{2} \quad \Phi(X_{\overline{P_2 \sqcup \bullet}}) = 12p^4\binom{n}{4} - 2p^3(1-p)\binom{n}{2}$$

$$\Phi(X_{C_4}) = 3p^4\binom{n}{4} - p^3(1-p)\binom{n}{2} \quad \Phi(X_{\overline{K_2 \sqcup 2\bullet}}) = 6p^5\binom{n}{4} - 2p^4(1-p)\binom{n}{2}$$

$$\Phi(X_{K_4}) = p^6\binom{n}{4} - \frac{1}{2}p^5(1-p)\binom{n}{2}$$

Suppose n is p-SPC and consider the quantity $N = 2p^5\binom{n}{4} - p^5(1-p)\binom{n}{2}$. If n is p-SPC, N has to be an integer as well, since the only way for it not to be an integer is that if it has factors of b in its denominator, and if that is the case then $pN = 2\Phi(X_{K_4})$ cannot be an integer. Therefore,

$$\Phi(X_{\overline{K_2 \sqcup 2\bullet}}) - 3N = p^4(1-p)\binom{n}{2}$$

has to be an integer as well, implying that $b^5 \mid \binom{n}{2}$.

Conversely, if $b^5 \mid \binom{n}{2}$, it follows that

$$b^5 \mid 3\binom{n}{3} = (n-2)\binom{n}{2} \quad \text{and} \quad b^5 \mid 6\binom{n}{4} = \binom{n}{2}\binom{n-2}{2}.$$

It follows that that $b^4 \mid \binom{n}{3}, \binom{n}{4}$, implying for all $H \in \mathcal{C}_2 \cup \mathcal{C}_3 \cup \mathcal{C}_4$ with $e(H) \le 4$, $\Phi(X_H)$ is an integer. Moreover,

$$\Phi(X_{\overline{K_2 \sqcup 2\bullet}}) = 6p^5\binom{n}{4} - 2p^4(1-p)\binom{n}{2} = p^5\binom{n}{2}\binom{n-2}{2} - 2p^4(1-p)\binom{n}{2}$$

is an integer as well. As a result, we find that n is p-SPC if and only if $b^5 \mid \binom{n}{2}$ and

$$\Phi(X_{K_4}) \in \mathbb{Z} \iff 2b^6 \mid a^5\left(2a\binom{n}{4} - (b-a)\binom{n}{2}\right).$$

We now take q-adic evaluations, where q ranges across the prime divisors of $2b$.

First, if $q \ge 3$, then $\nu_q(b^5) \le \nu_q(\binom{n}{2})$ is equivalent to $\max\{\nu_q(n), \nu_q(n-1)\} \ge 5\nu_q(b)$. Moreover, since $b^6 \mid b\binom{n}{2}$, the q-adic portion of the second condition is equivalent to

$$6\nu_q(b) \le \nu_q\left(a^5\left(2a\binom{n}{4} + a\binom{n}{2}\right)\right) = \nu_q\left(2\binom{n}{4} + \binom{n}{2}\right).$$

Now, we may compute

$$2\binom{n}{4} + \binom{n}{2} = \binom{n}{2}\left(\frac{(n-2)(n-3)+6}{6}\right).$$

If $n \equiv 0, 1 \pmod{q^{5\nu_q(b)}}$, we also have $(n-2)(n-3)+6 \equiv 12, 8 \pmod{q^{5\nu_q(b)}}$ (respectively). Thus $\nu_q((n-2)(n-3)+6)$ is 0 unless $q = 3$ and $n \equiv 0 \pmod{q^{5\nu_q(b)}}$, in which case it is 1, meaning that

$$\nu_q\left(2\binom{n}{4} + \binom{n}{2}\right) = \nu_q\left(\binom{n}{2}\right) + \begin{cases} -1 & \text{if } q = 3 \text{ and } n \equiv 1 \pmod{q^{5\nu_q(b)}} \\ 0 & \text{else} \end{cases}$$

Thus our final condition is $\max\{\nu_q(n), \nu_q(n-1)\} \ge 6\nu_q(b)$ if $q \ge 5$, and $\nu_q(n) \ge 6\nu_q(b)$ or $\nu_q(n-1) \ge 6\nu_q(b)+1$ if $q = 3$, matching conditions (4) and (5).

We now turn to the case $q = 2$. First of all, if b is odd, then the conditions are equivalent to

$$a^5 \left(2a \binom{n}{4} - (b - a) \binom{n}{2} \right)$$

being even. The first term is obviously even, while the second is as well since $a(b - a)$ must be even. Henceforth assume that b is even. Since a is odd, we find that assuming $5\nu_2(b) \leq \nu_2(\binom{n}{2})$,

$$\nu_2(2b^6) \leq \nu_2 \left(b(a/3 + 1) \binom{n}{2} \right),$$

meaning that we can replace the second condition with

$$6\nu_2(b) + 1 \leq \nu_2 \left(a^5 \left(2a \binom{n}{4} - (-ab/3 - a) \binom{n}{2} \right) \right) = \nu_2 \left(2 \binom{n}{4} + (b/3 + 1) \binom{n}{2} \right).$$

Now, since

$$2 \binom{n}{4} + (b/3 + 1) \binom{n}{2} = \binom{n}{2} \left(\frac{(n - 2)(n - 3) + 6 + 2b}{6} \right),$$

we have rewritten our conditions as

$$5\nu_2(b) \leq \nu_2 \left(\binom{n}{2} \right) \quad \text{and} \quad 6\nu_2(b) + 2 \leq \nu_2 \left(\binom{n}{2} \right) + \nu_2((n - 2)(n - 3) + 6 + 2b).$$

Now assume that n is even. Then $5\nu_2(b) + 1 \leq \nu_2(n)$, meaning that modulo $2^{5\nu_2(b)+1}$, we have $(n-2)(n-3)+6+2b \equiv 2b+12$, so $(n-2)(n-3)+6+2b$ has 2-adic evaluation at least 3 if $\nu_2(b) = 1$ and exactly 2 otherwise. Thus the conditions are satisfied exactly when $\nu_2(n) \geq 6$ if $\nu_2(b) = 1$ and $\nu_2(n) \geq 6\nu_2(b) + 1$ otherwise, matching conditions (1) to (3).

If n is odd, then $5\nu_2(b)+1 \leq \nu_2(n-1)$, so modulo $2^{5\nu_2(b)+1}$, we have $(n-2)(n-3)+6+2b \equiv 2b+8$, meaning that the 2-adic evaluation of $(n-2)(n-3)+6+2b$ is 2 if $\nu_2(b) = 1$, at least 4 if $\nu_2(b) = 2$, and 3 otherwise. Therefore the conditions are satisfied if and only if $\nu_2(n - 1) \geq 7$ for $\nu_2(b) = 1$, $\nu_2(n - 1) \geq 11$ for $\nu_2(b) = 2$, and $\nu_2(n - 1) \geq 6\nu_2(b)$ otherwise. This also matches conditions (1) to (3).

A.3. **Generalities surrounding $(p, \pm)$-HPC numbers.** In this section we develop a number of general results that will be useful in the proofs of Propositions 6.7 to 6.9, including proving part of Proposition 6.7.

Consider the evaluation $\Phi_\pm$ sending $\gamma_H \mapsto 0$ for all $H \in \mathcal{C}_3 \cup \mathcal{C}_4 \cup \mathcal{C}_5$ and sending

$$\gamma_{K_2} \mapsto \frac{\frac{1}{2} - p}{\sqrt{p(1 - p)}} \pm \sqrt{\binom{n}{2} + \frac{(\frac{1}{2} - p)^2}{p(1 - p)}}.$$

First we compute $\Phi_\pm(\gamma_H)$ for all NIV H. The NIV graphs with at most 5 vertices consist of the empty graph, K_2, members of $\mathcal{C}_3 \cup \mathcal{C}_4 \cup \mathcal{C}_5$, $2K_2$, $K_2 \sqcup P_2$, and $K_2 \sqcup K_3$. By construction, $\Phi_\pm(\gamma_{2K_2}) = 0$. Also, in a similar manner to (6.2) we obtain that in $\mathcal{R}_n$,

$$\gamma_{K_2} \gamma_{P_2} = \gamma_{K_2 \sqcup P_2} + 2\gamma_{P_3} + 3\gamma_{K_{1,3}} - 2\frac{2p - 1}{\sqrt{p(1 - p)}} \gamma_{P_2} + 2(n - 2)\gamma_{K_2}$$

$$\gamma_{K_2} \gamma_{K_3} = \gamma_{K_2 \sqcup K_3} + \gamma_{\overline{P_2 \sqcup \bullet}} - 3\frac{2p - 1}{\sqrt{p(1 - p)}} \gamma_{K_3} + \gamma_{P_2}.$$

It follows that

$$\Phi_\pm(\gamma_{K_2 \sqcup P_2}) = -2(n - 2)\Phi_\pm(\gamma_{K_2}) \quad \text{and} \quad \Phi_\pm(\gamma_{K_2 \sqcup K_3}) = 0.$$

Therefore, from (A.1) and (A.2), we obtain that for all $H \in \mathcal{C}_2 \cup \mathcal{C}_3 \cup \mathcal{C}_4 \cup \mathcal{C}_5$, we have

$$(A.3) \quad \Phi_\pm(X_H) = p^{e(H)} \frac{v(H)!}{\text{aut } H} \binom{n}{v(H)} + p^{e(H)-1} \sqrt{p(1-p)} \frac{2(v(H)-2)!e(H)}{\text{aut } H} \binom{n-2}{v(H)-2} \Phi_\pm(\gamma_{K_2})$$

$$- p^{e(H)-3}(p(1-p))^{3/2} \frac{4X_{K_2 \sqcup P_2}(H)}{\text{aut } H} 2(n-2)\Phi_\pm(\gamma_{K_2}),$$

where the last term only occurs for 5-vertex graphs.

We first need a lemma about these coefficients.

Lemma A.1. *If H' and H are graphs with $v(H) \geq v(H')$, then*

$$\frac{\text{aut } H' \cdot X_{H'}(H) \cdot (v(H) - v(H'))!}{\text{aut } H} \in \mathbb{Z}.$$

In particular, for $H \in \mathcal{C}_2 \cup \mathcal{C}_3 \cup \mathcal{C}_4 \cup \mathcal{C}_5$,

$$\frac{v(H)!}{\text{aut } H}, \quad \frac{2(v(H)-2)!e(H)}{\text{aut } H}, \quad \frac{4X_{K_2 \sqcup P_2}(H)}{\text{aut } H} \in \mathbb{Z}$$

(the above statement applied for $H' = \varnothing, K_2, K_2 \sqcup P_2$).

Proof. Consider the action of $\text{Aut } H$ on injective graph homomorphisms $H' \to H$. The stabilizer of each homomorphism must permute the vertices not in the image of the homomorphism, meaning that it must have size dividing $(v(H) - v(H'))!$. This implies that the size of each orbit must have size a multiple of $\text{aut } H/(v(H) - v(H'))!$. But there are exactly $\text{aut } H' \cdot X_{H'}(H)$ such homomorphisms, so the result follows. $\square$

We now start proving Proposition 6.7.

Claim A.2. *If there exists an integer $n \geq 5$ that is $(p,+)$- or $(p,-)$-HPC, then p is rational.*

Proof. First, consider some connected H with at most 4 vertices. Then, we find that

$$\frac{\text{aut } H}{v(H)!\binom{n}{v(H)}} \Phi_\pm(X_H) = p^{e(H)} + p^{e(H)-1}\sqrt{p(1-p)} \frac{2e(H)}{n(n-1)} \Phi_\pm(\gamma_{K_2})$$

$$= p^{e(H)}\left(1 + e(H)\sqrt{\frac{1-p}{p}} \frac{2\Phi_\pm(\gamma_{K_2})}{n(n-1)}\right)$$

is a rational number; since it depends only on $e(H)$ call it $\xi_{e(H)}$. Now we compute that

$$\xi_1^2 - \xi_2 = \frac{4p(1-p)}{n^2(n-1)^2} \Phi_\pm(\gamma_{K_2})^2 \quad \text{and} \quad \xi_1\xi_2 - \xi_3 = \frac{8p^2(1-p)}{n^2(n-1)^2} \Phi_\pm(\gamma_{K_2})^2$$

are both rational, so their quotient, which is equal to $2p$, must be rational as well. $\square$

Remark A.3. This argument in fact proves that if there exists a $(p, \mathcal{C}_3)\hat{\ }$-proportional graph, then p is rational.

For the remainder of the proof we will consider rational p; suppose that $p = a/b$ with $\gcd(a,b) = 1$. Define

$$h_\pm = \sqrt{a(b-a)}\Phi_\pm(\gamma_{K_2}) = \frac{b - 2a \pm \sqrt{D}}{2},$$

where we additionally define

$$D = 2a(b-a)n(n-1) + (b-2a)^2 = \frac{2a(b-a)(2n-1)^2 + 2(3a-b)(3a-2b)}{4}.$$

Now (A.3) becomes

$$(A.4) \quad \Phi_\pm(X_H) = \frac{a^{e(H)-2}}{b^{e(H)}} \left(\frac{v(H)!}{\mathrm{aut}\, H} a^2 \binom{n}{v(H)} + \frac{2(v(H)-2)!e(H)}{\mathrm{aut}\, H} a \binom{n-2}{v(H)-2} h_\pm \right.$$
$$\left. - 2\frac{4X_{K_2 \sqcup P_2}(H)}{\mathrm{aut}\, H}(b-a)(n-2)h_\pm \right).$$

Lemma A.4. *A necessary condition for n to be $(p, \pm)$-HPC is that $\frac{1}{b}(h_\pm + a\binom{n}{2})$ is an integer and $b^8 \mid 2h_\pm(n-2)$. A sufficient condition for n to be $(p, \pm)$-HPC is that $\frac{1}{b}(h_\pm + a\binom{n}{2})$ is an integer and $b^{12} \mid n - 2$.*

Proof. First assume that n is $(p, \pm)$-HPC. Since

$$\Phi_\pm(X_{K_2}) = \frac{a}{b}\binom{n}{2} + \frac{\sqrt{a(b-a)}}{b}\Phi_\pm(\gamma_{K_2}) = \frac{a\binom{n}{2} + h_\pm}{b},$$

the first condition is evident. Now, we consider the graphs K_5, $\overline{K_2 \sqcup 3\bullet}$, and $-(P_2 + 2\bullet)$. After computing

$$\mathrm{aut}\, K_5 = 120 \qquad\qquad \mathrm{aut}(\overline{K_2 \sqcup 3\bullet}) = 12 \qquad\qquad \mathrm{aut}(\overline{P_2 \sqcup 2\bullet}) = 4$$
$$e(K_5) = 10 \qquad\qquad e(\overline{K_2 \sqcup 3\bullet}) = 9 \qquad\qquad e(\overline{P_2 \sqcup 2\bullet}) = 8$$
$$X_{K_2 \sqcup P_2}(K_5) = 30 \qquad\qquad X_{K_2 \sqcup P_2}(\overline{K_2 \sqcup 3\bullet}) = 21 \qquad\qquad X_{K_2 \sqcup P_2}(\overline{P_2 \sqcup 2\bullet}) = 13$$

by (A.3) we have

$$(A.5) \qquad \Phi_\pm(X_{K_5}) = \frac{a^8}{b^{10}}\left(a^2\binom{n}{5} + a\binom{n-2}{3}h_\pm - 2(b-a)(n-2)h_\pm \right)$$

$$(A.6) \qquad \Phi_\pm(X_{\overline{K_2 \sqcup 3\bullet}}) = \frac{a^7}{b^9}\left(10a^3\binom{n}{5} + 9a\binom{n-2}{3}h_\pm - 14(b-a)(n-2)h_\pm \right)$$

$$\Phi_\pm(X_{\overline{P_2 \sqcup 2\bullet}}) = \frac{a^6}{b^8}\left(30a^3\binom{n}{5} + 24a\binom{n-2}{3}h_\pm - 26(b-a)(n-2)h_\pm \right).$$

Now we observe that

$$-30b^2\Phi_\pm(X_{K_5}) + 6ab\Phi_\pm(X_{\overline{K_2 \sqcup 3\bullet}}) - a^2\Phi_\pm(X_{\overline{P_2 \sqcup 2\bullet}}) = \frac{a^8}{b^8}(2(b-a)(n-2)h_\pm),$$

so $b^8 \mid 2(n-2)h_\pm$.

Conversely, assume that $\frac{1}{b}(h_\pm + a\binom{n}{2})$ is an integer and $b^{12} \mid n - 2$. Clearly $\Phi_\pm(X_{K_2})$ and $h_\pm$ are integers. Moreover, for $H \in \mathcal{C}_3 \cup \mathcal{C}_4 \cup \mathcal{C}_5$, it is straightforward to show that b^{10} divides $\binom{n}{v(H)}$, $\binom{n-2}{v(H)-2}$, and $n - 2$. We are now done by combining (A.4) and Lemma A.1. $\square$

A.4. **Proof of remainder of Proposition 6.7.** In light of Claim A.2, it suffices to consider rational p.

If $p \notin \{\frac{1}{3}, \frac{2}{3}\}$ and $2a(b-a)$ is a square, then there are only finitely many n for which $4D$ is a perfect square, since $2(3a - b)(3a - 2b) \neq 0$, the number $2a(b-a)(2n-1)^2$ is a square, and there are only finitely many pairs of perfect squares with a given nonzero difference. Thus there are only finitely many n such that $h_\pm$ is an integer, so by Lemma A.4 there are only finitely many $(p, \pm)$-HPC n.

Now suppose p is rational but not of the above form. We first need a lemma.

Lemma A.5. *For every positive integer m, there exist infinitely many positive integers $n \equiv 2$ (mod m) such that D is a square.*

Proof. If $p \in \{\frac{1}{3}, \frac{2}{3}\}$, we may compute that $D = (2n-1)^2$, so the conclusion is obvious. For other p, we will use the theory of Pell equations. In particular, we know that since $2a(b-a)$ is not a perfect square, there exist positive integers r and s such that $r^2 - 2a(b-a)s^2 = 1$. We claim that we may additionally choose r and s to be equivalent to 1 and 0 mod $2m$. To see this, observe that $r + s\sqrt{2a(b-a)}$ is a unit inside the ring $R = (\mathbb{Z}/2m\mathbb{Z})[\sqrt{2a(b-a)}]$. Since R is finite, some power of $r + s\sqrt{2a(b-a)}$ must be equal to 1 in R, yielding the desired solution. Now, since

$$(2b)^2 - 2a(b-a) \cdot 3^2 = 2(3a-b)(3a-2b)$$

(a reflection of the fact that K_2 would be hyperproportional were it not too small), by multiplying $2b + 3\sqrt{2a(b-a)}$ by arbitrarily large powers of $r + s\sqrt{2a(b-a)}$ we find arbitrarily large positive integers t and u, congruent to $2b$ and 3 mod $2m$, such that

$$t^2 - 2a(b-a) \cdot u^2 = 2(3a-b)(3a-2b).$$

We now claim that $n = \frac{u+1}{2}$ works. Indeed, by construction $u \equiv 2 \pmod{m}$ and $D = t^2/4$. Since t is even, we are done. $\qquad\square$

We now claim that if $n \equiv 2 \pmod{2b^{12}}$ and D is a square, n is both $(p, +)$- and $(p, -)$-HPC. First of all, we note that $D \equiv 4a(b-a) + (b-2a)^2 \equiv b^2 \pmod{2b^2}$, so $\sqrt{D}$ is b times an odd integer. Therefore $h_{\pm}$ is an integer that is $-a$ mod b. Moreover, $n(n-1) \equiv 2 \pmod{2b}$, so $\binom{n}{2} \equiv 1 \pmod{b}$. Therfore $\frac{1}{b}(h_{\pm} + \binom{n}{2})$ is an integer, meaning that we are done by Lemma A.4.

A.5. **Proof of Proposition 6.8.** The proof is in two parts: we first reduce to a divisibility condition on n, and then solve the Pell equation given these divisiblity conditions. We will also note that $h_{\pm} = \sqrt{\binom{n}{2}}$.

Claim A.6. An integer $n \geq 5$ is $(\frac{1}{2}, +)$-HPC if and only if it is $(\frac{1}{2}, -)$-HPC, which occurs exactly when

- $h_{\pm}$ is an integer, and
- $2^{13} \mid n$, $2^{21} \mid n-1$, or $2^{11} \mid n-2$.

Proof. Throughout this proof, we will frequently use Lemma A.1 implicitly. Also note that if $h_{\pm}$ is an integer, then $\nu_2(h_{\pm}) = \frac{1}{2}(\max\{\nu_2(n), \nu_2(n-1)\} - 1)$.

Assume $h_{\pm}$ is an integer. If $2^{13} \mid n$, then we always have $2^{10} \mid \binom{n}{v(H)}$, so the first term in (A.4) is an integer. Also $2^6 \mid h_{\pm}$, so the second and third terms are integers for all graphs with at most 4 vertices. For 5 vertex graphs, we note that $\binom{n-2}{3} \equiv -4 \pmod{2^{12}}$, so $\binom{n-2}{3}h_{\pm}$ and $2(n-2)h_{\pm}$ are both divisible by 2^8. It remains to check that for connected 5-vertex graphs H with at least 9 edges,

$$2^{e(H)-8} \mid -\frac{12e(H)}{\text{aut } H} + \frac{4X_{K_2 \sqcup P_2}(H)}{\text{aut } H}.$$

If H is K_5 minus an edge, then we need to show $2 \mid -9 + 7$, which is true. If $H = K_5$, then we need to show $4 \mid -1 + 1$, which is also true.

If $2^{21} \mid n-1$, then $2^{10} \mid h_{\pm}$ and $2^{10} \mid \binom{n}{v(H)}$, so n is indeed $(\frac{1}{2}, \pm)$-HPC.

If $2^{11} \mid n-2$, then we first note that since $h_{\pm}$ is odd, $\Phi_{\pm}(X_{K_2}) = \frac{1}{2}(\binom{n}{2} + h_{\pm})$ is an integer. Now, for $v(H) \geq 3$, we have $2^{10} \mid \binom{n-2}{v(H)-2}$, so we may ignore the second and third terms. Moreover, it is straightforward to compute that $2^{11} \mid \binom{n}{3}$, $2^9 \mid \binom{n}{4}$, and $2^{10} \mid \binom{n}{5}$, so the first term is integral as well. This finishes the proof of one direction.

In the other direction, we note that by Lemma A.4 we get that $h_{\pm}$ is an integer and $2^7 \mid h_{\pm}(n-2)$. We now split into cases depending on n mod 4.

If $n \equiv 0 \pmod{4}$, then we get that $2^6 \mid h_{\pm}$, so $2^{13} \mid n$, as desired.

If $n \equiv 1 \pmod 4$, then $2^7 \mid h_\pm$ and thus $2^{15} \mid n - 1$. This implies that $2^{10} \mid \binom{n}{5}$, so by considering $\Phi_\pm(X_{K_5})$ we get that

$$2^{10} \mid h_\pm\left(\binom{n-2}{3} - 2(n-2)\right).$$

But $\binom{n-2}{3} - 2(n-2)$ is odd, so $2^{10} \mid h_\pm$ and $2^{21} \mid n - 1$.

If $n \equiv 2 \pmod 4$, then $h_\pm$ is odd, immediately implying that $2^7 \mid n - 2$. This is enough to imply that

$$\nu_2\left(\binom{n}{5}\right) = \nu_2(n-2) - 1, \qquad \nu_2\left(\binom{n-2}{3}h_\pm\right) = \nu_2(n-2), \qquad \nu_2(2(n-2)h_\pm) = \nu_2(n-2) + 1.$$

Therefore $\nu_2(\Phi_\pm(X_{K_5})) = \nu_2(n-2) - 11$, implying that $2^{11} \mid n - 2$.

Finally, if $n \equiv 3 \pmod 4$, then $(n-2)h_\pm$ is odd, contradiction. This concludes the proof. $\square$

Now, since we can rewrite $h_\pm = \pm\sqrt{\binom{n}{2}}$ as $(2n-1)^2 - 2(2h_\pm)^2 = 1$, the values of $n \geq 1$ with $h_\pm$ integer correspond to solutions to the Pell equation $r^2 - 2s^2 = 1$ with r odd and positive and s even and nonnegative (which is all nonnegative solutions by mod 4 reasons). The fundamental solution is $(r, s) = (3, 2)$, so in general the nonnegative solutions are given by (r_a, s_a) where

$$r_a + s_a\sqrt{2} = (3 + 2\sqrt{2})^a = (1 + \sqrt{2})^{2a} \iff r_a = \frac{(1 + \sqrt{2})^{2a} + (1 + \sqrt{2})^{-2a}}{2}$$

for $a \geq 0$. This in turn means that

$$n = \frac{r_a + 1}{2} = \left(\frac{(1 + \sqrt{2})^a + (1 + \sqrt{2})^{-a}}{2}\right)^2.$$

It remains to show that for $n \geq 5$ (equivalent to $a \geq 2$) of this form, we have $2^{13} \mid n$, $2^{21} \mid n - 1$, or $2^{11} \mid n - 2$ if and only if $a \equiv 0, \pm 1, \pm 511 \pmod{1024}$. In other words, we need to show the following.

Claim A.7. Let $a \geq 2$ and let $r_a + s_a\sqrt{2} = (3 + 2\sqrt{2})^a$. The following are equivalent:

(1) $2^{14} \mid r_a + 1$, $2^{22} \mid r_a - 1$, or $2^{12} \mid r_a - 3$,
(2) $a \equiv 0, \pm 1, \pm 511 \pmod{1024}$.

Proof. We work inside the field $\mathbb{Q}_2(\sqrt{2})$, which has a valuation $\|-\|_2$ and ring of integers $\mathbb{Z}_2[\sqrt{2}]$. Let $\rho = 3 + 2\sqrt{2}$.

We add a condition (3): that $\rho^a \equiv \rho \pmod{2^{10}}$, $\rho^a \equiv \rho^{-1} \pmod{2^{10}}$, or $\rho^a \equiv 1 \pmod{2^{11}}$.

We first show that (1) and (3) are equivalent. We first claim that we cannot have $r_a \equiv 5, 7 \pmod 8$. Indeed, since $\rho^2 = 17 + 12\sqrt{2}$, it follows that the powers of ρ alternate $3 + 2\sqrt{2}$ and 1 modulo $4\sqrt{2}$. So $2^{14} \mid r_a + 1$ is in fact impossible. Now, if $2^{22} \mid r_a - 1$, then $r_a^2 \equiv 1 \pmod{2^{23}}$, implying that $s_a^2 \equiv 0 \pmod{2^{22}}$ and thus $2^{11} \mid s_a$. Therefore $\rho^a \equiv 1 \pmod{2^{11}}$. Moreover, if $r_a \equiv 3 \pmod{2^{12}}$, then $r_a^2 \equiv 9 \pmod{2^{13}}$, implying that $s_a^2 \equiv 4 \pmod{2^{12}}$, which then implies that $s_a \equiv \pm 2 \pmod{2^{10}}$. Thus $\rho^a \equiv \rho^{\pm 1} \pmod{2^{10}}$. This proves that (1) implies (3).

The opposite is similar. If $\rho^a \equiv 1 \pmod{2^{11}}$, then $2^{11} \mid s_a$, proving that $2^{23} \mid r_a^2 - 1$ and thus $2^{22} \mid r_a \pm 1$. But it is impossible to have $r_a \equiv 7 \pmod 8$, so $2^{22} \mid r_a - 1$. If $\rho^a \equiv \rho^{\pm 1} \pmod{2^{10}}$, then $s_a^2 \equiv 4 \pmod{2^{12}}$, so $r_a^2 \equiv 9 \pmod{2^{13}}$ and thus $r_a \equiv \pm 3 \pmod{2^{12}}$. But since $r_a \not\equiv 5 \pmod 8$, we must have $r_a \equiv 3 \pmod{2^{12}}$.

To show that (2) and (3) are equivalent, we note that $\rho^a \equiv \rho^b \pmod{2^e}$ if and only if $\rho^{a-b} \equiv 1 \pmod{2^e}$. So it suffices to show that ρ has order 2^9 mod 2^{10} and order 2^{10} mod 2^{11}. To see this, observe that if ρ has order 2 mod 4, so its order must be even. Moreover, since $\|\rho^2 - 1\|_2 = 2^{-5/2} < 2^{-1/(2-1)}$, it follows that for even a

$$\|\rho^a - 1\|_2 = \|\log(\rho^a)\|_2 = \|a/2\|_2\|\log(\rho^2)\|_2 = \|a/2\|_2\|\rho^2 - 1\|_2 = 2^{-3/2}\|a\|_2.$$

The result follows. $\qquad\square$

We remark that the above claim can also be proven by computer, since r_a is periodic with respect to any modulus.

A.6. **Proof of Proposition 6.9.** We first observe three facts. First, by (A.5) and (A.6), we have

$$10b\Phi_{\pm}(X_{K_5}) - a\Phi_{\pm}(X_{\overline{K_2 \sqcup 3\bullet}}) = \frac{a^8}{b^9}\left(a\binom{n-2}{3}h_{\pm} - 6(b-a)(n-2)h_{\pm}\right),$$

so by Lemma A.4, a necessary condition for n to be $(p, \pm)$-HPC is $3^9 \mid h_{\pm}\binom{n-2}{3}$.

Moreover, one can compute the following:

$$(A.7) \qquad n \equiv 0 \pmod{3^9} \implies \binom{n}{5} \equiv \frac{n}{5}, \quad \binom{n-2}{3}n \equiv -4n \pmod{3^{10}}$$

$$(A.8) \qquad n \equiv 1 \pmod{3^9} \implies \binom{n}{5} \equiv -\frac{n-1}{20}, \quad \binom{n-2}{3}(n-1) \equiv -(n-1) \pmod{3^{10}}$$

$$(A.9) \qquad n \equiv 2 \pmod{3^{10}} \implies \binom{n}{5} \equiv \frac{n-2}{30}, \quad \binom{n-2}{3} \equiv \frac{n-2}{3} \pmod{3^{10}}$$

Finally, note that when $b = 3$, $D = (2n-1)^2$, meaning that for $n \geq 1$, we have $h_+ = n$ and $h_- = 1 - n$ when $p = \frac{1}{3}$ and $h_+ = n - 1$ and $h_- = -n$ if $p = \frac{2}{3}$.

We now split into cases.

A.6.1. *Case 1: $n \equiv 0 \pmod 3$.* We aim to show that n is $(\frac{1}{3}, +)$- and $(\frac{2}{3}, -)$-HPC if and only if $3^9 \mid n$, and that n is never $(\frac{1}{3}, -)$- and $(\frac{2}{3}, +)$-HPC. The latter fact follows immediately from the fact that $3^8 \mid h_{\pm}(n-2)$. Now assume that we are working with the other two choices of p and sign, so that $h_{\pm} = \pm n$.

For the only if direction, note that the condition $3^9 \mid h_{\pm}\binom{n-2}{3}$ implies that $3^9 \mid n$. For the if direction, note that we automatically have that 3^8 divides $\binom{n}{v(H)}$, 3^9 divides $\binom{n}{5}$, and 3^9 divides $h_{\pm}$, so it suffices to check $H = K_5$. Since 3^9 divides n and $h_{\pm}$, this reduces to checking that modulo 3^{10}

$$0 \equiv a^2\binom{n}{5} \pm a\binom{n-2}{3}n \mp 4an \equiv \binom{n}{5} + \binom{n-2}{3}n - 4n.$$

This follows from (A.7).

A.6.2. *Case 2: $n \equiv 1 \pmod 3$.* As in the previous case, from the fact that $3^8 \mid h_{\pm}(n-2)$ we find that n is never $(\frac{1}{3}, +)$- or $(\frac{2}{3}, -)$-HPC. Now assume that we are in the other two cases, so that $h_{\pm} = \pm(n-1)$.

For the only if direction, note that the condition $3^9 \mid h_{\pm}\binom{n-2}{3}$ implies that $3^9 \mid n - 1$. Now, the integrality of $\Phi_{\pm}(X_{K_5})$ implies that, modulo 3^{10},

$$0 \equiv a^2\binom{n}{5} \pm a\binom{n-2}{3}(n-1) \mp 2a(n-1) \equiv \binom{n}{5} - \binom{n-2}{3}(n-1) + 2(n-1)$$

$$\stackrel{(A.8)}{\equiv} (n-1)\left(-\frac{1}{20} + 1 + 2\right),$$

which implies that $3^{10} \mid n - 1$ as $-\frac{1}{20} + 1 + 2 \not\equiv 0 \pmod 3$. The only if direction follows from the fact that we have $3^9 \mid \binom{n}{v(H)}$, $3^{10} \mid \binom{n}{5}$, and $3^{10} \mid h_{\pm}$.

A.6.3. *Case 3: $n \equiv 2 \pmod 3$.* We aim to show that n is $(p, \pm)$-hyperproportional if and only if $3^{10} \mid n - 2$, for both choices of p and both signs.

For the only if direction, note that since $3 \nmid h_\pm$, we must have $3^9 \mid \binom{n-2}{3}$, so $3^{10} \mid n - 2$. For the if direction, observe that $\Phi_\pm(X_{K_2}) = \frac{1}{3}(a\binom{n}{2} + h_\pm)$ is an integer since $\binom{n}{2} \equiv 1 \pmod 3$ and $h_\pm \equiv -a$ (mod 3). Moreover, for $v(H) \geq 3$, we have that 3^9 divides $\binom{n}{v(H)}$ and $\binom{n-2}{v(H)-2}$, so it suffices to check $H = K_5$. To do this, we need to show that

$$3^{10} \mid a^2 \binom{n}{5} + a \binom{n-2}{3} h_\pm,$$

which since $h_\pm \equiv -a \pmod 3$ is equivalent to $\binom{n}{5} \equiv \binom{n-2}{3} \pmod{3^{10}}$. This follows from (A.9).

References

[1] Ross Berkowitz. A Local Limit Theorem for Cliques in $G(n, p)$. Preprint. 2018. arXiv: 1811.03527. 1, 3, 12, 16

[2] Ross Berkowitz. A Quantitative Local Limit Theorem for Triangles in Random Graphs. Preprint. 2016. arXiv: 1610.01281. 1

[3] Jacob Fox, Matthew Kwan, and Lisa Sauermann. Anti-concentration for subgraph counts in random graphs. *Ann. Probab.* **49**, 1515–1553. 2021. 1

[4] Justin Gilmer and Swastik Kopparty. A local central limit theorem for triangles in a random graph. *Random Structures Algorithms* **48**, 732–750. 2016. 1

[5] Svante Janson. A functional limit theorem for random graphs with applications to subgraph count statistics. *Random Structures Algorithms* **1**, 15–37. 1990. 2

[6] Svante Janson. A graph Fourier transform and proportional graphs. In "Proceedings of the Sixth International Seminar on Random Graphs and Probabilistic Methods in Combinatorics and Computer Science (Poznań, 1993)", *Random Structures Algorithms* **6**, 341–351. 1995. 23

[7] Svante Janson. Orthogonal decompositions and functional limit theorems for random graph statistics. *Mem. Amer. Math. Soc.* **111**, no. 534. 1994. 2, 3, 23

[8] Svante Janson and Jan Kratochvíl. Proportional graphs. *Random Structures Algorithms* **2**, 209–224. 1991. 23

[9] Svante Janson and Joel Spencer. Probabilistic construction of proportional graphs. *Random Structures Algorithms* **3**, 127–137. 1992. 23

[10] Jan Kärrman. An example of a superproportional graph. In "Proceedings of the Fifth International Seminar on Random Graphs and Probabilistic Methods in Combinatorics and Computer Science (Poznań, 1991)", *Random Structures Algorithms* **5**, 95–98. 1994. 23

[11] Jan Kärrman. Existence of proportional graphs. *J. Graph Theory* **17**, 207–220. 1993. 23

[12] Elizabeth Meckes. On Stein's method for multivariate normal approximation. In *High Dimensional Probability V: The Luminy Volume*, 153–178. Edited by Christian Houdré, Vladimir Koltchinskii, David M. Mason, and Magda Peligrad. Inst. Math. Statist., 2009. 3, 14

[13] Krzysztof Nowicki. Asymptotic normality of graph statistics. *J. Statist. Plann. Inference* **21**, 209–222. 1989. Originally a 1985 University of Lund technical report. 1, 2

[14] Krzysztof Nowicki and John C. Wierman. Subgraph counts in random graphs using incomplete *U*-statistics methods. In "Proceedings of the First Japan Conference on Graph Theory and Applications (Hakone, 1986)", *Discrete Math.* **72**, 299–310. 1988. 1

[15] Ryan O'Donnell. *Analysis of Boolean functions.* Cambridge University Press, 2014. 12

[16] Alexander A. Razborov. Flag algebras. *J. Symbolic Logic* **72**, 1239–1282. 2007. 3

[17] Andrzej Ruciński. When are small subgraphs of a random graph normally distributed? *Probab. Theory Related Fields* **78**, 1–10. 1988. 1

[18] Ashwin Sah and Mehtaab Sawhney. Local limit theorems for subgraph counts. *J. Lond. Math. Soc.* 2nd ser. **105**, 950–1011. 2022. 1, 3, 12