COMPOSITIO MATHEMATICA

Subgraph distributions in dense random regular graphs

Ashwin Sah and Mehtaab Sawhney

Compositio Math. **159** (2023), 2125–2148.

doi:10.1112/S0010437X23007364

FOUNDATION COMPOSITIO MATHEMATICA

LONDON MATHEMATICAL SOCIETY
EST. 1865

https://doi.org/10.1112/S0010437X23007364 Published online by Cambridge University Press

Compositio Math. 159 (2023) 2125–2148
doi:10.1112/S0010437X23007364

Subgraph distributions in dense random regular graphs

Ashwin Sah and Mehtaab Sawhney

Abstract

Given a connected graph H which is not a star, we show that the number of copies of H in a dense uniformly random regular graph is asymptotically Gaussian, which was not known even for H being a triangle. This addresses a question of McKay from the 2010 International Congress of Mathematicians. In fact, we prove that the behavior of the variance of the number of copies of H depends in a delicate manner on the occurrence and number of cycles of $3, 4, 5$ edges as well as paths of 3 edges in H. More generally, we provide control of the asymptotic distribution of certain statistics of bounded degree which are invariant under vertex permutations, including moments of the spectrum of a random regular graph. Our techniques are based on combining complex-analytic methods due to McKay and Wormald used to enumerate regular graphs with the notion of graph factors developed by Janson in the context of studying subgraph counts in $\mathbb{G}(n, p)$.

1. Introduction

The study of the asymptotic distribution of small subgraph counts in the Erdős–Rényi random graphs $\mathbb{G}(n, p)$ and $\mathbb{G}(n, m)$ has been a topic of central interest in random graph theory. In particular, following a long series of papers, Ruciński [Ruc88] established the optimal conditions under which X_H, the number of unlabeled copies of H in $\mathbb{G}(n, p)$, satisfies a central limit theorem. Furthermore, in general the distribution of small subgraphs in $\mathbb{G}(n, p)$ is known to a substantial degree of precision. We refer the reader, in particular, to the book [JŁR00] and references therein for a more complete account.

With regards to asymptotic distributions, the state of affairs for random d-regular graphs is substantially less satisfactory. Let $\mathbb{G}(n, d)$ denote a uniformly random d-regular graph. Note that unlike $\mathbb{G}(n, p)$ or $\mathbb{G}(n, m)$, the edges in $\mathbb{G}(n, d)$ exhibit strong and non-obvious correlations and therefore even the question of determining the number of d-regular graphs has a rich history drawing on techniques ranging from switchings developed by McKay [McK85] (and refined by McKay and Wormald [MW91]), a complex-analytic technique of McKay and Wormald [MW90], and recent breakthroughs using fixed-point iteration due to Liebenau and Wormald [LW17]. We refer the reader to the excellent survey of Wormald [Wor18] where the extensive history of this problem and various related enumeration problems are discussed.

Received 6 September 2022, accepted in final form 3 May 2023, published online 25 August 2023.
2020 Mathematics Subject Classification 60F05, 05C80, 05A16 (primary).
Keywords: random regular graphs, central limit theorem.
Sah and Sawhney were supported by NSF Graduate Research Fellowship Program DGE-2141064. Sah was supported by the PD Soros Fellowship.

McKay [McK10] in his 2010 International Congress of Mathematicians (ICM) survey on graphs with a fixed degree sequence asked for an understanding of the asymptotic distribution of subgraph counts in dense random regular graphs, noting that 'there is almost nothing known about the distribution of subgraph counts' for these models; the state of affairs has remained unchanged since. In particular, the only result which applies in this regime is work of McKay [McK11] which computes the expectation of the number of subgraphs of a fixed size in $\mathbb{G}(n,d)$ (see [IM18] for an extension to more exotic degree sequences). Our main result establishes a central limit theorem for counting copies of connected graphs H in $\mathbb{G}(n,d)$ for $\min(d,n-d)\ge n/\log n$, and a consequence of our methods demonstrates a joint central limit theorem for so-called 'graph factors' in the sense of Janson [Jan94b]. We additionally apply our techniques to show an analogous result for moments of the spectrum.

Though such a result for dense graphs has until now been out of reach, there is a rich literature regarding sparser graphs. A variety of results have been proven based on applications of the moment method and taking sufficiently fast growing moments. When d is constant, the cycle count distribution was shown to asymptotically converge to a Poisson distribution independently by Bollobás [Bol80] and Wormald [Wor81]. This result was extended to strictly balanced graphs near the threshold for existence by Kim, Sudakov, and Vu [KSV07], establishing a Poisson limit theorem in general. For results regarding asymptotic normality, McKay, Wormald, and Wysocka [MWW04] proved asymptotic normality of cycle counts for d tending to infinity sufficiently slowly, in particular proving normality of triangle counts when $d=o(n^{1/5})$. Gao and Wormald [GW08] later improved this result for a variety of subgraph structures H by counting isolated copies, including extending the regime for triangle counts to $d=o(n^{2/7})$. This was further improved by Gao [Gao20] who improved the range of normality for triangle counts to $d=O(n^{1/2})$. Finally, we note that the study of the asymptotic distribution of the number of spanning structures in random regular graphs has also been of interest (see [Gao20] for further discussion).

Before stating our results let us formally define a random graph with a specified degree sequence.

Definition 1.1. Given nonnegative sequence $\mathbf{d}=(d_1,\ldots,d_n)$, let $\mathbb{G}(\mathbf{d})$ be a uniformly random simple graph G with degree sequence $\mathbf{d}$. When $2\,|\,dn$ let $\mathbb{G}(n,d)$ be a uniformly random simple graph on n vertices which is d-regular, and let $G(n,d)$ be the set of possible outcomes. Given $G\sim\mathbb{G}(n,d)$ we define its density $p=p(G)=e(G)/\binom{v(G)}{2}=d/(n-1)$.

Our results provide a complete understanding of the small subgraph distribution for dense random regular graphs. We first state a corollary of our main result regarding the distribution of subgraph statistics in $\mathbb{G}(n,d)$. Note first that the number of stars with $s\ge 2$ leaves in a d-regular graph on n vertices is trivially always $n\binom{d}{s}$, so we exclude this case from consideration. In addition, given graphs H and F let $N(H,F)$ be the number of unlabeled copies of F in H (or, more precisely, the number of distinct, not necessarily induced, subgraphs of H which are isomorphic to F). In particular, for this definition we have that $N(C_5,P_5)=5$, where we write C_k and P_k for a cycle and path respectively on k vertices.

Theorem 1.2. *Fix a nonempty connected graph H which is not a star and let X_H denote the number of unlabeled copies of H in $G\sim\mathbb{G}(n,d)$. If $n/\log n\le\min(d,n-d)$, $2\,|\,dn$, and $G\sim\mathbb{G}(n,d)$ we have the following.*

- *If H contains a C_3, then*

$$\bigg(\frac{X_H-\mathbb{E}X_H}{\sqrt{\operatorname{Var}[X_H]}}\bigg)\xrightarrow{d.}\mathcal{N}(0,1)$$

https://doi.org/10.1112/S0010437X23007364 Published online by Cambridge University Press

with

$$\mathrm{Var}[X_H] = 6N(H, C_3)^2 p^{2e(H)-3}(1-p)^3 \frac{n^{2v(H)-3}}{\mathrm{aut}(H)^2} + O(n^{2v(H)-3-1/6}).$$

- *If H contains a C_4 and no C_3, then*

$$\left(\frac{X_H - \mathbb{E}X_H}{\sqrt{\mathrm{Var}[X_H]}}\right) \xrightarrow{d.} \mathcal{N}(0,1)$$

with

$$\mathrm{Var}[X_H] = 8N(H, C_4)^2 p^{2e(H)-4}(1-p)^4 \frac{n^{2v(H)-4}}{\mathrm{aut}(H)^2} + O(n^{2v(H)-4-1/6}).$$

- *If H does not contain a C_3 or C_4, then it contains a P_4 and*

$$\left(\frac{X_H - \mathbb{E}[X_H]}{\sqrt{\mathrm{Var}[X_H]}}\right) \xrightarrow{d.} \mathcal{N}(0,1)$$

with

$$\begin{aligned}\mathrm{Var}[X_H] = (&10p^{2e(H)-5}(1-p)^5 N(H, C_5)^2 \\ &+ 6p^{2e(H)-3}(1-p)^3 N(H, P_4)^2)\frac{n^{2v(H)-5}}{\mathrm{aut}(H)^2} + O(n^{2v(H)-5-1/6}).\end{aligned}$$

Remark 1.3. We note that $\mathbb{E}X_H = (1+o(1))n^{v(H)}p^{e(H)}/\,\mathrm{aut(H)}$ is known due to [McK11] and, in fact, our method can be used to compute the expectation to accuracy $o(\sqrt{\mathrm{Var}[X_H]})$, but the resulting expressions are rather involved. In addition, the above result implies that for H being a triangle we have that the variance of X_H is on the order of p^3n^3 for $1/\log n \le p \le 1/2$, whereas in $\mathbb{G}(n,p)$ the variance is of order $\max(p^3n^3, p^5n^4)$. Therefore, for the range of p under consideration $\mathrm{Var}[X_H]$ is substantially lower than in $\mathbb{G}(n,p)$; this is unlike the results of McKay *et al.* [MWW04] and Gao and Wormald [GW08] when p is sufficiently sparse. We also note that for H not containing a triangle, $\mathrm{Var}[X_H]$ is, in fact, asymptotically smaller than the corresponding variance in $\mathbb{G}(n,m)$. Finally, we note that the subgraph counts are not, in general, asymptotically independent.

In general, our results are sufficiently powerful to deduce the asymptotic distribution of statistics of fixed degree which is invariant under vertex permutation. To state our main result we will need the notion of graph factors as defined by Janson [Jan94b]. Let x_e be the indicator random variable for whether an edge e is in included in random graph $G \sim \mathbb{G}(n,d)$ and let $\chi_e = (x_e - p)/\sqrt{p(1-p)}$. Note that, by symmetry, marginally each x_e is distributed as $\mathrm{Ber}(p)$ and, thus, χ_e has mean 0 and variance 1. However, as G is a random regular graph there are substantial correlations between different edges χ_e.

Definition 1.4. Fix a graph H with no isolated vertices and an integer $n \ge |v(H)|$. Then define

$$\gamma_H(\mathbf{x}) = \sum_{\substack{E \subseteq K_n \\ E \simeq H}} \prod_{e \in E} \chi_e.$$

Here $\simeq$ denotes graph isomorphism, specifically between H and the graph spanned by the edges E. We frequently adopt the shorthand that $\chi_S = \prod_{e \in S} \chi_e$. We call $\gamma_H(\mathbf{x})$ the *graph factor* corresponding to the graph H. When H is connected and its minimum degree is at least 2, define

https://doi.org/10.1112/S0010437X23007364 Published online by Cambridge University Press

the *normalized graph factor* to be

$$\widetilde{\gamma}_H(G) = (\gamma_H(\mathbf{x}) - E_H)/\sigma_H$$

where $\sigma_H = (n^{v(H)}/\mathrm{aut}(H))^{1/2}$ and $E_H = 0$ if H is not an even cycle and

$$E_H = \frac{2n^{v(H)/2}}{\mathrm{aut}(H)} = \frac{n^{v(H)/2}}{v(H)}$$

if it is.

Remark 1.5. In the original notion [Jan94b], all connected graphs H are needed to express symmetric functions of graphs. However, as we show in § 4, d-regularity means that γ_H for H with a degree 1 vertex can be expressed as a linear combination of the smaller $\gamma_{H'}$ (in terms of d). In addition, the E_H (approximate expectation) term for even cycles makes an appearance due to regularity. This is another departure from $\mathbb{G}(n,p)$ behavior, since in the independent setting the expectation of every $\gamma_H(\mathbf{x})$ is 0.

We now are in position to state our main result.

Theorem 1.6. *Fix a collection of nonisomorphic connected graphs $\mathcal{H} = \{H_i : 1 \le i \le k\}$ each of minimum degree at least 2. Let $n/\log n \le \min(d, n-d)$, $2 \mid dn$, and $G \sim \mathbb{G}(n,d)$. Then as $n \to \infty$ (uniformly in d), we have*

$$(\widetilde{\gamma}_{H_i}(G))_{1 \le i \le k} \xrightarrow{d.} \mathcal{N}(0,1)^{\otimes k}.$$

Furthermore, $\mathrm{Var}[\gamma_H(G)] = (1 + O(n^{-1/6}))\sigma_H^2$ *and* $\mathbb{E}\gamma_H(G) = E_H + O(n^{-1/6}\sigma_H)$.

Remark 1.7. The convergence in distribution can be made quantitative in terms of Kolmogorov distance or Wasserstein distance by quantifying the convergence in moments (see, e.g., [RW19, Theorem 4]); however, the associated rates are quantitatively quite poor.

Although Theorem 1.6 is stated in terms of graph factors, a straightforward computation allows one to deduce the asymptotic distribution of any symmetric statistic of the edges of bounded degree; one can think of these as the 'building blocks' for all such statistics in $\mathbb{G}(n,d)$.

Our results also imply that the traces of fixed powers of A_G (the adjacency matrix of G), or equivalently the moments of the spectrum, satisfy a joint central limit theorem. Using techniques of Sinai and Soshnikov [SS98] along with a suitable modifications one could likely extend the result to prove normal fluctuations for sufficiently nice test functions (e.g. analytic functions with suitably large radius of convergence). However, given that substantially stronger results are likely plausible using Green's function estimates established by He [He22] and results connecting such estimates with functional central limit theorems (see [LS20] and reference therein), we omit such an extension. In addition, we remark that direct spectral techniques are insufficient to recover Theorem 1.6 since graph factors which do not correspond to cycles are not purely determined by the spectrum.

Corollary 1.8. *Given $k \ge 3$, there exists a positive-definite matrix $\Sigma_k \in \mathbb{R}^{(k-2)\times(k-2)}$ such that the following holds. For $G \sim \mathbb{G}(n,d)$ with $n/\log n \le \min(d, n-d)$ and $2 \mid dn$, $E_\ell = \mathbb{E}(\mathrm{tr}(A_G^\ell))$, and $\sigma_\ell^2 = \mathrm{Var}[\mathrm{tr}(A_G^\ell)]$, we have*

$$(\sigma_\ell^{-1}(\mathrm{tr}(A_G^\ell) - E_\ell))_{3 \le \ell \le k} \xrightarrow{d.} \mathcal{N}(0, \Sigma_k).$$

Remark 1.9. Note that $\mathrm{tr}(A_G) = 0$ and $\mathrm{tr}(A_G^2) = dn$ deterministically.

https://doi.org/10.1112/S0010437X23007364 Published online by Cambridge University Press

1.1 Proof techniques

Our proof uses techniques from the enumeration of dense graphs with a specific degree sequence given by McKay and Wormald [MW90] combined with the notion of graph factors introduced by Janson [Jan94b]. The crucial technical point is that previous work regarding asymptotic normality relied on computing the raw moments of X_H and, therefore, requires taking a number of moments which grows with n. In our approach, one instead notices that any symmetric statistic on d-regular graphs can be expressed in terms of simple building blocks, and we can directly prove a joint central limit theorem for this collection. In order to prove the necessary limit theorem, we only require an arbitrarily slowly growing moment of these graph factors and they are particularly well-behaved when using the complex-analytic techniques developed by McKay and Wormald [MW90]. In particular, the necessary moments of graph factors can be given a natural complex-analytic expression using the multidimensional Cauchy integral formula. Then desired estimates can be computed directly. In fact, certain comparisons to $|G(n, d)|$ simplify the situation, allowing us to avoid repeating a careful saddle point analysis as in the work of McKay and Wormald [MW90]. The nontrivial expectation contributions E_H when H is an even cycle come into play due to counting certain even-power monomials in a polynomial expansion associated to the edges of H.

We further believe combining the general method of considering graph factors along with recent work of Liebenau and Wormald [LW17], which enumerates graphs of degrees of intermediate sparsity, can likely be used to address asymptotic distribution for regular graphs of all sparsities, a direction we plan to pursue in future work. Finally, we note that while Theorem 1.2 can handle subgraph counts of mildly growing size, the understanding of the asymptotic distributions of spanning subgraphs in dense random regular graphs is also of interest. In particular, do analogues of the asymptotic normality results of Janson [Jan94a] regarding the number of perfect matchings in $\mathbb{G}(n, m)$ exist for $\mathbb{G}(n, d)$?

1.2 Organization

In § 2 we prove the main estimates regarding the expectation of χ_S for a fixed set of edges S via contour integration techniques. In § 3, we deduce Theorem 1.6 via the method of moments and a graph-theoretic argument which guarantees that the estimates in § 2 are of sufficient accuracy. In § 4 we develop the theory of graph factors in d-regular graphs and prove that any symmetric graph statistic of fixed degree, when evaluated on d-regular graphs, can be expressed as a polynomial of graph factors that are of the type described in Definition 1.4. Finally in § 5 we deduce Theorem 1.2 and Corollary 1.8 as straightforward consequences of our main results and the proofs in § 4.

1.3 Notation

We use standard asymptotic notation throughout, as follows. For functions $f = f(n)$ and $g = g(n)$, we write $f = O(g)$ or $f \lesssim g$ to mean that there is a constant C such that $|f(n)| \le C|g(n)|$ for sufficiently large n. Similarly, we write $f = \Omega(g)$ or $f \gtrsim g$ to mean that there is a constant $c > 0$ such that $f(n) \ge c|g(n)|$ for sufficiently large n. Finally, we write $f \asymp g$ or $f = \Theta(g)$ to mean that $f \lesssim g$ and $g \lesssim f$, and we write $f = o(g)$ or $g = \omega(f)$ to mean that $f(n)/g(n) \to 0$ as $n \to \infty$. We write $O_H(1)$ for some unspecified constant that can be chosen as some bounded value depending only on H. In addition, we set $k!! = 2^{k/2} \cdot (k/2)!$ for even integers $k \ge 0$. Finally, we let $[n] = \{1, \ldots, n\}$ and $\binom{[n]}{2} = \{(i, j) : 1 \le i < j \le n\}$.

https://doi.org/10.1112/S0010437X23007364 Published online by Cambridge University Press

2. Cancellation estimates based on contour integrals

2.1 Preliminary estimates

We first recall a number of estimates from the work of McKay and Wormald [MW90].

Lemma 2.1 [MW90, Lemma 1]. *Let $0 \le \lambda \le 1$ and $|x| \le \pi$. Then we have that*

$$|1+\lambda(e^{ix}-1)| = (1-2\lambda(1-\lambda)(1-\cos x))^{1/2} \le \exp\big(-\tfrac{1}{2}\lambda(1-\lambda)x^2 + \tfrac{1}{24}\lambda(1-\lambda)x^4\big).$$

Lemma 2.2 [MW90, (3.3)]. *We have for $x_j \in \mathbb{R}$ that*

$$\sum_{1\le j<k\le \ell}(x_j+x_k)^2 \ge (\ell-2)\sum_{1\le j\le \ell} x_j^2, \qquad \sum_{1\le j<k\le \ell}(x_j+x_k)^4 \le 8(\ell-1)\sum_{1\le j\le\ell} x_j^4.$$

We also require the following elementary estimate (a variant of which appears in [MW90, p. 8]); we provide a proof for the sake of completeness.

Lemma 2.3. *We have for $m \ge m_{2.3}$ that*

$$\int_{-\pi/16}^{\pi/16} \exp(-mx^2+mx^4)\,dx = (1\pm 2m^{-1})\sqrt{\pi/m}.$$

Proof. Note that for m larger than an absolute constant,

$$\begin{aligned}\int_{-\pi/16}^{\pi/16} \exp(-mx^2+mx^4)\,dx &= \int_{-m^{-2/5}}^{m^{-2/5}} \exp(-mx^2+mx^4)\,dx \pm \pi/8\cdot\exp(-m^{1/5})\\ &= \int_{-m^{-2/5}}^{m^{-2/5}} (1\pm 2mx^4)\exp(-mx^2)\,dx \pm \pi/8\cdot\exp(-m^{1/5})\\ &= (1\pm 2m^{-1})\sqrt{\pi/m}.\end{aligned}$$

□

We also use an elementary estimate bounding large moments in the following twisted Gaussian integral.

Lemma 2.4. *We have*

$$\int_{-\pi/16}^{\pi/16} |x|^k \exp(-mx^2+mx^4)\,dx \le \sqrt{2\pi}k^{k/2}m^{-(k+1)/2}.$$

Proof. Note

$$\begin{aligned}\int_{-\pi/16}^{\pi/16} |x|^k\exp(-mx^2+mx^4)\,dx &\le \int_{-\infty}^{\infty}|x|^k\exp(-mx^2/2)\,dx\\ &= m^{-(k+1)/2}\int_{-\infty}^{\infty}|x|^k\exp(-x^2/2)\,dx\\ &= \sqrt{2\pi}m^{-(k+1)/2}\mathbb{E}_{Z\sim\mathcal{N}(0,1)}|Z|^k \le k^{k/2}\sqrt{2\pi}m^{-(k+1)/2}.\end{aligned}$$

□

We will need another polynomial inequality in the real numbers.

Lemma 2.5. *For $x_1,\ldots,x_\ell \in \mathbb{R}$ we have*

$$\begin{aligned}k!\sum_{1\le j_1<\cdots<j_k\le\ell} x_{j_1}^2\cdots x_{j_k}^2 &\le \Big(\sum_{1\le j\le\ell} x_j^2\Big)^k\\ &\le k!\sum_{1\le j_1<\cdots<j_k\le\ell} x_{j_1}^2\cdots x_{j_k}^2 + \binom{k}{2}\Big(\max_{j\in[\ell]} x_j^2\Big)\Big(\sum_{1\le j\le\ell}x_j^2\Big)^{k-1}.\end{aligned}$$

https://doi.org/10.1112/S0010437X23007364 Published online by Cambridge University Press

Proof. The first inequality is trivial. For the second, consider expanding $(\sum_{j=1}^{\ell} x_j^2)^k$ and removing the terms which have no repeated index. For the remaining terms, remove the first term in the sequence that later repeats and bound it by $\max_{j\in[\ell]} x_j^2$. It is easy to check that the resulting map on index sequences has fibers of size at most $\binom{k}{2}$. □

Finally, we require the main result of [MW90, Theorem 1] which provides a sharp estimate for $|G(n,d)|$.

THEOREM 2.6. *There exists $\varepsilon = \varepsilon_{2.6} > 0$ such that for $n/\log n \le \min(d, n-d)$, $2|dn$, $\lambda = d/(n-1)$, and $r = \sqrt{\lambda/(1-\lambda)}$, we have*

$$|G(n,d)| = 2^{1/2}(2\pi\lambda^{d+1}(1-\lambda)^{n-d}n)^{-n/2}\exp\left(\frac{-1+10\lambda-10\lambda^2}{12\lambda(1-\lambda)} + O(n^{-\varepsilon})\right)$$
$$= \frac{(1+r^2)^{\binom{n}{2}}}{(2\pi r^d)^n}\left(\frac{2\pi}{\lambda(1-\lambda)n}\right)^{n/2}\left(2^{1/2}\exp\left(\frac{-1+10\lambda-10\lambda^2}{12\lambda(1-\lambda)} + O(n^{-\varepsilon})\right)\right).$$

2.2 Graph factor estimates

The crucial estimates for the remainder of the proof will be the following inequalities controlling the behavior of the constituent expectations in a graph factor.

PROPOSITION 2.7. *There is $C = C_{2.7} > 0$ so that for a set of distinct edges $S \subseteq K_n$ the following holds. Let $p = d/(n-1)$, $n/\log n \le \min(d, n-d)$, and $2|dn$. Recall the notation χ_S from Definition 1.4.*

- *For any S such that $|S| \le \sqrt{\log n}$ we have*

$$|\mathbb{E}_{G\sim\mathbb{G}(n,d)}\chi_S| \le Cn^{-|S|/2+1/4}.$$

- *For S such that $|S| \le \sqrt{\log n}$, and there is a connected component which is an odd cycle, we have*

$$|\mathbb{E}_{G\sim\mathbb{G}(n,d)}\chi_S| \le Cn^{-1/4}n^{-|S|/2}.$$

- *For S such that $|S| \le \sqrt{\log n}$, the set of edges form a set of vertex disjoint even cycles, and there are ℓ disjoint cycles, we have*

$$|\mathbb{E}_{G\sim\mathbb{G}(n,d)}\chi_S - 2^{\ell}n^{-|S|/2}| \le Cn^{-1/5}n^{-|S|/2}.$$

As mentioned previously, the initial reduction in the proof closely mimics that of the proof of [MW90, Theorem 1].

Proof. Note that by complementing $\mathbb{G}(n,d)$ and replacing p by $1-p$, we have that

$$\mathbb{E}_{G\sim\mathbb{G}(n,d)}\chi_S = (-1)^{|S|}\mathbb{E}_{G\sim\mathbb{G}(n,n-d-1)}\chi_S.$$

Therefore, it suffices to treat the case where $p \le 1/2$.

https://doi.org/10.1112/S0010437X23007364 Published online by Cambridge University Press

By Cauchy's integral formula and taking the contours for z_j to be circles of radius $r = \sqrt{p/(1-p)}$ around the origin we have

$$\mathbb{E}_{G\sim\mathbb{G}(n,d)}\chi_S = \frac{(2\pi i)^{-n}}{|G(n,d)|}\oint\cdots\oint \frac{\prod_{(j,k)\notin S}(1+z_jz_k)\prod_{(j,k)\in S}(-p+(1-p)z_jz_k)/\sqrt{p(1-p)}}{\prod_{j\in[n]}z_j^{d+1}}dz$$
$$= \frac{(1+r^2)^{\binom{n}{2}}}{(2\pi r^d)^n|G(n,d)|}\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}\frac{\prod_{(j,k)\notin S}(1+p(e^{i(\theta_j+\theta_k)}-1))\prod_{(j,k)\in S}(p(1-p))^{1/2}(e^{i(\theta_j+\theta_k)}-1)}{\exp(id\sum_{j\in[n]}\theta_j)}d\theta, \tag{2.1}$$

where $dz = \prod_{j\in[n]}dz_j$ and $d\theta = \prod_{j\in[n]}d\theta_j$, and the product is over unordered pairs (j,k) which can be thought of as edges of the complete graph K_n.

Step 1: localizing θ. As in [MW90, Theorem 1], which corresponds to the case $S=\emptyset$, the first maneuver is to localize near the origin, and the techniques are similar. Let $t=\pi/8$ and fix ε to be a small numerical constant to be chosen later ($\varepsilon = 10^{-10}$ will suffice). We divide indices based on where they lie on the circle: $S_1 = \{j\colon \theta_j\in[-t,t]\}$, $S_2=\{j\colon \theta_j\in[t,\pi-t]\}$, $S_3=\{j\colon\theta_j\in[\pi-t,\pi]\cup[-\pi,-\pi+t]\}$, and $S_4=\{j\colon\theta_j\in[-\pi+t,-t]\}$. Let $\mathbf{R}$ denote the set of θ such at least one of $|S_1||S_3|\ge n^{1+\varepsilon}$, $|S_2|^2\ge n^{1+\varepsilon}$, or $|S_4|^2\ge n^{1+\varepsilon}$ holds. We have by Lemma 2.1 that

$$\left|\int_{\mathbf{R}}\frac{\prod_{(j,k)\notin S}(1+p(e^{i(\theta_j+\theta_k)}-1))\prod_{(j,k)\in S}(p(1-p))^{1/2}(e^{i(\theta_j+\theta_k)}-1)}{\exp(id\sum_{j\in[n]}\theta_j)}d\theta\right|$$
$$\le\int_{\mathbf{R}}(1-2p(1-p)(1-\cos(2t)))^{n^{1+\varepsilon}/3-|S|}d\theta\le\exp(-\Omega(n^{1+\varepsilon/2})) \tag{2.2}$$

and, therefore, it will suffice to consider $\theta\notin\mathbf{R}$. Thus, $|S_2|,|S_4|\le n^{1/2+\varepsilon/2}$. Furthermore, note that $\theta\notin\mathbf{R}$ implies that $|S_1|\le n^{\varepsilon}$ or $|S_3|\le n^{\varepsilon}$. As the integrand is invariant under $\theta\to\theta+\pi$ (since $2|dn$) it suffices to consider when $|S_3|\le n^{\varepsilon}$ and multiply the resulting integral by a factor of 2.

Let $\mathbf{R}'$ denote the set of θ such that $\theta\notin\mathbf{R}$, $|S_3|\le n^{\varepsilon}$, and there is $\theta_j\notin[-n^{-1/2+\varepsilon},n^{-1/2+\varepsilon}]$. We have

$$\left|\int_{\theta\in\mathbf{R}'}\frac{\prod_{(j,k)\notin S}(1+p(e^{i(\theta_j+\theta_k)}-1))\prod_{(j,k)\in S}(p(1-p))^{1/2}(e^{i(\theta_j+\theta_k)}-1)}{\exp(id\sum_{j\in[n]}\theta_j)}d\theta\right|$$
$$\le\int_{\theta\in\mathbf{R}'}\prod_{(j,k)\notin S}|1+p(e^{i(\theta_j+\theta_k)}-1)|d\theta\le e^{O(|S|)}\left(\frac{2\pi}{\lambda(1-\lambda)n}\right)^{n/2}\exp(-\Omega(n^{\varepsilon})), \tag{2.3}$$

where we used a slight modification of [MW90, (3.4), (3.5)] in the second inequality (namely, the analogy to the intermediate upper bound given in [MW90] is multiplicatively stable with respect to removal of the terms corresponding to $(j,k)\in S$).

https://doi.org/10.1112/S0010437X23007364 Published online by Cambridge University Press

Finally, let $\mathbf{U}$ denote the set of θ such that $|\theta_j| \le n^{-1/2+\varepsilon}$ for all j. Combining (2.1) to (2.3), the above symmetry observation, and Theorem 2.6 yields

$$\begin{aligned}&\mathbb{E}_{G\sim\mathbb{G}(n,d)}\chi_S \pm \exp(-\Omega(n^\varepsilon))\\ &\quad= \frac{2(1+r^2)^{\binom{n}{2}}}{(2\pi r^d)^n|G(n,d)|}\int_{\mathbf{U}} \frac{\prod_{(j,k)\notin S}(1+p(e^{i(\theta_j+\theta_k)}-1))\prod_{(j,k)\in S}(p(1-p))^{1/2}(e^{i(\theta_j+\theta_k)}-1)}{\exp(id\sum_{j\in[n]}\theta_j)}d\theta.\end{aligned} \tag{2.4}$$

Step 2: reducing the S contribution to a polynomial. We next apply a Taylor series transformation in order to reduce to a more symmetric integral where the terms depending on S are polynomial factors within the integrand. First, note that if θ_j,θ_k are sufficiently small, then

$$|(e^{i(\theta_j+\theta_k)}-1)(1+p(e^{i(\theta_j+\theta_k)}-1))^{-1} - i(\theta_j+\theta_k)| \le |\theta_j+\theta_k|^2.$$

Therefore, for $\theta\in\mathbf{U}$, we have

$$\bigg|\prod_{(j,k)\in S}(e^{i(\theta_j+\theta_k)}-1)(1+p(e^{i(\theta_j+\theta_k)}-1))^{-1} - \prod_{(j,k)\in S} i(\theta_j+\theta_k)\bigg| \le 2^{|S|}(2n^{-1/2+\varepsilon})\prod_{(j,k)\in S}|\theta_j+\theta_k|.$$

Define

$$P_1(\theta) = \prod_{(j,k)\notin S}(1+p(e^{i(\theta_j+\theta_k)}-1))\prod_{(j,k)\in S}(e^{i(\theta_j+\theta_k)}-1)$$

and

$$P_2(\theta) = \prod_{(j,k)\in\binom{[n]}{2}}(1+p(e^{i(\theta_j+\theta_k)}-1))\prod_{(j,k)\in S}(i(\theta_j+\theta_k)).$$

Note the above analysis implies

$$\begin{aligned}&\bigg|\int_{\mathbf{U}}\frac{P_1(\theta)-P_2(\theta)}{\exp(id\sum_{j\in[n]}\theta_j)}d\theta\bigg|\\ &\quad\le 2^{|S|}\int_{\mathbf{U}} 2n^{-1/2+\varepsilon}\prod_{(j,k)\in S}|\theta_j+\theta_k|\prod_{(j,k)\in\binom{[n]}{2}}|1+p(e^{i(\theta_j+\theta_k)}-1)|d\theta\\ &\quad\le 8^{|S|}n^{-1/2+\varepsilon}\int_{\mathbf{U}}|\theta_1|^{|S|}\prod_{(j,k)\in\binom{[n]}{2}}|1+p(e^{i(\theta_j+\theta_k)}-1)|d\theta\\ &\quad\le 8^{|S|}n^{-1/2+\varepsilon}\int_{\mathbf{U}}|\theta_1|^{|S|}\prod_{(j,k)\in\binom{[n]}{2}}\exp\bigg(-\frac{1}{2}p(1-p)(\theta_j+\theta_k)^2+\frac{1}{24}p(1-p)(\theta_j+\theta_k)^4\bigg)d\theta\\ &\quad\le 8^{|S|}n^{-1/2+\varepsilon}\int_{\mathbf{U}}|\theta_1|^{|S|}\exp\bigg(\sum_{1\le j\le n}-(n-2)\frac{p(1-p)}{2}\theta_j^2+(n-1)\frac{p(1-p)}{3}\theta_j^4\bigg)d\theta\\ &\quad\le 8^{|S|}n^{-1/2+\varepsilon}\int_{\mathbf{U}}|\theta_1|^{|S|}\exp\bigg(\sum_{1\le j\le n}-(n-2)\frac{p(1-p)}{2}\theta_j^2+(n-2)\frac{p(1-p)}{2}\theta_j^4\bigg)d\theta\\ &\quad\lesssim 16^{|S|}n^{-1/2+\varepsilon}n^{-|S|/2}|S|^{|S|/2}(2\pi/(p(1-p)n))^{n/2},\end{aligned}$$

https://doi.org/10.1112/S0010437X23007364 Published online by Cambridge University Press

where we have applied Lemmas 2.1–2.4. By (2.4) and Theorem 2.6 it follows that

$$\mathbb{E}_{G\sim\mathbb{G}(n,d)}\chi_S = \frac{2(1+r^2)^{\binom{n}{2}}}{(2\pi r^d)^n|G(n,d)|}\int_{\mathbf{U}}\frac{P_2(\theta)(p(1-p))^{|S|/2}}{\exp(id\sum_{j\in[n]}\theta_j)}d\theta \pm n^{-(|S|+1)/2+2\varepsilon}. \tag{2.5}$$

Step 3: uniform bound on the integral. We now prove the first bullet point in Proposition 2.7. Note that Lemmas 2.1 and 2.2 give

$$\begin{aligned}
&\bigg|\int_{\mathbf{U}}\frac{P_2(\theta)}{\exp(id\sum_{j\in[n]}\theta_j)}d\theta\bigg|\\
&\le \int_{\mathbf{U}}\prod_{(j,k)\in S}|\theta_j+\theta_k|\prod_{(j,k)\in\binom{[n]}{2}}|1+p(e^{i(\theta_j+\theta_k)}-1)|d\theta\\
&\le 2^{|S|}\int_{\mathbf{U}}|\theta_1|^{|S|}\prod_{(j,k)\in\binom{[n]}{2}}\exp\bigg(-\frac{1}{2}p(1-p)(\theta_j+\theta_k)^2+\frac{1}{24}p(1-p)(\theta_j+\theta_k)^4\bigg)d\theta\\
&\le 2^{|S|}\int_{\mathbf{U}}|\theta_1|^{|S|}\exp\bigg(\sum_{1\le j\le n}-(n-2)\frac{p(1-p)}{2}\theta_j^2+(n-1)\frac{p(1-p)}{3}\theta_j^4\bigg)d\theta\\
&\le 2^{|S|}\int_{\mathbf{U}}|\theta_1|^{|S|}\exp\bigg(\sum_{1\le j\le n}-(n-2)\frac{p(1-p)}{2}(\theta_j^2-\theta_j^4)\bigg)d\theta\\
&\lesssim n^{-|S|/2}(4|S|)^{|S|}(2\pi/(p(1-p)n)))^{n/2},
\end{aligned}$$

which immediately gives the desired initial estimate noting that the final term in enumeration count from Theorem 2.6 is bounded by $n^{1/5}$ and since $|S|$ is small.

Step 4: cancellation from odd degree terms. We next prove that any polynomial factor in terms of the θ coefficients which is not an even polynomial exhibits additional cancellation. This will immediately imply the second bullet point as there are at most $2^{|S|}$ terms in $\prod_{(j,k)\in S}(\theta_j+\theta_k)$ and since there is an odd cycle component (implying every term has an index of degree 1). In particular, it suffices to bound

$$\int_{\mathbf{U}}\frac{\prod_{j\in[k]}\theta_j^{\ell_j}\prod_{(j,k)\in\binom{[n]}{2}}(1+p(e^{i(\theta_j+\theta_k)}-1))}{\exp(id\sum_{j\in[n]}\theta_j)}d\theta,$$

where $\ell_k=1$, and $k\le 2|S|$. For this, note that by symmetry

$$\begin{aligned}
&\bigg|\int_{\mathbf{U}}\frac{\prod_{j\in[k]}\theta_j^{\ell_j}\prod_{(j,k)\in\binom{[n]}{2}}(1+p(e^{i(\theta_j+\theta_k)}-1))}{\exp(id\sum_{j\in[n]}\theta_j)}d\theta\bigg|\\
&=\frac{1}{n-k+1}\bigg|\int_{\mathbf{U}}\frac{(\sum_{k\le j\le n}\theta_j)\prod_{j\in[k-1]}\theta_j^{\ell_j}\prod_{(j,k)\in\binom{[n]}{2}}(1+p(e^{i(\theta_j+\theta_k)}-1))}{\exp(id\sum_{j\in[n]}\theta_j)}d\theta\bigg|\\
&\le\frac{1}{n-k+1}\int_{\mathbf{U}}\bigg|\sum_{k\le j\le n}\theta_j\bigg|\prod_{j\in[k-1]}|\theta_j|^{\ell_j}\prod_{(j,k)\in\binom{[n]}{2}}|1+p(e^{i(\theta_j+\theta_k)}-1)|d\theta\\
&\le\frac{2}{n}\int_{\mathbf{U}}\bigg|\sum_{k\le j\le n}\theta_j\bigg|\prod_{j\in[k-1]}|\theta_j|^{\ell_j}\exp\bigg(\sum_{1\le j\le n}-(n-2)\frac{p(1-p)}{2}(\theta_j^2-\theta_j^4)\bigg)d\theta
\end{aligned}$$

https://doi.org/10.1112/S0010437X23007364 Published online by Cambridge University Press

$$
\begin{aligned}
&= \frac{2}{n}\int_{\mathbf{U}} \mathbb{E}_{s\sim\mathrm{Rad}^{\otimes n}}\bigg|\sum_{k\le j\le n} s_j\theta_j\bigg| \prod_{j\in[k-1]}|\theta_j|^{\ell_j}\exp\bigg(\sum_{1\le j\le n} -(n-2)\frac{p(1-p)}{2}(\theta_j^2-\theta_j^4)\bigg)d\theta\\
&\le \frac{2}{n}\int_{\mathbf{U}} \bigg(\mathbb{E}_{s\sim\mathrm{Rad}^{\otimes n}}\bigg(\sum_{k\le j\le n} s_j\theta_j\bigg)^2\bigg)^{1/2} \prod_{j\in[k-1]}|\theta_j|^{\ell_j}\exp\bigg(\sum_{1\le j\le n} -(n-2)\frac{p(1-p)}{2}(\theta_j^2-\theta_j^4)\bigg)d\theta\\
&\le \frac{2}{n}\int_{\mathbf{U}} \bigg(\sum_{k\le j\le n}\theta_j^2\bigg)^{1/2} \prod_{j\in[k-1]}|\theta_j|^{\ell_j}\exp\bigg(\sum_{1\le j\le n} -(n-2)\frac{p(1-p)}{2}(\theta_j^2-\theta_j^4)\bigg)d\theta\\
&\le \frac{2}{n^{1-\varepsilon}}\int_{\mathbf{U}} \prod_{j\in[k-1]}|\theta_j|^{\ell_j}\exp\bigg(\sum_{1\le j\le n} -(n-2)\frac{p(1-p)}{2}(\theta_j^2-\theta_j^4)\bigg)d\theta\\
&\lesssim n^{-\sum_{j\in[k]}\ell_j/2-1/2+2\varepsilon}(2\pi/(p(1-p)n)))^{n/2}
\end{aligned}
$$

as desired, where in the last line we apply Lemma 2.4 and use that $|S|$ is small.

Step 5: even cycles. We now handle the third bullet point, proving that the integral is sufficiently close to the desired quantity. Using the technique in the previous step, and noting that given a set S of ℓ disjoint even cycles there are 2^ℓ terms in the expansion of $\prod_{(j,k)\in S}(\theta_j+\theta_k)$ where every vertex has even degree, we have

$$
\begin{aligned}
&\bigg|\mathbb{E}_{G\sim\mathbb{G}(n,d)}\chi_S - \frac{2^{\ell+1}(1+r^2)^{\binom{n}{2}}(p(1-p))^{|S|/2}}{(2\pi r^d)^n|G(n,d)|}\\
&\qquad\times\int_{\mathbf{U}}\frac{\prod_{j\in[|S|/2]}\theta_j^2\prod_{(j,k)\in\binom{[n]}{2}}(1+p(e^{i(\theta_j+\theta_k)}-1))}{\exp(id\sum_{j\in[n]}\theta_j)}d\theta\bigg|\\
&\lesssim n^{-1/4-|S|/2}.
\end{aligned}
\tag{2.6}
$$

Note that $\mathbb{E}_{G\sim\mathbb{G}(n,d)}\chi_\emptyset = 1$ by definition and (2.6) applies with S empty. Subtracting, it therefore suffices to prove

$$
\begin{aligned}
&\bigg|\frac{2^{\ell+1}(1+r^2)^{\binom{n}{2}}}{(2\pi r^d)^n|G(n,d)|}\int_{\mathbf{U}}\frac{(\prod_{j\in[|S|/2]}\theta_j^2-(p(1-p)n)^{-|S|/2})\prod_{(j,k)\in\binom{[n]}{2}}(1+p(e^{i(\theta_j+\theta_k)}-1))}{\exp(id\sum_{j\in[n]}\theta_j)}d\theta\bigg|\\
&\quad\lesssim n^{-1/4-|S|/2}.
\end{aligned}
$$

From Lemma 2.5 we have

$$
\frac{(\sum_j x_j^2)^k-k^2(\max_j x_j^2)(\sum_j x_j^2)^{k-1}}{k!}\le\sum_{j_1<\cdots<j_k}x_{j_1}^2\cdots x_{j_k}^2\le\frac{(\sum_j x_j^2)^k}{k!},
$$

and using our initial bounds from earlier it follows immediately that

$$
\begin{aligned}
&\bigg|\frac{2^{\ell+1}(1+r^2)^{\binom{n}{2}}\binom{n}{|S|/2}^{-1}}{(2\pi r^d)^n|G(n,d)|(|S|/2)!}\\
&\qquad\times\int_{\mathbf{U}}\frac{((|S|/2)^2n^{-1+2\varepsilon}(\sum_j\theta_j^2)^{|S|/2-1})\prod_{(j,k)\in\binom{[n]}{2}}(1+p(e^{i(\theta_j+\theta_k)}-1))}{\exp(id\sum_{j\in[n]}\theta_j)}d\theta\bigg|\\
&\lesssim n^{-1/4-|S|/2}.
\end{aligned}
$$

https://doi.org/10.1112/S0010437X23007364 Published online by Cambridge University Press

Therefore, symmetrizing over all permutations of $[n]$ and trivially bounding some lower-order contributions, we see it suffices to bound

$$\frac{2^{\ell+1}(1+r^2)^{\binom{n}{2}}\binom{n}{|S|/2}^{-1}}{(2\pi r^d)^n|G(n,d)|(|S|/2)!} \times \int_{\mathbf{U}} \frac{((\sum_{1\le j\le n}\theta_j^2)^{|S|/2}-(p(1-p))^{-|S|/2})\prod_{(j,k)\in\binom{[n]}{2}}(1+p(e^{i(\theta_j+\theta_k)}-1))}{\exp(id\sum_{j\in[n]}\theta_j)}\,d\theta.$$

Note that, once again,

$$\begin{aligned}
&\int_{\mathbf{U}} \frac{((\sum_{1\le j\le n}\theta_j^2)^{|S|/2}-(p(1-p))^{-|S|/2})\prod_{(j,k)\in\binom{[n]}{2}}(1+p(e^{i(\theta_j+\theta_k)}-1))}{\exp(id\sum_{j\in[n]}\theta_j)}\,d\theta\\
&\le \int_{\mathbf{U}} \bigg|\bigg(\sum_{1\le j\le n}\theta_j^2\bigg)^{|S|/2}-(p(1-p))^{-|S|/2}\bigg|\prod_{(j,k)\in\binom{[n]}{2}}|1+p(e^{i(\theta_j+\theta_k)}-1)|\,d\theta\\
&\le \int_{\mathbf{U}} \bigg|\bigg(\sum_{1\le j\le n}\theta_j^2\bigg)^{|S|/2}-(p(1-p))^{-|S|/2}\bigg|\\
&\qquad\times\exp\bigg(\sum_{1\le j\le n}-(n-2)\frac{p(1-p)}{2}(\theta_j^2-\theta_j^4)\bigg)\,d\theta. \qquad (2.7)
\end{aligned}$$

We now proceed via splitting (2.7) based on the size of $\sum_{1\le j\le n}\theta_j^2$. Defining the region $\mathbf{S}=\{\theta\colon |\sum_{1\le j\le n}\theta_j^2-(p(1-p))^{-1}|\ge n^{-1/3}\}$ we have that

$$\begin{aligned}
&\int_{\mathbf{U}} \mathbb{1}_{\theta\in\mathbf{S}}\bigg|\bigg(\sum_{1\le j\le n}\theta_j^2\bigg)^{|S|/2}-(p(1-p))^{-|S|/2}\bigg|\exp\bigg(\sum_{1\le j\le n}-(n-2)\frac{p(1-p)}{2}(\theta_j^2-\theta_j^4)\bigg)\,d\theta\\
&\le \int_{\mathbf{U}} \mathbb{1}_{\theta\in\mathbf{S}}(2n^{2\varepsilon})^{|S|/2}\exp\bigg(\sum_{1\le j\le n}-(n-2)\frac{p(1-p)}{2}(\theta_j^2-\theta_j^4)\bigg)\,d\theta\\
&\le \int_{\mathbf{U}} \mathbb{1}_{\theta\in\mathbf{S}}(2n^{2\varepsilon})^{|S|/2}\exp\bigg(\sum_{1\le j\le n}\frac{-(n-2n^{2\varepsilon})p(1-p)}{2}\theta_j^2\bigg)\,d\theta\\
&\le (2n^{2\varepsilon})^{|S|/2}(1/(p(1-p)(n-2n^{2\varepsilon})))^{n/2}\int_{\mathbb{R}^n}\mathbb{1}_{|\sum_{j\in[n]}x_j^2-(n-2n^{2\varepsilon})|\ge p(1-p)n^{2/3}/2}\\
&\qquad\times\exp\bigg(-\frac{1}{2}\sum_{j\in[n]}x_j^2\bigg)dx\\
&\le \exp(O(n^{2\varepsilon}))\cdot(2\pi/(p(1-p)n))^{n/2}\mathbb{P}_{Z\sim\mathcal{N}(0,1)^{\otimes n}}\bigg[\bigg|\sum_{1\le j\le n}Z_j^2-n\bigg|\ge n^{3/5}\bigg]\\
&\le \exp(n^{-1/10})\cdot(2\pi/(p(1-p)n))^{n/2},
\end{aligned}$$

https://doi.org/10.1112/S0010437X23007364 Published online by Cambridge University Press

which is sufficiently small as desired. For the remaining portion of (2.7) note that

$$\begin{aligned}
&\int_{\mathbf{U}} \mathbb{1}_{\theta\notin\mathbf{S}}\bigg|\bigg(\sum_{1\le j\le n}\theta_j^2\bigg)^{|S|/2} - (p(1-p))^{|S|/2}\bigg| \exp\bigg(\sum_{1\le j\le n} -(n-2)\frac{p(1-p)}{2}(\theta_j^2-\theta_j^4)\bigg)\,d\theta \\
&\quad\le \int_{\mathbf{U}} \mathbb{1}_{\theta\notin\mathbf{S}}(4p(1-p))^{|S|/2}(n^{-1/3})\exp\bigg(\sum_{1\le j\le n} -(n-2)\frac{p(1-p)}{2}(\theta_j^2-\theta_j^4)\bigg)\,d\theta \\
&\quad\le \int_{\mathbf{U}} (4p(1-p))^{|S|/2}(n^{-1/3})\exp\bigg(\sum_{1\le j\le n} -(n-2)\frac{p(1-p)}{2}(\theta_j^2-\theta_j^4)\bigg)\,d\theta \\
&\quad\lesssim n^{-2/7}(2\pi/(p(1-p)n))^{n/2},
\end{aligned}$$

where we have used Lemma 2.3 in the final step. The desired result follows immediately. □

3. Deduction of Theorem 1.6

In order to prove Theorem 1.6 we proceed via the method of moments. We require the following standard result regarding converting control on moments to distributional control. This follows immediately from the standard univariate method of moments via the Cramér–Wold device (see, e.g., [Dur19, Theorem 3.10.6]) which shows that in order to prove convergence of a sequence of random variables $X_n \to \mu \in \mathbb{R}^d$ in distribution, it suffices to prove convergence of $X_n\cdot\theta \to \mu\cdot\theta$ for all $\theta\in\mathbb{R}^d$. (A proof of the univariate case of the method of moments is standard; see e.g. [Dur19, Section 3.3.5, Theorem 3.3.25].)

LEMMA 3.1. *Fix a vector $\mu\in\mathbb{R}^d$ and a positive-definite matrix in $\Sigma\in\mathbb{R}^{d\times d}$. Given a sequence of random vectors $X_n\in\mathbb{R}^d$, suppose that for any sequence of nonnegative integers $(\ell_i)_{1\le i\le d}$ that*

$$\mathbb{E}\bigg[\prod_{i=1}^d ((X_n)_i)^{\ell_i}\bigg] \to \mathbb{E}_{G\sim\mathcal{N}(\mu,\Sigma)}\bigg[\prod_{i=1}^d (G_i)^{\ell_i}\bigg]$$

as $n\to\infty$. Then it follows that

$$X_n \xrightarrow{d.} \mathcal{N}(\mu,\Sigma).$$

We also require the following graph-theoretic input which will be used when applying the method of moments. For a multigraph G, let $E_{\mathrm{sing}}(G)$ be the set of edges of multiplicity 1.

LEMMA 3.2. *Let $\mathcal{H}=(H_i)_{1\le i\le k}$ be a sequence of connected graphs each of minimum degree at least 2 (not necessarily distinct). Consider overlaying the H_i in order to obtain a multigraph G (so overlaying two or more edges would give the corresponding multiplicity in G). Let $E_{\mathrm{sing}} = E_{\mathrm{sing}}(G)$. Then we have*

$$v(G) - \frac{1}{2}|E_{\mathrm{sing}}| \le \frac{1}{2}\sum_{i=1}^k v(H_i)$$

with equality if and only if each connected component of G is either (i) a cycle of multiplicity 1 which is isolated or (ii) a multigraph with all multiplicities 2. Furthermore, in choice (i) the said connected component arises from a single H_i which is a cycle while in choice (ii) the connected component arises from two H_i, H_j which are isomorphic and perfectly overlaid.

Proof. Note that E_{sing} contains precisely the edges of G without multiplicity, which therefore arise from a single graph in $\mathcal{H}$. In addition, it trivially suffices to prove the claim for each

https://doi.org/10.1112/S0010437X23007364 Published online by Cambridge University Press

connected component of G individually. Equivalently, we may assume G is connected. We have

$$\begin{aligned}
&\frac{1}{2}\sum_{i=1}^{k} v(H_i) - v(G) + \frac{|E_{\mathrm{sing}}|}{2}\\
&= \sum_{v\in V(G)}\bigg(\bigg(\frac{1}{2}\sum_{i=1}^{k}\mathbb{1}_{v\in H_i}\bigg) - 1\bigg) + \frac{|E_{\mathrm{sing}}|}{2}\\
&\geq -\frac{1}{2}\sum_{v\in V(G)}\mathbb{1}[|\{i\in[k]\colon v\in H_i\}| = 1] + \frac{|E_{\mathrm{sing}}|}{2}\\
&\geq -\frac{1}{2}\sum_{v\in V(G)}\mathbb{1}[|\{i\in[k]\colon v\in H_i\}| = 1] + \sum_{v\in V(G)}\frac{2\mathbb{1}[|\{i\in[k]\colon v\in H_i\}| = 1]}{4} \geq 0.
\end{aligned}$$

In the last line, the first inequality is justified as follows: consider distributing a mass of $1/2$ on each edge in E_{sing} into $1/4$ on both its vertices. Note that every vertex that appears in exactly one H_i must be contributed by at least 2 such edges, since the minimum degree is at least 2 and such edges clearly must be singletons.

For equality to occur note that every vertex must have $\mathcal{H}$-multiplicity 1 or 2 (i.e. appears in 1 or 2 of the H_i), each singleton edge must occur between two vertices of $\mathcal{H}$-multiplicity 1, and no $\mathcal{H}$-multiplicity 1 vertex has degree larger than 2. Note that as we assumed G is connected, we must have that either all vertices have $\mathcal{H}$-multiplicity 1 or all have $\mathcal{H}$-multiplicity 2: if there is an edge between the two different types of vertex, then it must be a singleton (since one of the endpoints is $\mathcal{H}$-multiplicity 1) and, hence, we have a contradiction to the required property of singleton edges in the equality case. Now, if all vertices are $\mathcal{H}$-multiplicity 1, then G must arise from a single graph H_1, and the equality case is immediately seen to be a cycle using our assumption that $G = H_1$ is connected and also that every vertex has degree exactly 2.

We now focus on the complementary case that every vertex has $\mathcal{H}$-multiplicity 2 and, hence, there are no singleton edges. Consider a vertex v of G and suppose without loss of generality that it is in H_1 and H_2. Every G-neighbor w of v has the property that edge (v, w) is not singleton, which implies this edge must be present in both H_1 and H_2. Thus, w is in H_1 and H_2 (and no other H_i since it has $\mathcal{H}$-multiplicity 2). Iterating this argument, and using that G is connected, we see that every vertex of G is in H_1 and H_2, implying that $k = 2$. Since there are no singleton edges, this must be a direct overlay of equal graphs.

Finally, we easily check that choices (i) and (ii) are easily seen to indeed give equality. □

We now prove Theorem 1.6. Given the results proven so far this is essentially a routine computation with the method of moments.

Proof of Theorem 1.6. Fix a collection of connected graphs $\mathcal{H} = \{H_i\colon 1 \leq i \leq k\}$ of minimum degree at least 2. In order to apply the method of moments consider fixed values $\ell_1, \ldots, \ell_k$ and write

$$\mathbb{E}_{G\sim\mathbb{G}(n,d)}\bigg[\prod_{i=1}^{k}\gamma_{H_i}(G)^{\ell_i}\bigg] = \bigg(\prod_{i=1}^{k}\sigma_{H_i}^{-\ell_i}\bigg)\sum_{\substack{1\leq i\leq k\\ 1\leq j\leq \ell_i\\ H_{i,j}\simeq H_i}}\mathbb{E}_{G\sim\mathbb{G}(n,d)}\bigg[\prod_{i=1}^{k}\prod_{j=1}^{\ell_i}\chi_{H_{i,j}}\bigg]. \tag{3.1}$$

Here the $H_{i,j}$ are embedded into K_n, and we are summing over possible simultaneous choices of such unlabeled copies. Recall the definition of E_H, σ_H from Definition 1.4, and note this means $\prod_{i=1}^{k}\sigma_{H_i}(G)^{\ell_i} = \Theta(n^{\sum_{i=1}^{k}\ell_i v(H_i)/2}) = \Theta(n^{\sum_{i,j} v(H_{i,j})/2})$.

https://doi.org/10.1112/S0010437X23007364 Published online by Cambridge University Press

We consider the terms based on the isomorphism class of $G = \bigcup_{1\le i\le k}\bigcup_{1\le j\le \ell_i} H_{i,j}$, treating G as a multigraph. Let $E_{\mathrm{sing}} = E_{\mathrm{sing}}(G)$, the set of singleton edges in the isomorphism class. Note that the contribution of terms based on G is bounded by $O(n^{v(G)}n^{-|E_{\mathrm{sing}}|/2+1/3})$ using the first bullet of Proposition 2.7 (and using that if an edge is repeated multiple times, the corresponding χ_e^t term can be reduced to a linear combination of $1, \chi_e$ with coefficients depending only on t and p). Thus, if $v(G) - |E_{\mathrm{sing}}|/2 < \sum_{i,j} v(H_{i,j})/2$ the terms contribute negligibly (namely, $O(n^{-1/6})$) to the quantity (3.1), since this implies $v(G) - |E_{\mathrm{sing}}|/2 \le -1/2 + \sum_{i,j} v(H_{i,j})/2$.

However, recall that by Lemma 3.2 we have $v(G) - |E_{\mathrm{sing}}|/2 \le \sum_{i,j} v(H_{i,j})/2$, and equality occurs only in certain specialized cases where each graph $H_{i,j}$ either (i) is a cycle and its vertices are not used by any other $H_{i',j'}$ or (ii) is perfectly overlaid with another $H_{i,j'}$ (with the same isomorphism type) as equal copies. The earlier analysis shows we may restrict attention to such equality cases, so now we more closely characterize which such terms contribute. Without loss of generality, let us assume that $H_1, \ldots, H_m$ are cycles, if any, while $H_{m+1}, \ldots, H_k$ are not cycles. Note that if any ℓ_t for $t \in [m+1, k]$ is odd then it is impossible to pair up and overlay all the $H_{t,i}$ for $i \in [\ell_t]$. This violates the equality condition, so is not possible. Thus, if ℓ_t for some $t \in [m+1, k]$ is odd, then the total contribution to (3.1) is $O(n^{-1/6})$.

Now consider H_t with $1 \le t \le m$. If H_t is an odd cycle and is not overlaid with another, then by case (i) it is isolated within G. The second bullet of Proposition 2.7 again shows the total contribution of terms with such an unpaired H_t to (3.1) is $O(n^{-1/6})$.

Finally, we have a situation where all graphs $H_{i,j}$ except even cycles must be paired among themselves and the even cycles $H_{i,j}$ are either isolated in G or paired with another even cycle $H_{i,j'}$ of the same size and overlaid. Without loss of generality let $H_1, \ldots, H_{m'}$ be the even cycles, if any.

The number of choices for pairing up the graphs other than even cycles is $\prod_{i=m'+1}^{k} \ell_i!!$. The number of choices for pairing up $s_i \le \ell_i/2$ even cycles for $i \in [m']$ is $\prod_{i=1}^{m'} \binom{\ell_i}{2s_i}(2s_i)!!$. In such a pairing, let $\mathcal{U}_i \subseteq [\ell_i]$ be the list of unpaired indices for $i \in [m']$. We find that $\prod_{i=1}^{k}\prod_{j=1}^{\ell_i} \chi_{H_{i,j}}$ is a product of various terms of the form χ_e for $e \in H_{i,j}$ where $i \in [m']$ and $j \in \mathcal{U}_i$, as well as terms of the form χ_e^2 in certain connected components of G. There are $\sum_{i=1}^{m'}(\ell_i - 2s_i)v(H_i)$ vertices of the former type and $\sum_{i=1}^{m'} s_i v(H_i) + \sum_{i=m'+1}^{k}(\ell_i/2)v(H_i)$ of the latter type. Note that $\chi_e^2 = 1 - (2p-1)\chi_e/\sqrt{p(1-p)}$, and expanding out the repeated terms in such a way yields one term of the form $\prod_{i=1}^{m'}\prod_{j\in\mathcal{U}_i} \chi_{H_{i,j}}$ and others with additional terms of the form $(2p-1)\chi_e/\sqrt{p(1-p)}$ tacked on. The contribution of such other terms totals at most, by the first bullet of Proposition 2.7,

$$O\big(n^{\sum_{i=1}^{m'}(\ell_i - s_i)v(H_i) + \sum_{i=m'+1}^{k}(\ell_i/2)v(H_i)} \cdot n^{-\sum_{i=1}^{m'}\sum_{j\in\mathcal{U}_i} e(H_{i,j})/2 - 1/2 + 1/3}\big).$$

The exponent is bounded by $v(G)/2 - 1/3 + \sum_{i=1}^{m'}\sum_{j\in\mathcal{U}_i}(v(H_{i,j}) - e(H_{i,j})/2) = \sum_{i=1}^{k} \ell_i v(H_i)/2 - 1/6$ since $H_{i,j}$ for $i \in [m']$ is a cycle, so in (3.1) this amounts to a total contribution of $O(n^{-1/6})$.

Finally, what remains is

$$\begin{aligned}
&\mathbb{E}_{G\sim\mathbb{G}(n,d)}\left[\prod_{i=1}^{k} \gamma_{H_i}(G)^{\ell_i}\right] \\
&\quad = \left(\prod_{i=1}^{k} \sigma_{H_i}^{-\ell_i} \prod_{i=m'+1}^{k} \ell_i!!\right) \sum_{s_i \le \ell_i/2} \sum_{H_{i,j}}{}^{*} \mathbb{E} \prod_{i=1}^{m'} \left(\binom{\ell_i}{2s_i} (2s_i)!! \prod_{j=2s_i+1}^{\ell_i} \chi_{H_{i,j}} \right) + O(n^{-1/6}),
\end{aligned}$$

where $\sum_{s_i\le \ell_i/2}$ denotes a simultaneous choice of such nonnegative integers s_i for $i\in[m']$ and where $\sum^*$ denotes a sum over choices of $H_{i,j}$ such that they are all vertex-disjoint other than pairs $H_{i,2j-1}=H_{i,2j}$ for $1\le j\le s_i/2$ when $1\le i\le m'$ as well as for $1\le j\le \ell_i/2$ when $m'+1\le i\le k$. This equation basically means that we can validly pair up the necessary graphs and then replace the χ_e^2 terms by 1. Furthermore, it is not hard to see based on the considerations so far that we can remove the vertex-disjointness condition between different $H_{i,j}$ without changing the error rate, and thus we can write

$$\begin{aligned}\mathbb{E}_{G\sim\mathbb{G}(n,d)}\bigg[\prod_{i=1}^k\gamma_{H_i}(G)^{\ell_i}\bigg] &= \bigg(\prod_{i=1}^k\sigma_{H_i}^{-\ell_i}\prod_{i=m'+1}^k\ell_i!!\bigg(\binom{n}{v(H_i)}\frac{v(H_i)!}{\mathrm{aut}(H_i)}\bigg)^{\ell_i/2}\bigg)\\ &\qquad\times\sum_{s_i\le\ell_i/2}\bigg(\prod_{i=1}^{m'}\binom{\ell_i}{2s_i}(2s_i)!!\bigg(\binom{n}{v(H_i)}\frac{v(H_i)!}{\mathrm{aut}(H_i)}\bigg)^{s_i}(\mathbb{E}\gamma_{H_i})^{\ell_i-2s_i}\bigg)\\ &\qquad+O(n^{-1/6}),\\ &=\prod_{i=m'+1}^k\ell_i!!\sum_{s_i\le\ell_i/2}\prod_{i=1}^{m'}\binom{\ell_i}{2s_i}(2s_i)!!(\sigma_{H_i}^{-1}\mathbb{E}\gamma_{H_i}(G))^{\ell_i-2s_i}+O(n^{-1/6}),\end{aligned}$$

using the formula for σ_H in the second step. Finally, the third bullet of Proposition 2.7 shows that $\mathbb{E}\gamma_{H_i}(G)=(1+O(n^{-1/5}))2n^{-v(H_i)/2}\cdot\binom{n}{v(H_i)}(v(H_i)!/\mathrm{aut}(H_i))=(1+O(n^{-1/5}))E_H$ for $i\in[m']$. We also know that $E_H=(2/v(H))^{1/2}\sigma_H$, hence we find

$$\mathbb{E}_{G\sim\mathbb{G}(n,d)}\bigg[\prod_{i=1}^k\gamma_{H_i}(G)^{\ell_i}\bigg]=\prod_{i=m'+1}^k\ell_i!!\sum_{s_i\le\ell_i/2}\prod_{i=1}^{m'}\binom{\ell_i}{2s_i}(2s_i)!!(E_H/\sigma_H)^{\ell_i-2s_i}+O(n^{-1/6}),$$

which can be seen to match the moments of $\mathcal{N}(E_H/\sigma_H,1)^{\otimes m'}\otimes\mathcal{N}(0,1)^{\otimes(k-m')}$. Using Lemma 3.1 and shifting appropriately, this implies the desired

$$(\widetilde{\gamma}_{H_i}(G))_{1\le i\le k}\xrightarrow{d.}\mathcal{N}(0,1)^{\otimes k}.$$

Finally, we briefly note that the moment computations above where $\ell_i\in\{1,2\}$ and all $\ell_j=0$ for $j\ne i$ show that the means and variances are as claimed. □

4. Computations with graph factors

We now prove that any fixed degree polynomial in the indicator functions $x_e\in\{0,1\}$ which is symmetric under vertex permutation can be rewritten (so that it agrees on the set of d-regular graphs) as a function of connected graph factors of the form in Definition 1.4. The reduction specifically to connected graph factors appears essentially in the work of Janson [Jan95, p. 347].

Lemma 4.1. *Given a disconnected graph H (with no isolated vertices) with connected components $H_1,\ldots,H_k$, $\gamma_H(\mathbf{x})-\prod_{i=1}^k\gamma_{H_i}(G)$ can be expressed (as a function on graphs) as a sum of $\gamma_{H'}$ with $v(H')<v(H)$ (though H' may be itself disconnected). Furthermore, the coefficients of the sum are bounded by $O(1/(p(1-p))^{O_H(1)})$.*

This can clearly be inductively applied to show that the connected graph factors generate all graph factors using polynomial expressions. The crucial lemma for our work is that given a connected graph H with a vertex of degree 1, the graph factor $\gamma_H(\mathbf{x})$ can be simplified further (since our input graphs are regular).

https://doi.org/10.1112/S0010437X23007364 Published online by Cambridge University Press

LEMMA 4.2. *Given a graph H (with no isolated vertices) with a vertex of degree 1, then $\gamma_H(\mathbf{x})$ can be expressed, as a function on d-regular graphs, as a sum of $\gamma_{H'}(\mathbf{x})$ with $v(H') < v(H)$. Furthermore, the coefficients of the sum are bounded by $O(1/(p(1-p))^{O_H(1)})$.*

Proof. Let v be a vertex in H of degree 1 and (u,v) be the unique edge in H connected to v. Note that, considering this as a sum over possible choices of v, we have $\sum_{v \neq u} \chi_{(v,u)} = 0$ by d-regularity. Therefore, it follows that

$$\gamma_H(\mathbf{x}) = \sum_{\substack{E \subseteq K_n \\ E \simeq H}} \prod_{e \in E} \chi_e = \sum_{\substack{E \subseteq K_n \\ E \simeq H}} \chi_{(u,v)} \prod_{e \in E \setminus \{(u,v)\}} \chi_e$$
$$= \sum_{\substack{E \subseteq K_n \\ E \simeq H}} \Bigg(- \sum_{u \in V(E) \setminus \{v\}} \chi_{(u,v)} \Bigg) \prod_{e \in E \setminus \{(u,v)\}} \chi_e$$

and the desired result follows immediately using that $\chi_e^2 = 1 - (2p-1)\chi_e/\sqrt{p(1-p)}$. $\square$

Note that iterating Lemmas 4.1 and 4.2 shows we can write any graph factor on d-regular graphs as a function (in terms of d) of ones that are connected and with minimum degree at least 2. (In particular, any graph factor corresponding to a tree can be expressed in terms of graph factors with cycles as well as the constant $\gamma_\emptyset = 1$.) However, we will not explicitly need this fact, but rather its implication that variances of graph factors satisfy a reasonable uniform bound. Furthermore, having a degree 1 vertex (such as with trees) leads to a natural power-saving in this bound.

LEMMA 4.3. *Suppose that $n/\log n \leq \min(d, n-d)$. Given a graph H (with no isolated vertices) we have $\mathrm{Var}_{G \sim \mathbb{G}(n,d)}(\gamma_H(G)) = O(n^{v(H)})$. Furthermore, if H has a degree 1 vertex, we have $\mathrm{Var}_{G \sim \mathbb{G}(n,d)}(\gamma_H(G)) \leq n^{v(H)-2/3}$.*

Proof. We induct on $v(H)$. Note $v(H) \leq 2$ is trivial, as in fact $\gamma_H(G)$ is deterministic, so both parts of the lemma are satisfied. For H being a connected graph with minimum degree at least 2, the desired result follows immediately from the moments calculation given in the proof of Theorem 1.6. In the remaining cases, for G a d-regular graph we find that if H has connected components $H_1, \ldots, H_k$, then $\gamma_H(G) - \prod_{i=1}^k \gamma_{H_i}(G)$ can be written as a sum of graph factors each involving coefficients bounded by $1/(p(1-p))^{O_H(1)}$ and with at most $v(H) - 1$ vertices by Lemma 4.1. If there is a vertex of degree 1 in H, and hence some H_i, we can apply Lemma 4.2 and then we obtain a sum of graph factors with at most $v(H) - 1$ vertices after expanding (with similar bounds on coefficients). Thus, the total variance, by induction, is $(p(1-p))^{-O_H(1)} \cdot O(n^{v(H)-1}) \leq n^{v(H)-2/3}$, which satisfies the desired stronger bound in the case where H has a degree 1 vertex.

Finally, if all the H_i are minimum degree at least 2, then we see that the 'lower' portion corresponding to graph factors on at most $v(H) - 1$ vertices contributes $O(n^{v(H)-2/3})$ by induction similar to before. Hence, the problem reduces to understanding the variance of $\prod_{i=1}^k \gamma_{H_i}(G)$. We have

$$\mathrm{Var}\Bigg(\prod_{i=1}^k \gamma_{H_i}(G) \Bigg) \leq \mathbb{E} \prod_{i=1}^k \gamma_{H_i}(G)^2 \leq \prod_{i=1}^k (\mathbb{E}\gamma_{H_i}(G)^{2k})^{1/k}.$$

Again, the moment-based proof of Theorem 1.6 gives a bound of $O(n^{v(H_1)+\cdots+v(H_k)})$ for this. $\square$

https://doi.org/10.1112/S0010437X23007364 Published online by Cambridge University Press

5. Deduction of subgraph count and trace count normality

We now consider a subgraph count X_H for $G \sim \mathbb{G}(n,d)$ and prove the desired normality as in Theorem 1.2. This is essentially an immediate consequence of Theorem 1.6 and expanding into the appropriate graph factors. The precise nature of the contributing terms however depends in an intricate manner on the precise structure of H.

Proof of Theorem 1.2. Let H be a connected graph at least 2 vertices which is not a star. For $G \sim \mathbb{G}(n,d)$ write

$$W = X_H = \sum_{\substack{H' \subseteq K_n \\ H' \simeq H}} \prod_{e \in E(H')} x_e.$$

Letting $\chi_e = (x_e - p)/\sqrt{p(1-p)}$ as usual, we find that

$$\begin{aligned} W &= \sum_{\substack{H' \subseteq K_n \\ H' \simeq H}} \prod_{e \in E(H')} (p + \sqrt{p(1-p)}\chi_e) \\ &= \sum_{S \subseteq H} p^{e(H)-e(S)} (\sqrt{p(1-p)})^{e(S)} c_{S,H} d_{S,H} \binom{n - v(S)}{v(H) - v(S)} \gamma_S(\mathbf{x}), \end{aligned} \tag{5.1}$$

where $c_{S,H} = (v(H) - v(S))!\mathrm{aut}(S)/\mathrm{aut}(H)$, $d_{S,H} = N(H,S)$ (the number of times S appears as a subgraph of H), and the sum is over subgraphs S (lacking isolated vertices) of H up to isomorphism. For the empty graph, we have $c_{\emptyset,H} = v(H)!/\mathrm{aut} H$ and $d_{\emptyset,H} = 1$.

Note that the graph factors γ_S with $e(S) \le 2$ (the empty graph, an edge, a star with two edges, and two disjoint edges) are deterministic since G is a d-regular graph. If H contains a C_3 note that all other graph factors γ_S in the expansion have $v(S) \ge 4$, hence the corresponding terms have variance bounded by $O(n^{2(v(H)-v(S))} \cdot n^{v(S)}) = O(n^{2v(H)-4})$ by Lemma 4.3 while the γ_{C_3} term has variance

$$\left(\frac{6(v(H)-3)!}{\mathrm{aut}(H)} N(H, C_3) p^{e(H)-3/2} (1-p)^{3/2} \binom{n-3}{v(H)-3} \right)^2 \mathrm{Var}[\gamma_{C_3}].$$

Since the variance determination in Theorem 1.6 allows us to compute $\mathrm{Var}[\gamma_{C_3}] = (1 + O(n^{-1/6}))n^3/6$, we easily obtain the first bullet point of Theorem 1.2: we can write $W = X + Y$ where X is the term coming from γ_{C_3} and Y is the rest. We have that $\mathrm{Var}[Y] = O(n^{-1}\mathrm{Var}[X])$. Thus, since X satisfies a central limit theorem, so does $X + Y$. Furthermore, the variance can be written

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y)$$

and $|\mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y)| \le \sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]}$ by Cauchy–Schwarz, which gives an appropriate bound for the change in the variance going from X to $X + Y$. In particular, $\mathrm{Var}[X + Y] = (1 + O(n^{-1/2}))\mathrm{Var}[X]$.

Next suppose that H contains a C_4 but no C_3. Then all potential contributing graph factors which are not deterministic are on at least 4 vertices. Note that any graph factor γ_S with $v(S) \ge 5$ has corresponding variance at most $O(n^{2v(H)-5})$ by Lemma 4.3. Furthermore for $v(S) = 4$, note that if some vertex has degree 1, then by Lemma 4.3 we obtain corresponding variance of order $O(n^{2v(H)-4-2/3})$. All remaining S must have 4 vertices, minimum degree at least 2, and contain

https://doi.org/10.1112/S0010437X23007364 Published online by Cambridge University Press

no C_3, so $S = C_4$. The variance of the γ_{C_4} term is

$$\left(\frac{8(v(H)-4)!}{\mathrm{aut}(H)}N(H,C_4)p^{e(H)-2}(1-p)^2\binom{n-4}{v(H)-4}\right)^2\mathrm{Var}[\gamma_{C_4}].$$

The second bullet point of Theorem 1.2 follows similar to before.

The last case is when H contains neither C_3 nor C_4. Since H is not a star, H contains a P_4, i.e. a path on 4 vertices. First note that graph factors γ_S with $v(S) \geq 6$ have corresponding variance $O(n^{2v(H)-6})$ and graph factors γ_S with $v(S) = 5$ and some vertex of degree 1 have corresponding variance $O(n^{2v(H)-5-2/3})$ by Lemma 4.3. Furthermore, since H has no C_3 and no C_4, we see that the only possible S with $v(S) \leq 4$ for which γ_S is not deterministic is $S = P_4$. In addition, the possible S with $v(S) = 5$ are those with minimum degree at least 2 and no C_3 and no C_4, which is easily seen to force $S = C_5$. The variance of the γ_{C_5} term is

$$\begin{aligned}&\left(\frac{10(v(H)-5)!}{\mathrm{aut}(H)}N(H,C_5)p^{e(H)-5/2}(1-p)^{5/2}\binom{n-5}{v(H)-5}\right)^2\mathrm{Var}[\gamma_{C_5}]\\&\quad= (1+O(n^{-1/6}))\frac{10N(H,C_5)^2}{\mathrm{aut}(H)^2}p^{2e(H)-5}(1-p)^5n^{2v(H)-5}\end{aligned}$$

by Theorem 1.6. The graph factor γ_{P_4} is more delicate, as we must use the observation in Lemma 4.2 that we can reduce its complexity using $\sum_{v\neq u}\chi_{(v,u)} = 0$ for all fixed u. We obtain

$$\begin{aligned}\gamma_{P_4}(\mathbf{x}) &= \frac{1}{2}\sum_{u,v,w}\left(\chi_{(u,v)}\chi_{(v,w)}\sum_{u'\neq u,v,w}\chi_{(w,u')}\right) = \frac{1}{2}\sum_{u,v,w}\chi_{(u,v)}\chi_{(v,w)}(-\chi_{(w,u)}-\chi_{(w,v)})\\&= -\frac{1}{2}\sum_{u,v,w}(\chi_{(u,v)}\chi_{(v,w)}\chi_{(w,u)}+\chi_{(u,v)}\chi^2_{(v,w)})\\&= -3\gamma_{C_3} - \frac{1}{2}\sum_{u,v,w}\chi_{(u,v)}\left(1-\frac{(2p-1)\chi_{(v,w)}}{\sqrt{p(1-p)}}\right),\end{aligned}$$

where the sum is over tuples of distinct $u, v, w \in [n]$. Here we have used that $\chi_e^2 = 1-(2p-1)\chi_e/\sqrt{p(1-p)}$ and given that $\gamma_{K_2}, \gamma_{P_3}$ are deterministic, we find that $\gamma_{P_4} + 3\gamma_{C_3}$ is deterministic. This can alternatively be deduced by noting that $X_{P_4} + 3X_{C_3}$ is a deterministic function in a d-regular graph.[1] Therefore, the variance of the γ_{P_4} term is

$$\begin{aligned}&\left(\frac{2(v(H)-4)!}{\mathrm{aut}(H)}N(H,P_4)p^{e(H)-3/2}(1-p)^{3/2}\binom{n-4}{v(H)-4}\right)^2\cdot 9\mathrm{Var}[\gamma_{C_3}]\\&\quad= (1+O(n^{-1/6}))\frac{6N(H,P_4)^2}{\mathrm{aut}(H)^2}p^{2e(H)-3}(1-p)^3n^{2v(H)-5}.\end{aligned}$$

Finally, writing X_1 for the γ_{C_5} term and X_2 for the γ_{P_4} term, using the moment computations in the proof of Theorem 1.6 (applied to C_3 and C_5) we easily find that $\mathrm{Cov}(X_1, X_2) = O(n^{2v(H)-5-1/6})$. (Or, we can directly see this from the joint central limit theorem satisfied by $\gamma_{C_3}, \gamma_{C_5}$ in Theorem 1.6.) The third bullet of Theorem 1.2 follows similar to before. ∎

[1] We thank the referee for this remark, which provides a check for the formulas in Theorem 1.2 since we must have $\mathrm{Var}[X_{P_4}] = 9\mathrm{Var}[X_{C_3}]$.

https://doi.org/10.1112/S0010437X23007364 Published online by Cambridge University Press

Finally, we prove Corollary 1.8. Again, this is mostly rearranging terms in order to apply Theorem 1.6. Our analysis is more complicated than typical trace expansion arguments as one cannot trivially rule out walks where an edge appears with multiplicity 1 in various expectation computations. To perform the necessary analysis, we will need the following modified version of Lemma 3.2 which allows for some of the H_i to be a doubled edge (but we may otherwise restrict to cycles). As a consequence, the equality case is more complicated. Recall that for a multigraph G, $E_{\mathrm{sing}}(G)$ is the set of edges of multiplicity 1.

Lemma 5.1. *Let $\mathcal{H} = (H_i)_{1\le i\le k}$ be a sequence of cycles or multigraphs consisting of a doubled edge. Consider overlaying the H_i in order to obtain a multigraph G. Let $E_{\mathrm{sing}} = E_{\mathrm{sing}}(G)$. Suppose that every connected component of G contains at least one participating cycle of $\mathcal{H}$. Then we have*

$$v(G) - \frac{1}{2}|E_{\mathrm{sing}}| \le \frac{1}{2}\sum_{i=1}^{k} v(H_i) = \frac{e(G)}{2}$$

with equality only if every connected component of G can be obtained by first taking a cycle of $\mathcal{H}$ or perfectly overlaying two cycles of $\mathcal{H}$, and second attaching pendant trees of doubled edges from $\mathcal{H}$ (in particular, removing the cycle portion leaves a forest of doubled edges). Here $e(G)$ is computed with multiplicity.

Remark 5.2. We note that for any H_i which is a doubled edge, the corresponding multiedges in G are not contained in E_{sing}.

Proof. Without loss of generality we may assume G is connected, as this clearly preserves the inequality as well as equality cases. In addition, the equality $\sum_{i=1}^{k} v(H_i) = e(G)$ is trivial since cycles and doubled edges have the same edge and vertex counts. Now let $H_1, \ldots, H_{k'}$ be the cycles and the rest the doubled edges. Let G' be the multigraph overlay of $H_1, \ldots, H_{k'}$ and define $E'_{\mathrm{sing}} = E_{\mathrm{sing}}(G')$. These are the edges contained in a single H_i for $i \in [k']$. By Lemma 3.2, we have

$$v(G') - \frac{1}{2}|E'_{\mathrm{sing}}| \le \frac{1}{2}\sum_{i=1}^{k'} v(H_i)$$

and equality can only occur if every connected component of G' is either a single cycle H_i for $i \in [k']$ or an overlay of 2 equal cycles $H_i, H_{i'}$ for distinct $i, i' \in [k']$. Furthermore, by initial assumption $k' \ge 1$ so G' is nonempty.

Now consider adding in the doubled edges in a specified order, starting at $G_{k'} = G'$ and ending at $G_k = G$. We choose the order as follows: at time $k' \le i \le k-1$, once we have G_i, since we know G is connected there must be a doubled edge to add which shares a vertex with G_i; add one of those edges. Define $E^{(i)}_{\mathrm{sing}}$ in the obvious way. We see that

$$(v(G_{i+1}) - v(G_i)) - \tfrac{1}{2}(|E^{(i+1)}_{\mathrm{sing}}| - |E^{(i)}_{\mathrm{sing}}|) \le 1 = \tfrac{1}{2}v(H_{i+1})$$

since either we add 0 vertices and at worst reduce the number of singleton edges by 1, or we add 1 vertex and, thus, leave the number of singleton edges unchanged (here we are using that G' is nonempty and the connected components of G_i each contain a cycle, otherwise it could be possible to add 2 vertices at the beginning). Equality occurs here only if we add 1 new pendant vertex.

Adding these inequalities over all i, we obtain the desired inequality. Furthermore, equality can only occur if the connected components of G' are single or doubled cycles, and then we only add pendant trees of doubled edges. But since the final multigraph G is connected, this

https://doi.org/10.1112/S0010437X23007364 Published online by Cambridge University Press

means we must have started with at most one component as we cannot bridge between connected components of some G_i using a doubled edge while simultaneously increasing the vertex count by 1. The result follows. $\square$

Finally, we demonstrate Corollary 1.8.

Proof of Corollary 1.8. Note that deterministically we have that the all 1 vector is an eigenvector with eigenvalue d. Therefore, we have that $M := A_G - pJ + pI$ (where J is the all 1 matrix) has eigenvalues $\lambda_i + p$ for $2 \le i \le n$ and one eigenvalue of 0.

In order to prove Corollary 1.8 it suffices to prove that if $E_\ell^* = \mathbb{E}\operatorname{tr}(M^\ell)$, $\sigma_\ell^{*2} = \operatorname{Var}[\operatorname{tr}(M^\ell)]$, then

$$(\sigma_\ell^{*-1/2}(\operatorname{tr}(M^\ell) - E_\ell^*))_{3\le \ell\le k} \xrightarrow{d.} \mathcal{N}(0, \Sigma_k)$$

and $\sigma_\ell^* = \Theta((p(1-p)n)^{\ell/2})$ for fixed $\ell \ge 3$. To see that this implies the desired result note that each term of $\sum_{i=2}^n (\lambda_i + p)^\ell - \sum_{i=2}^n \lambda_i^\ell$ can be represented as a degree at most $\ell - 1$ polynomial in $\lambda_i + p$ with coefficients bounded by $O_k(p^{O_k(1)})$. These terms are lower order due to the order of the variance (and using that the first two moments of the eigenvalues are deterministic).

Given $\ell \ge 3$, note that

$$\operatorname{tr}((A_G - pJ + pI)^\ell/(p(1-p))^{\ell/2}) = \sum_{u_1,\ldots,u_\ell\in[n]} \prod_{i=1}^{\ell} \chi_{(u_i,u_{i+1})}$$

where we define $\chi_{(u,u)} = 0$ and take indices modulo ℓ. The sum is over closed walks of length ℓ.

Consider the closed walk $u_1, \ldots, u_\ell$. As $\chi_{(u,u)} = 0$, we have that the walk has no self-loops corresponding to $u_{t+1} = u_t$. The edges traced out thus form a multigraph when superimposed. Let $\mathcal{W}_\ell$ be the collection of possible isomorphism types of multigraphs and for $(u, v) \in G$ and $G \in \mathcal{W}_\ell$ let $G(u, v)$ be the multiplicity of (u, v) in G. We see

$$\operatorname{tr}((A_G - pJ + pI)^\ell/(p(1-p))^{\ell/2}) = \sum_{G\in\mathcal{W}_\ell} c_G \Bigg(\sum_{\substack{V(G')\subseteq V(K_n)\\ G'\simeq G}} \prod_{(u,v)\in G} \chi_{(u,v)}^{G(u,v)} \Bigg),$$

where c_G is the number of choices of vertices in G and closed walks of length ℓ starting at that vertex and traversing each edge $(u, v) \in G$ in either direction exactly $G(u, v)$ times. If G is a simple graph, the term on the inside is just $\gamma_G(\mathbf{x})$. We therefore abusively define

$$\gamma_G(\mathbf{x}) = \sum_{\substack{V(G')\subseteq V(K_n)\\ G'\simeq G}} \prod_{(u,v)\in G} \chi_{(u,v)}^{G(u,v)}$$

for multigraphs G without isolated vertices. However, we will later use $\chi_e^2 = 1 - (2p - 1)\chi_e/\sqrt{p(1-p)}$ and similar relations for higher powers to reduce to a linear combination of graph factors γ_F.

Furthermore, every multigraph in $\mathcal{W}_\ell$ can be decomposed (with multiplicity preserved) into a collection of cycles and doubled edges: move along the walk until the first vertex repetition, then remove a portion corresponding to a doubled edge or cycle, and keep doing this. We can further further turn the doubled edges into a multitree by iteratively removing cycles of doubled edges and turning them into two cycles. Given $G \in \mathcal{W}_\ell$, let $\mathcal{H}_G$ be the sequence of multigraphs thus generated. Let $\mathcal{T}_\ell$ be the collection of $G \in \mathcal{W}_\ell$ that are composed only of doubled edges, which therefore compose a tree as G is connected.

First consider $G \in \mathcal{W}_\ell \setminus \mathcal{T}_\ell$, so that $\mathcal{H}_G$ contains at least one cycle. Let $\mathcal{W}'_G$ be all possible isomorphism classes $G_1 \cup G_2$ for the multigraph union of two copies $G_1, G_2 \simeq G$. We see

$$\operatorname{Var}[\gamma_G] \le \mathbb{E}\gamma_G^2 \lesssim \sum_{G' \in \mathcal{W}'_G} n^{v(G')} \cdot n^{-|E_{\text{sing}}(G')|/2+1/3}$$

by expansion and Proposition 2.7. Now consider the collection of cycles and doubled edges which make up G'. Since they are overlaid in a way that form two copies of G, our condition on $\mathcal{H}_G$ implies that every connected component of G' has at least one cycle participating in its creation. Thus, Lemma 5.1 applies. For cases where equality does not hold, we have $v(G') - |E_{\text{sing}}(G')|/2 < \ell$ since $e(G') = 2\ell$. This implies $v(G') - |E_{\text{sing}}(G')|/2 + 1/3 \le \ell - 1/6$. For cases where equality does hold, by Lemma 5.1 every connected component of G' must consist of a cycle or doubled cycle (which come from our specified list of cycles that create G_1, G_2) and then pendant trees of doubled edges. Note that G_1, G_2 are each connected, so we either have that these are disjoint and of this form or they are connected and form such a graph. In the former case G is clearly either a cycle or doubled cycle with pendant trees of doubled edges. In the latter case we easily deduce that G is a single cycle with pendant trees of doubled edges (recalling $G \notin \mathcal{T}_\ell$). Let $\mathcal{C}_\ell$ be the set of isomorphism classes of these more special forms, so that $\operatorname{Var}[\gamma_G] = O(n^{\ell-1/6})$ for $G \in \mathcal{W}_\ell \setminus (\mathcal{T}_\ell \cup \mathcal{C}_\ell)$.

Next we study $G \in \mathcal{T}_\ell$, in which case ℓ must be even (thus, $\ell \ge 4$) and G has $\ell/2$ doubled edges and at most $\ell/2+1$ vertices (being a multitree). We write

$$\gamma_G(\mathbf{x}) = \sum_{F \subseteq G} c_{F,G}(p) \Bigg(\sum_{\substack{F' \subseteq K_n \\ F' \simeq F}} \prod_{(u,v) \in F'} \chi_{(u,v)} \Bigg),$$

where the sum is over graphs F up to isomorphism obtained by either including 1 or 0 edges for each edge in G (without multiplicity). Here $c_{F,G}(p)$ are appropriately computed constants. This is shown by expanding via $\chi_e^2 = 1 - (2p-1)\chi_e/\sqrt{p(1-p)}$ and similar for higher powers, and collecting the patterns that can result. Note also that F may have isolated vertices, and that it is a forest since G is a multitree. Regardless of these isolated vertices, let us abusively denote the inside term as $\gamma_F(\mathbf{x})$ (this agrees with the usual definition). If $e(F) \le 2$, then the term corresponding to F is deterministic since we are considering d-regular graphs. Hence, we may restrict to just terms with $e(F) \ge 3$. We have for such F that

$$\operatorname{Var}[\gamma_F] \le \mathbb{E}\gamma_F^2 \lesssim \max_{v \ge 0} n^{2(\ell/2+1)-v} \cdot n^{-(2e(F)-\max(v-1,0))/2+1/3} = \max_{v \ge 0} n^{\ell+2-e(F)+\max(-v-1,-2v)/2+1/3}$$

by Proposition 2.7: two copies of F with v overlapping vertices can share at most $\max(v-1,0)$ edges. This clearly yields $\operatorname{Var}[\gamma_F] = O(n^{\ell-2/3})$ and, thus, we find $\operatorname{Var}[\gamma_G] = O(n^{\ell-1/6})$ for $G \in \mathcal{T}_\ell$.

Finally, consider $G \in \mathcal{C}_\ell$. In $\gamma_G(\mathbf{x})$, the highest degree of any term χ_e is 2. Using $\chi_e^2 = 1 - (2p-1)\chi_e/\sqrt{p(1-p)}$ and expanding out, it is easy to see similar to above that the sum of the terms involving any $-(2p-1)\chi_e/\sqrt{p(1-p)}$ in the expansion, call this γ'_G, has total variance bounded by $O(n^{\ell-1/6})$. Finally, if G is a doubled cycle with pendant trees of double edges, then the remaining term is deterministic, while if G is a single cycle with such pendant trees then the remaining term is a cycle of say length r with $(\ell-r)/2$ isolated vertices. Finally, recall that the variance of γ_{C_r} is $O(n^r)$ by Theorem 1.6.

Overall, combining all this information and noting that a γ_{C_ℓ} term only comes from a walk that repeats no vertices, we see

$$\operatorname{tr}((A_G - pJ + pI)^\ell/(p(1-p))^{\ell/2}) = 2\ell\gamma_{C_\ell}(\mathbf{x}) + \sum_{\substack{3 \le r < \ell \\ r \equiv \ell \pmod 2}} \alpha_{\ell,r} n^{(\ell-r)/2} \gamma_{C_r}(\mathbf{x}) + X_\ell$$

for some random variable X_ℓ satisfying $\mathrm{Var}[X_\ell] = O(n^{\ell-1/6})$ and for appropriate combinatorially definable rational numbers $\alpha_{\ell,r}$ independent of n, p.

Equivalently,

$$\operatorname{tr}((A_G - pJ + pI)^\ell/(p(1-p)n)^{\ell/2}) = 2\ell(n^{-\ell/2}\gamma_{C_\ell}) + \sum_{\substack{3\le r<\ell \\ r\equiv \ell \pmod 2}} -\alpha_{\ell,r,n}(n^{-r/2}\gamma_{C_r}) + n^{-\ell/2}X_\ell$$

and now the result clearly follows from Theorem 1.6 as the error terms X_ℓ are negligible (using a similar argument as in the proof of Theorem 1.2). Since this representation in terms of the cycle graph factors is triangular, we furthermore see that the resulting Σ_k that arises in the limit is indeed positive definite; here we are using that the coefficient of γ_{C_ℓ} is a strictly positive constant, that the $\alpha_{\ell,r}$ are independent of p, n and of bounded size in terms of ℓ, and that the graph factors corresponding to cycles are jointly independently normally distributed by Theorem 1.6. □

Acknowledgements

We thank Vishesh Jain for several useful discussions which played a key role in the development of the project. We also thank Brendan McKay and Nick Wormald for helpful comments. Finally, we are very grateful to the anonymous referee for a number of useful comments, including finding errors in the initial version of the manuscript with respect to the statements of the main results, and for providing us with numerical data which helped to corroborate various claims.

Conflicts of Interest

None.

References

Bol80 B. Bollobás, *A probabilistic proof of an asymptotic formula for the number of labelled regular graphs*, European J. Combin. **1** (1980), 311–316.

Dur19 R. Durrett, *Probability—theory and examples*, fifth edition, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 49 (Cambridge University Press, Cambridge, 2019).

Gao20 P. Gao, *Triangles and subgraph probabilities in random regular graphs*, Preprint (2020), arXiv:2012.01492.

GW08 Z. Gao and N. C. Wormald, *Distribution of subgraphs of random regular graphs*, Random Structures Algorithms **32** (2008), 38–48.

He22 Y. He, *Spectral gap and edge universality of dense random regular graphs*, Preprint (2022), arXiv:2203.07317.

IM18 M. Isaev and B. D. McKay, *Complex martingales and asymptotic enumeration*, Random Structures Algorithms **52** (2018), 617–661.

Jan94a S. Janson, *The numbers of spanning trees, Hamilton cycles and perfect matchings in a random graph*, Combin. Probab. Comput. **3** (1994), 97–126.

Jan94b S. Janson, *Orthogonal decompositions and functional limit theorems for random graph statistics*, Mem. Amer. Math. Soc. **111** (1994).

Jan95 S. Janson, *A graph Fourier transform and proportional graphs*, Random Structures Algorithms **6** (1995), 341–351.

JŁR00 S. Janson, T. Łuczak and A. Rucinski, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization (Wiley, New York, 2000).

KSV07 J. H. Kim, B. Sudakov and V. Vu, *Small subgraphs of random regular graphs*, Discrete Math. **307** (2007), 1961–1967.

LS20 B. Landon and P. Sosoe, *Applications of mesoscopic CLTs in random matrix theory*, Ann. Appl. Probab. **30** (2020), 2769–2795.

LW17 A. Liebenau and N. Wormald, *Asymptotic enumeration of graphs by degree sequence, and the degree sequence of a random graph*, J. Eur. Math. Soc. (JEMS), to appear.

McK85 B. D. McKay, *Asymptotics for symmetric 0-1 matrices with prescribed row sums*, Ars Combin. **19** (1985), 15–25.

McK10 B. D. McKay, *Subgraphs of random graphs with specified degrees*, Proceedings of the International Congress of Mathematicians, vol. IV (Hindustan Book Agency, New Delhi, 2010), 2489–2501.

McK11 B. D. McKay, *Subgraphs of dense random graphs with specified degrees*, Combin. Probab. Comput. **20** (2011), 413–433.

MW90 B. D. McKay and N. C. Wormald, *Asymptotic enumeration by degree sequence of graphs of high degree*, European J. Combin. **11** (1990), 565–580.

MW91 B. D. McKay and N. C. Wormald, *Asymptotic enumeration by degree sequence of graphs with degrees* $o(n^{1/2})$, Combinatorica **11** (1991), 369–382.

MWW04 B. D. McKay, N. C. Wormald and B. Wysocka, *Short cycles in random regular graphs*, Electron. J. Combin. **11** (2004), 12.

RW19 P. Rigollet and J. Weed, *Uncoupled isotonic regression via minimum Wasserstein deconvolution*, Inf. Inference **8** (2019), 691–717.

Ruc88 A. Ruciński, *When are small subgraphs of a random graph normally distributed?*, Probab. Theory Related Fields **78** (1988), 1–10.

SS98 Y. Sinai and A. Soshnikov, *Central limit theorem for traces of large random symmetric matrices with independent matrix elements*, Bull. Braz. Math. Soc. (N.S.) **29** (1998), 1–24.

Wor18 N. Wormald, *Asymptotic enumeration of graphs with given degree sequence*, Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. IV. Invited Lectures (World Scientific, Hackensack, NJ, 2018), 3245–3264.

Wor81 N. C. Wormald, *The asymptotic distribution of short cycles in random regular graphs*, J. Combin. Theory Ser. B **31** (1981), 168–182.

Ashwin Sah asah@mit.edu

Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

Mehtaab Sawhney msawhney@mit.edu

Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

Compositio Mathematica is owned by the Foundation Compositio Mathematica and published by the London Mathematical Society in partnership with Cambridge University Press. All surplus income from the publication of *Compositio Mathematica* is returned to mathematics and higher education through the charitable activities of the Foundation, the London Mathematical Society and Cambridge University Press.