Combinatorics, Probability and Computing (2000) 9, 125–148. Printed in the United Kingdom

Spanning Subgraphs of Random Graphs

OLIVER RIORDAN

SFB 343, Universität Bielefeld,
Postfach 10 01 31, 33501 Bielefeld, Germany

and

Department of Pure Mathematics and Mathematical Statistics,
16 Mill Lane, Cambridge CB2 1SB, England
(e-mail: omr10@dpmms.cam.ac.uk)

Received 1 May 1998; revised 30 August 1999

Let G_p be a random graph on 2^d vertices where edges are selected independently with a fixed probability $p > \frac{1}{4}$, and let H be the d-dimensional hypercube Q^d. We answer a question of Bollobás by showing that, as $d \to \infty$, G_p almost surely has a spanning subgraph isomorphic to H. In fact we prove a stronger result which implies that the number of d-cubes in $G \in \mathcal{G}(n, M)$ is asymptotically normally distributed for M in a certain range. The result proved can be applied to many other graphs, also improving previous results for the lattice, that is, the 2-dimensional square grid. The proof uses the second moment method – writing X for the number of subgraphs of G isomorphic to H, where G is a suitable random graph, we expand the variance of X as a sum over all subgraphs of H itself. As the subgraphs of H may be quite complicated, most of the work is in estimating the various terms of this sum.

1. Introduction

In this paper we consider the following general question: when is a random graph G likely to have a spanning subgraph isomorphic to a given graph H? For the graph H we concentrate on two special cases, that of the *cube* and that of the *lattice*, although we prove a more general result which covers both these cases as well as many others. For the random graph G we consider the two standard models $\mathcal{G}(n, p)$ and $\mathcal{G}(n, M)$. As usual, a random $G \in \mathcal{G}(n, p)$ is a graph with vertex set $[n] = \{1, 2, \ldots, n\}$ obtained by selecting each possible edge independently with probability p. Writing $N = \binom{n}{2}$, a random $G \in \mathcal{G}(n, M)$ is obtained by selecting each of the $\binom{N}{M}$ graphs on $[n]$ with M edges with equal probability.

Let Q^d, the *d-cube*, be the graph with vertex set $\{0, 1\}^d$ where two sequences are adjacent if they differ in precisely one place. The following question due to Bollobás (see

[1]) provides the motivation for this paper: for what range of p does $G \in \mathscr{G}(2^d, p)$ almost surely contain a copy of Q^d?

Alon and Füredi [1] gave a partial answer to this question, showing that any fixed $p > \frac{1}{2}$ suffices. Writing $n = 2^d$, the expected number of copies of Q^d in $G \in \mathscr{G}(n,p)$ is $n!p^{dn/2}/(nd!)$, which is much smaller than $(np^{d/2})^n = (4p)^{dn/2}$. Thus, as noted in [1], for $p \leqslant \frac{1}{4}$ the expected number of d-cubes in G tends to zero, so $p > \frac{1}{4}$ is a necessary condition for almost every $G \in G(2^d, p)$ to contain a d-cube. We shall prove the following result, which is thus close to best possible.

Theorem 1.1. *Let $p = p(d) = \frac{1}{4} + 6\log d/d$. Then with probability tending to 1 as $d \to \infty$ a random $G \in \mathscr{G}(2^d, p)$ contains a copy of Q^d.*

Another graph for which such a question has been asked is the *lattice* L_k, the graph on $[k] \times [k]$ in which two points are adjacent if the Euclidean distance between them is 1. The question of which random graphs contain spanning lattices was raised by Venkatesan and Levin [8]. Writing $n = k^2$, Alon and Füredi [1] showed that $p(n) = O((\log n/n)^{1/4})$ will do, noting that $p = n^{-1/2}$ is a lower bound. A different partial answer to this question was given by Fernandez de la Vega and Manoussakis [5], who gave conditions under which $G \in \mathscr{G}(n,p)$ almost surely contains a lattice covering some constant fraction of the vertices. We shall considerably strengthen both these results, again showing that the lower bound given by the expectation is essentially correct.

Theorem 1.2. *Let $p = p(n) = \omega(n)n^{-1/2}$, where $\omega(n) \to \infty$, and let $n = k^2$. Then with probability tending to 1 as $k \to \infty$ a random $G \in \mathscr{G}(n,p)$ contains a copy of L_k.*

Note that, together with the lower bound given by the expectation, Theorem 1.2 states that $p = n^{-1/2}$ is a threshold for the property of containing a spanning lattice, in the sense of Erdős and Rényi (see [2]).

We shall deduce Theorems 1.1 and 1.2 from a common generalization which could also be applied to many other graphs. As the statement of this result is rather cumbersome, we postpone it to the next section. We note here, however, that we shall be forced to work in $\mathscr{G}(n,M)$ rather than $\mathscr{G}(n,p)$. The results we obtain actually give the asymptotic distribution of the number of cubes or grids in $G \in \mathscr{G}(n,M)$, for M in a certain range.

For a random variable X, let us write X^* for its *standardization*, that is, for $(X - \mathbb{E}(X))/\sqrt{\mathrm{Var}(X)}$. We shall prove the following results, writing N for $\binom{n}{2}$ as before.

Theorem 1.3. *Suppose that $M/N \geqslant \frac{1}{4} + 5\log d/d$, where $n = 2^d$, and let $X = X(G)$ be the number of copies of Q^d in $G \in \mathscr{G}(n,M)$. Then $\mathbb{P}(X = 0) \to 0$ as $d \to \infty$.*

If in addition $(1 - M/N)\log n \to \infty$, then X^* converges in distribution to a standard normal distribution as $d \to \infty$.

Theorem 1.4. *Suppose that $(M/N)n^{1/2} \to \infty$, where $n = k^2$, and let $X = X(G)$ be the number of copies of L_k in $G \in \mathscr{G}(n,M)$. Then $\mathbb{P}(X = 0) \to 0$ as $k \to \infty$.*

If in addition $(M/N)\log n \to 0$, then X^* converges in distribution to a standard normal distribution as $k \to \infty$.

Note that the first parts of the results above immediately imply Theorems 1.1 and 1.2, as a random $G \in \mathcal{G}(n,p)$ almost certainly has at least M edges, whenever $(pN - M)/\sqrt{Np(1-p)} \to \infty$. Also, for each of Theorems 1.3 and 1.4 the second part implies the first part. However, we shall deduce the first parts separately from a simpler result than that needed to prove the distribution results.

Before getting down to work, we give a very brief outline of the method used. Previous approaches ([1], [5]) to the above questions involved finding a copy of $H = Q^d$ or L_k in G one piece at a time. In contrast, we look at the random variable X, the number of full copies of H in G. We bound the variance of X in terms of its expectation, and use this to show that $\mathbb{P}(X = 0) \to 0$. More precisely, we express the variance of X as the expectation of a certain function of $e(H_1 \cap H_2)$, where H_1 and H_2 are two copies of H on $[n]$ chosen independently at random. This is similar to the method used by Janson [6] to find the asymptotic distributions of the numbers of Hamilton cycles, spanning trees and perfect matchings in certain random graphs. A simple example of the expansion used here occurs in [7], where it is used to find the variance of the number of Hamilton cycles in $G \in \mathcal{G}(n, \frac{1}{2})$.

2. Statement of the main results

In what follows we consider the limit of the probability that $G \in \mathcal{G}(n, pN)$ contains a 'fixed' spanning subgraph H. Of course, this makes no sense – we really consider a fixed sequence $H^{(i)}$ of graphs, such as Q^i, with $n = |H^{(i)}| \to \infty$. All the parameters we consider will depend on i, but we shall suppress this dependence throughout. All limits will be taken as $i \to \infty$, or, equivalently, as $n \to \infty$. As usual, we write $f = o(g)$ if $f/g \to 0$, and $f \sim g$ if $f/g \to 1$. We also write $f \lesssim g$ if $f \leqslant (1+o(1))g$, that is, if $\limsup f/g \leqslant 1$, and $f = \Omega(g)$ if f/g is bounded away from zero for n sufficiently large.

To state our results we need the following parameters of H: $n = |H|$, $e(H)$, which we shall often write as αN, $\Delta = \Delta(H)$, the maximum degree of H, and

$$e_H(v) = \max\{e(F) : F \subset H, |F| = v\}.$$

We also consider the ratio

$$\gamma = \gamma(H) = \max_{3 \leqslant v \leqslant n}\{e_H(v)/(v-2)\}.$$

Note that, in the definition of γ, if we divided by v instead of by $v-2$ we would just have twice the maximum average degree of a subgraph of H. If H is balanced, then this modified maximum would be achieved by H. If in addition the average degree of H is large, the same is likely to be true of γ. However, as we shall see, for the lattice, $e_H(v)/(v-2)$ attains its maximum when $v = 3$ or $v = 4$.

With the above notation we are now ready to state our main result.

Theorem 2.1. *Let $H(= H^{(i)})$ be a fixed sequence of graphs with $n = |H| \to \infty$, and let $\gamma = \gamma(H)$ be as above. Let $e(H) = \alpha N = \alpha(n)N$, and let $p = p(n) \in (0,1)$ with pN an* *integer. Suppose that the following conditions hold:*

$$\alpha N \geqslant n, \text{ and } pN, \ (1-p)\sqrt{n} \to \infty, \tag{2.1}$$

and

$$np^{\gamma}/\Delta^4 \to \infty. \tag{2.2}$$

Then, with probability tending to 1 as $n \to \infty$, a random $G \in \mathcal{G}(n, pN)$ has a spanning subgraph isomorphic to H.

Remark. As we shall see from the proof, the condition we actually need in place of (2.2) is slightly weaker. When Δ is fairly small, the condition actually needed is very similar to (2.2). Since Theorem 2.1 is potentially applicable with Δ nearly as large as $n^{1/4}$, however, we note that we can instead use the conditions

$$\Delta = o(n^{1/4}), \ p \sim 1, \ \Delta \to \infty \text{ and } np^{\gamma}/\Delta^{2+\epsilon} \to \infty, \tag{2.2$'$}$$

where $\epsilon > 0$ is arbitrarily small.

Remark. The conditions in (2.1) were chosen because they are convenient in proving Theorem 2.1, and some of them can probably be weakened or dropped. However, they all hold in the range where Theorem 2.1 is likely to give good bounds. Also, we *shall* make essential use of two consequences of the above conditions, namely

$$np^2/\Delta^4 \to \infty \tag{2.3}$$

and

$$\alpha^3 Np^{-2} \to 0. \tag{2.4}$$

Clearly, if (2.2′) holds, then so does (2.3). Also, $\alpha N \geqslant n$ implies that $\Delta \geqslant 2$, and hence that $\gamma \geqslant 2$. Together with (2.2), this implies (2.3). Since $\alpha \lesssim \Delta/n$, condition (2.3) implies (2.4). Thus (2.3) and (2.4) hold under the conditions of the theorem, whether we take (2.2) or (2.2′).

Under slightly stronger (and rather messier) conditions we can strengthen the conclusion of Theorem 2.1, showing that the number of copies of H present in $G \in \mathcal{G}(n, pN)$ is asymptotically normally distributed. To state the next result we write $\alpha_2 = \alpha_2(H)$ for the proportion of all P_2s (*i.e.*, pairs of adjacent edges) on $[n]$ contained in a particular copy of H, and $X_H(G)$ for the number of subgraphs of G isomorphic to H.

Theorem 2.2. *Suppose that the conditions of Theorem 2.1 are satisfied, and that H is triangle-free. Let $\alpha_2 = \alpha_2(n)$ be as above, and let $j_0 = \frac{1-p}{p}\alpha^2 N$. Suppose in addition that*

$$j_0/\log n \to \infty, \tag{2.5}$$

$$\Delta^6 n^{-1} p^{-2} = o((1-p)^2), \tag{2.6}$$

$$\alpha_2 - \alpha^2 = \Omega(\Delta n^{-2}), \quad \alpha_2 = O(\alpha^2), \tag{2.7}$$

and that either (2.2′) holds or

$$np^{\gamma'}/\Delta^2 \to \infty, \tag{2.8}$$

https://doi.org/10.1017/S0963548399004150 Published online by Cambridge University Press

where $\gamma' = \max_{v>4} \frac{e_H(v)-4}{v-4}$. *Then, writing* X *for* $X_H(G)$, $G \in \mathcal{G}(n, pN)$, *we have*

$$\frac{\mathrm{Var}(X)}{\mathbb{E}(X)^2} \sim (\alpha_2 - \alpha^2)^2 \left(\frac{1-p}{p}\right)^2 \frac{n^3}{2}, \tag{2.9}$$

and $X^* = (X - \mathbb{E}(X))/\sqrt{\mathrm{Var}(X)}$ *converges in distribution to a standard normal distribution.*

Remark. The conditions given above represent one possible way of simplifying (and strengthening) the conditions actually used in the proof. Thus, for example, one can check that the lower bound on $\alpha_2 - \alpha^2$ taken in (2.7) can be weakened provided condition (2.6) is strengthened accordingly, and *vice versa*. However, some lower bound on $\alpha_2 - \alpha^2$ is definitely needed: if this quantity is zero, then (2.9) cannot possibly hold. The bounds in (2.7) are natural: if H is Δ-regular then $\alpha = \Delta/(n-1)$ while $\alpha_2 = \binom{\Delta}{2}/\binom{n-1}{2}$, so $\alpha_2 - \alpha^2 \sim -\Delta n^{-2}$ and (2.7) holds.

We shall prove Theorem 2.1 in the next two sections and Theorem 2.2 in Section 5. For the moment, we assume the results, and deduce Theorems 1.3 and 1.4, and hence Theorems 1.1 and 1.2.

Proof of Theorem 1.3. Let $H = Q^d$ and $n = 2^d$: as above we write $N = \binom{n}{2}$ and $e(H) = \alpha N$, so $\alpha = d/(n-1)$. To prove the first part of Theorem 1.3 it suffices to verify (2.1) and (2.2) for $p = \lceil(\frac{1}{4} + 5\log d/d)N\rceil/N$. It is easy to see that (2.1) holds. For (2.2) we use the fact from [3] that $e_H(v) \leqslant v \log_4 v$, with equality when $v = n$, or any other power of two. It is easy to see (and one can check by differentiating) that for d large this implies that $\gamma = \gamma(H) = e(H)/(n-2) = dn/(2(n-2))$. As $1 + x \geqslant e^{x/2}$ for $x > 0$ sufficiently small, we see that, when d is large enough,

$$p^{-\gamma} \leqslant \left[\frac{1}{4}\left(1 + \frac{20\log d}{d}\right)\right]^{-\frac{d}{2}\frac{n}{n-2}} \leqslant (2^d e^{-5\log d})^{\frac{n}{n-2}} = (nd^{-5})^{\frac{n}{n-2}}.$$

As $n^{2/(n-2)} = o(d)$ we thus have $p^{-\gamma} = o(nd^{-4})$. As $\Delta = d$, we deduce that (2.2) holds, proving the first part of the theorem.

For the second part, suppose that $p = M/N$ is bounded below by $\frac{1}{4} + 5\log d/d$, and that $(1-p)\log n \to \infty$. Then the conditions of Theorem 2.1 are still satisfied, and it suffices to verify the additional conditions of Theorem 2.2. For (2.5) and (2.6) this is easy. As noted after the statement of Theorem 2.2, condition (2.7) holds as Q^d is Δ-regular. The argument for (2.8) is similar to that for (2.2) above: we first show that $\gamma' = (e(Q^d) - 4)/(n-4)$. □

Proof of Theorem 1.4. This time we take $H = L_k$, $n = k^2$ and $p = M/N$ such that $pn^{1/2} \to \infty$ and $p\log n \to 0$. As $e(H) = 2n(1 - n^{-1/2})$ we have $\alpha \sim 4n^{-1}$, and it is easy to see that (2.1) holds. The function $e_H(v)$ is implicitly given by Bollobás and Leader [4], who give an ordering on the vertices of the d-dimensional grid whose initial segments span as many edges as possible given their cardinality. In the special case of the lattice, these initial segments include the sublattices L_s, $s \leqslant k$, so, as noted in [4], $e_H(s^2) = 2s^2(1 - s^{-1})$. Also, from the precise description of the order it follows that for $1 \leqslant k \leqslant s$ we have $e_H(s^2 + k) = e_H(s^2) + 2k - 1$ and $e_H(s^2 + s + k) = e_H(s^2 + s) + 2k - 1$. To check (2.2) we use the observation that $e_H(v+1) \leqslant e_H(v) + 2$, which follows from these formulae, or directly

https://doi.org/10.1017/S0963548399004150 Published online by Cambridge University Press

from the nature of the ordering. Thus, since $e_H(v)/(v-2) = 2$ at $v = 3$, we have $\gamma(H) = 2$, and (2.2) holds. This shows that the conditions of Theorem 2.1 are satisfied, proving the first part of Theorem 1.4.

For the second part it is easy to verify (2.5)–(2.7). As $e_H(5) = 5$ we have $e_H(v) - 4 < 2(v-4)$ for $v = 5$. Since $e_H(v+1) \leqslant e_H(v) + 2$, this implies that $e_H(v) - 4 < 2(v-4)$ for all $v > 4$, so $\gamma' < 2$ and (2.8) holds. The second part of Theorem 1.4 thus follows from Theorem 2.2. □

In the next section we give an expansion of the variance of $X = X_H(G)$ as a sum over subgraphs of H. This expansion will be used in the proofs of both Theorems 2.1 and 2.2.

3. Expanding the second moment

The basic method we use, that of second moments, is very simple. We write $X_H(G)$ for the number of subgraphs of a graph G isomorphic to H. With G a random graph and $X = X_H(G)$, we apply Markov's inequality to bound $\mathbb{P}(X = 0)$ in terms of $\mathbb{E}(X)$ and $\mathrm{Var}(X)$. In turn, we bound $\mathrm{Var}(X)$ in terms of the extent to which the presence in G of one copy of H makes other copies more likely; this is made precise in the lemma below. In fact, we shall consider two graphs H and L, and the covariance of $X = X_H(G)$ and $Y = Y_L(G)$. This greater generality complicates the proof very little, and will be needed in Section 5.

Given graphs H, L on n vertices, let $H_1, \ldots, H_h$ be all copies of H with vertex set $[n]$, and $L_1, \ldots, L_l$ all copies of L. Thus h is $n!$ divided by the number of automorphisms of H, and similarly for l. For G a random graph, chosen from either $\mathscr{G}(n, p)$ or $\mathscr{G}(n, M)$, let X_i, Y_j be the indicator functions of the events $H_i \subset G$, $L_j \subset G$. Let $X = \sum_{i=1}^h X_i$, $Y = \sum_{j=1}^l Y_j$, so that, as before, X is the number of subgraphs of G isomorphic to H, and Y is the number isomorphic to L.

Lemma 3.1. *With the notation above, let*

$$f_{ij} = \mathbb{P}(X_i = 1, Y_j = 1)/(\mathbb{P}(X_i = 1)\mathbb{P}(Y_j = 1)),$$

and let $f = \frac{1}{hl}\sum_{i=1}^h \sum_{j=1}^l f_{ij}$. *Then*

$$\mathrm{Cov}(X, Y) = \mathbb{E}(X)\mathbb{E}(Y)(f - 1). \tag{3.1}$$

Furthermore, if $H = L$ *and* $f \lesssim 1$ *as* $n \to \infty$, *then*

$$\mathbb{P}(X = 0) \to 0.$$

Proof. Since neither $\mathbb{P}(X_i = 1)$ nor $\mathbb{P}(Y_i = 1)$ depends on i, we have

$$\begin{aligned} \mathbb{E}(XY) &= \mathbb{E}\left(\sum_{i=1}^h \sum_{j=1}^l X_i Y_j\right) = \sum_{i=1}^h \sum_{j=1}^l \mathbb{P}(X_i = 1, Y_j = 1) \\ &= \sum_{i=1}^h \sum_{j=1}^l f_{ij}\mathbb{P}(X_i = 1)\mathbb{P}(Y_j = 1) = hlf\mathbb{P}(X_1 = 1)\mathbb{P}(Y_1 = 1), \end{aligned}$$

while $\mathbb{E}(X) = \sum_{i=1}^h \mathbb{P}(X_i = 1) = h\mathbb{P}(X_1 = 1)$ and $\mathbb{E}(Y) = l\mathbb{P}(Y_1 = 1)$. Thus $\mathbb{E}(XY) = f\mathbb{E}(X)\mathbb{E}(Y)$, and (3.1) follows.

https://doi.org/10.1017/S0963548399004150 Published online by Cambridge University Press

For the second part, suppose that $H = L$. Writing μ for $\mathbb{E}(X)$ we have from (3.1) that $\mathrm{Var}(X)/\mu^2 = f - 1$. Note that $f \geqslant 1$, as $\mathrm{Var}(X) \geqslant 0$, so the assumption that $f \lesssim 1$ gives $f \to 1$. Since $(X-\mu)^2$ has expectation $\mathrm{Var}(X)$, we have from Markov's inequality that $\mathbb{P}(X = 0) \leqslant \mathbb{P}((X-\mu)^2 \geqslant \mu^2) \leqslant \mathrm{Var}(X)/\mu^2 = f - 1$, which tends to zero. □

Remark. For the case when $H = L$ is the d-dimensional cube Q^d, for example, we cannot hope to get reasonable bounds by applying Lemma 3.1 to the $\mathcal{G}(n, p)$ model. The reason is that in this model, the factors f_{ij} are given exactly by $p^{-e(H_i \cap H_j)}$. Since we want $p \sim \frac{1}{4}$, and most pairs of cubes on $[n]$ share about $d^2/2$ edges, the factor f will tend to infinity, not to 1. Another way of saying this is that $H = Q^d$ contains enough edges that the event $H_j \subset G$ pushes up the probability of $H_i \subset G$ just by pushing up the number of edges of G. For the rest of the proof we shall thus consider only the model $\mathcal{G}(n, M)$, writing $M = pN$, where $N = \binom{n}{2}$. The next step is to estimate the factors f_{ij} in this model.

Let us write $P(m)$ for the probability that some fixed set of m edges is present in $G \in \mathcal{G}(n, pN)$, so

$$P(m) = \binom{N-m}{pN-m}\binom{N}{pN}^{-1}.$$

Then f_{ij} depends only on $e(H_i \cap L_j)$ and can be written as $f(e(H_i \cap L_j))$, with

$$f(m) = P(e(H) + e(L) - m)P(e(H))^{-1}P(e(L))^{-1}.$$

While we could start estimating the f_{ij} directly, it turns out to be more convenient to make a simple transformation first. We shall think of H_i and L_j as independent random variables chosen uniformly from $H_1, \ldots, H_h$ and $L_1, \ldots, L_l$ respectively, writing $\mathbb{P}_{ij}$, $\mathbb{E}_{ij}$ for the probability and expectation in this probability space. Using this notation we have

$$f = \mathbb{E}_{ij} f_{ij} = \sum_{E \subset K_n} f(e(E))\mathbb{P}_{ij}(H_i \cap L_j = E),$$

where the sum is over all 2^N graphs E on $[n]$. Now the probability that the intersection of H_i and L_j is exactly some fixed E is much harder to calculate than the probability that this intersection contains some fixed F. We can express f in terms of these latter probabilities using the following simple lemma.

Lemma 3.2. *For any graphs H and L on $[n]$ we have*

$$f = \mathbb{E}_{ij} f_{ij} = \sum_{F \subset K_n} \overline{f}(e(F))\mathbb{P}_{ij}(F \subset H_i)\mathbb{P}_{ij}(F \subset L_j), \tag{3.2}$$

where

$$\overline{f}(m) = \binom{N - e(H) - e(L)}{pN - e(H) - e(L) + m}\binom{N}{pN}^{-1} P(e(H))^{-1}P(e(L))^{-1}. \tag{3.3}$$

Proof. For any integers $a, b \geqslant 0$ we have

$$\binom{a+m}{b+m} = \sum_k \binom{m}{m-k}\binom{a}{b+k} = \sum_k \binom{m}{k}\binom{a}{b+k}.$$

https://doi.org/10.1017/S0963548399004150 Published online by Cambridge University Press

Taking $a = N - e(H) - e(L)$, $b = pN - e(H) - e(L)$, it follows that $f(m) = \sum_k \binom{m}{k}\overline{f}(k)$. Now

$$\begin{aligned} f &= \sum_{E \subset K_n} f(e(E))\mathbb{P}_{ij}(H_i \cap L_j = E) \\ &= \sum_{E \subset K_n} \sum_k \binom{e(E)}{k} \overline{f}(k)\mathbb{P}_{ij}(H_i \cap L_j = E) \\ &= \sum_{E \subset K_n} \sum_{F \subset E} \overline{f}(e(F))\mathbb{P}_{ij}(H_i \cap L_j = E) \\ &= \sum_{F \subset K_n} \overline{f}(e(F))\mathbb{P}_{ij}(H_i \cap L_j \supset F). \end{aligned}$$

As H_i and L_j are chosen independently, the result follows. □

Note that $\mathbb{P}_{ij}(F \subset L_j) = X_F(L)/X_F(K_n)$, where, as before, $X_F(L)$ is the number of subgraphs of L isomorphic to F. An alternative expression for f is thus

$$f = \sum_{F \subset L} \overline{f}(e(F))X_F(H)/X_F(K_n). \tag{3.4}$$

In this section all we have done is to rewrite $\mathrm{Cov}(X, Y)$ in a different form. In the next two sections we use this expression to obtain the estimates we need.

4. Proof of Theorem 2.1

Up to this point we have been considering variables X, Y counting the number of copies of two graphs H, L in $G \in \mathcal{G}(n, M)$. Throughout this section we take $L = H$ and consider $\mathrm{Var}(X)$, aiming to show that under the conditions of Theorem 2.1 we have $f = \mathrm{Var}(X)/\mathbb{E}(X)^2 \sim 1$. As before, we are thus considering a sequence $H^{(i)}$ of fixed graphs with $n = n(i) = |H^{(i)}| \to \infty$. Throughout we shall write $e(H)$ as αN, where $N = \binom{n}{2}$, so α depends on i. We shall consider G from $\mathcal{G}(n, M)$, writing M as pN, $p = p(i)$. For brevity we suppress all dependence on i. All limits or statements involving $O()$, $o()$, $\sim$, $\lesssim$ are as $i \to \infty$, or, equivalently, as $n \to \infty$. The implicit functions involved depend only on the sequence $H^{(i)}$ and on i, not on any other parameters involved.

We now estimate the factors $P(m)$ and $\overline{f}(m)$ defined as in the previous section, taking $L = H$.

Lemma 4.1. *Suppose that αN, $pN \to \infty$, and that $\alpha^3 N p^{-2} \to 0$. Then*

$$P(\alpha N) \sim p^{\alpha N} e^{-\frac{1-p}{2p}\alpha^2 N},$$

and, for $0 \leqslant m \leqslant \alpha N$,

$$\overline{f}(m) \lesssim e^{-\frac{1-p}{p}\alpha^2 N} \left(\frac{1-p}{p-2\alpha}\right)^m. \tag{4.1}$$

Proof. Since $\alpha^2 p^{-2} = \alpha^3 N p^{-2}/(\alpha N) \to 0$, we have $\alpha = o(p)$, so $N - \alpha N \geqslant pN - \alpha N \to \infty$.

https://doi.org/10.1017/S0963548399004150 Published online by Cambridge University Press

We can thus apply Stirling's formula in the form $r! \sim r^r e^{-r}\sqrt{2\pi r}$ to obtain

$$P(\alpha N) = \binom{N-\alpha N}{pN-\alpha N}\binom{N}{pN}^{-1} = \frac{(N-\alpha N)!(pN)!}{(pN-\alpha N)!N!}$$
$$\sim \sqrt{\frac{(1-\alpha)p}{p-\alpha}}\left[(1-\alpha)^{1-\alpha}p^p(p-\alpha)^{-(p-\alpha)}\right]^N.$$

As $\alpha = o(p)$, the factor under the square root is $1+o(1)$. The first part of the lemma follows by expanding the logarithm of the factor in square brackets, neglecting terms of order $\alpha^3 p^{-2}$ or smaller.

For the second part note that $\overline{f}(0) = f(0) = P(2\alpha N)/P(\alpha N)^2 \sim e^{-\frac{1-p}{p}\alpha^2 N}$, by the first part. Also, as m varies $\overline{f}(m)$ is proportional to $\binom{N-2\alpha N}{pN-2\alpha N+m}$, so

$$\frac{\overline{f}(m+1)}{\overline{f}(m)} = \frac{N-pN-m}{pN-2\alpha N+m+1} < \frac{1-p}{p-2\alpha};$$

so (4.1) follows by induction on m. □

Together (3.4) (with $L = H$) and (4.1) give us a bound for f involving the numbers $X_F(H)$, $X_F(K_n)$ and $c = \frac{1-p}{p-2\alpha}$, namely

$$f \lesssim e^{-\frac{1-p}{p}\alpha^2 N}\sum_{F\subset H} c^{e(F)}\frac{X_F(H)}{X_F(K_n)}. \tag{4.2}$$

We would like to replace c by $\frac{1-p}{p} = p^{-1}-1$, but there is a small cost, which we express in terms of $r(F)$, the *rank* of F. Here, as usual, $r(F) = n - k(F)$, where $k(F)$ is the number of components of F. The quantity we shall consider is $S_H = S_H(p)$ given by

$$S_H = \sum_{F\subset H}(p^{-1}-1)^{e(F)}(1+n^{-\frac{1}{2}})^{r(F)}\frac{X_F(H)}{X_F(K_n)}.$$

We state what follows as a lemma even though it follows from (4.2) by straightforward calculation.

Lemma 4.2. *Suppose that conditions* (2.1), (2.3) *and* (2.4) *of Section 2 are satisfied. Then*

$$f \lesssim e^{-\frac{1-p}{p}\alpha^2 N} S_H.$$

Proof. Let $c = \frac{1-p}{p-2\alpha}$ as above. Under the given assumptions we have $\alpha = o(p)$, so

$$\log\left(\frac{c}{p^{-1}-1}\right) = \log\left(\frac{p}{p-2\alpha}\right) \leqslant \frac{3\alpha}{p}$$

for n large enough. Now as $\alpha \leqslant \Delta/(n-1)$ we have

$$\alpha p^{-1}\Delta \lesssim n^{-1}p^{-1}\Delta^2 = n^{-\frac{1}{2}}(np^2/\Delta^4)^{-\frac{1}{2}} = o(n^{-\frac{1}{2}}).$$

Thus, for n large enough,

$$\left(\frac{c}{p^{-1}-1}\right)^{e(F)} \leqslant e^{3\alpha p^{-1}e(F)} \leqslant e^{3\alpha p^{-1}\Delta r(F)} \leqslant (1+n^{-\frac{1}{2}})^{r(F)}.$$

Combined with (4.2) this completes the proof of the lemma. □

https://doi.org/10.1017/S0963548399004150 Published online by Cambridge University Press

The above result tells us that to estimate f it suffices to estimate S_H. We do this in two stages, noting that since we want $f \sim 1$ we must bound the large terms in S_H reasonably accurately.

Suppose, for example, that $H = Q^d$. Then most copies of H intersect in about $d^2/2$ edges, so we expect the main contribution to S_H to come from F with $e(F) = O(d^2)$. Since d^2 is very small with respect to n, almost all such F will consist of independent edges. With this in mind, we say that an edge e of a graph F is *isolated* if it forms a component of F of order 2, and that a graph F on $[n]$ is *good* if F has no isolated edges. We expect the main terms in S_H to come from adding about $O(d^2)$ isolated edges to the (good) empty graph. Let

$$S'_H = \sum_{F\subset H}{}' (p^{-1}-1)^{e(F)} 2^{r(F)} \frac{X_F(H)}{X_F(K_n)},$$

where from now on $\sum'$ represents summation only over good graphs. The next lemma bounds f in terms of S'_H. Note that the factor $2^{r(F)}$ will turn out to be irrelevant later, and is included to give us a little space when comparing S_H and S'_H.

Lemma 4.3. *If H is any graph with maximum degree Δ, and $np^2/\Delta^4 \to \infty$, then*

$$S_H \lesssim e^{\frac{1-p}{p}\alpha^2 N} S'_H.$$

Proof. We start by writing S_H and S'_H as sums over the isomorphism class $[F]$ of $F \subset H$. Thus,

$$S_H = \sum_{[F]:F\subset H} (p^{-1}-1)^{e(F)} (1+n^{-\frac{1}{2}})^{r(F)} \frac{X_F(H)^2}{X_F(K_n)}, \tag{4.3}$$

and

$$S'_H = \sum_{[F]:F\subset H}{}' (p^{-1}-1)^{e(F)} 2^{r(F)} \frac{X_F(H)^2}{X_F(K_n)}, \tag{4.4}$$

where the prime denotes restriction to good F.

Let F be any good graph with v isolated vertices, and let F_t be obtained from F by adding $t \leqslant \frac{v}{2}$ new edges, isolated in F_t. We shall write $S[F,t]$ for the contribution of $[F_t]$ to (4.3), and $S[F]$ for $\sum_{t\leqslant v/2} S[F,t]$. We shall also write $S'[F]$ for the contribution of $[F]$ to (4.4). Thus $S'_H = \sum'_{[F]} S'[F]$ by definition, and $S_H = \sum'_{[F]} S[F]$ since any $E \subset H$ can be obtained from a good $F \subset H$ by adding isolated edges. It follows that any upper bound for the ratio $S[F]/S'[F]$ which depends only on n (not on F) is also an upper bound for S_H/S'_H. Now

$$X_{F_t}(K_n) = X_F(K_n)\frac{1}{t!}\binom{v}{2}\binom{v-2}{2}\cdots\binom{v-2t+2}{2},$$

while

$$X_{F_t}(H) \leqslant X_F(H)\frac{1}{t!} e_H(v)e_H(v-2)\cdots e_H(v-2t+2).$$

Writing β_w for $e_H(w)^2/\binom{w}{2}$, we thus have that

$$\frac{X_{F_t}(H)^2}{X_{F_t}(K_n)} \leqslant \frac{X_F(H)^2}{X_F(K_n)} \frac{\beta_v\beta_{v-2}\cdots\beta_{v-2t+2}}{t!}.$$

https://doi.org/10.1017/S0963548399004150 Published online by Cambridge University Press

Hence, as $e(F_t) = e(F) + t$ and $r(F_t) = r(F) + t$, we have

$$\frac{S[F,t]}{S'[F]} \leqslant 2^{-r(F)}(p^{-1}-1)^t(1+n^{-\frac{1}{2}})^{r(F)+t}\frac{\beta_v\beta_{v-2}\cdots\beta_{v-2t+2}}{t!}.$$

Summing over t we obtain, for n large enough,

$$\frac{S[F]}{S'[F]} \leqslant 1.5^{-r(F)}\sum_{t=0}^{\infty}(p^{-1}-1)^t(1+n^{-\frac{1}{2}})^t\frac{\beta_v\beta_{v-2}\cdots\beta_{v-2t+2}}{t!},$$

where we take $\beta_w = 0$ for $w < 2$.

From now on we suppose that $np^2/\Delta^4 \to \infty$. As H has maximum degree Δ we have $e_H(w) \leqslant \min\{\binom{w}{2}, \frac{w\Delta}{2}\}$, and hence that $\beta_w \leqslant \min\{\binom{w}{2}, (\frac{w\Delta}{2})^2\binom{w}{2}^{-1}\}$. As the first term increases and the second decreases, and they are equal when $w = \Delta+1$, we have $\beta_w \leqslant \binom{\Delta+1}{2} \leqslant \Delta^2$ for all w. Hence $p^{-1}\beta_w \leqslant p^{-1}\Delta^2 = o(\sqrt{n})$, and the error in truncating the sum above at $\sqrt{n}$ tends to zero. Thus

$$\frac{S[F]}{S'[F]} \lesssim 1.5^{-r(F)}\sum_{t=0}^{\sqrt{n}}(p^{-1}-1)^t(1+n^{-\frac{1}{2}})^t\frac{\beta_v\beta_{v-2}\cdots\beta_{v-2t+2}}{t!}. \tag{4.5}$$

We claim that $S[F]/S'[F] \lesssim e^{\frac{1-p}{p}\alpha^2 N}$ for any F. To show this, suppose first that $v \leqslant n-\sqrt{n}$. Using $\beta_w \leqslant \Delta^2$ we have from (4.5) that

$$\begin{aligned}\frac{S[F]}{S'[F]} &\lesssim 1.5^{-r(F)}\exp\{(p^{-1}-1)(1+n^{-\frac{1}{2}})\Delta^2\}\\ &\leqslant 1.5^{-r(F)}e^{2\Delta^2p^{-1}}.\end{aligned}$$

Since $r(F) \geqslant \frac{1}{2}(n-v) \geqslant \frac{1}{2}\sqrt{n}$ and $\Delta^2p^{-1} = o(\sqrt{n})$, we have $S[F]/S'[F] = o(1)$. As $e^{\frac{1-p}{p}\alpha^2N} \geqslant 1$, this proves the claim in this case.

Suppose next that $v \geqslant n-\sqrt{n}$. This time we use the bound $\beta_w \leqslant e(H)^2/\binom{w}{2} = \alpha^2N\binom{n}{2}/\binom{w}{2}$, which gives $\beta_w = \alpha^2N(1+O(n^{-1/2}))$ for $w \geqslant n-3\sqrt{n}$. Ignoring, as we may, the factor $1.5^{-r(F)}$, we thus have from (4.5) that

$$\frac{S[F]}{S'[F]} \lesssim \exp\{(p^{-1}-1)\alpha^2N(1+O(n^{-\frac{1}{2}}))\} \sim e^{\alpha^2N\frac{1-p}{p}},$$

as $p^{-1}\alpha^2N = O(p^{-1}\Delta^2) = o(\sqrt{n})$. This proves the claim.

Since $S[F]/S'[F] \lesssim e^{\alpha^2N\frac{1-p}{p}}$ for any F, we have $S_H/S'_H \lesssim e^{\alpha^2N\frac{1-p}{p}}$, completing the proof of the lemma. □

Our next aim is to give a simple bound for the ratio $X_F(H)/X_F(K_n)$ in terms of the rank of F. Since we expect the main term in S'_H to come from the empty graph, which contributes 1, we shall be fairly generous with the other terms.

Lemma 4.4. *Let H be any graph on $[n]$ with maximum degree Δ, and $F \subset H$. Then*

$$\frac{X_F(H)}{X_F(K_n)} \leqslant \left(\frac{2e\Delta}{n}\right)^{r(F)}.$$

https://doi.org/10.1017/S0963548399004150 Published online by Cambridge University Press

Proof. For a graph L on $[n]$ let $Y_F(L)$ be the number of bijections from $[n]$ to $[n]$ mapping all edges of F into edges of L. Since $Y_F(L)$ is just $X_F(L)$ multiplied by the number of automorphisms of F, we have $Y_F(H)/Y_F(K_n) = X_F(H)/X_F(K_n)$. Now $Y_F(K_n) = n!$. What about $Y_F(H)$? We shall go through the vertices of F one by one in some order, bounding the number of choices for the image of each vertex given our previous choices. Suppose F has i isolated vertices and $k' = k(F) - i$ other components, so $r(F) = n - k' - i$. We order the vertices of F as follows. Start with one vertex v_i from each nontrivial component C_i of F. Then go through the components C_i listing their remaining vertices such that each is preceded by at least one of its neighbours. (For example, write the vertices of C_i in order of increasing distance from v_i.) Then list the isolated vertices in any order. For the first k' vertices we have $(n)_{k'}$ choices for the images. If v is one of the next $r(F)$ vertices we have at most Δ choices, as we have already chosen the image of a neighbour of v. For the last i vertices we have $i!$ choices. Thus

$$Y_F(H) \leqslant (n)_{k'}\Delta^{r(F)}i! = \frac{\Delta^{r(F)}}{(n-k')_{r(F)}} Y_F(K_n).$$

Now for any $0 \leqslant b \leqslant a$ we have $(a)_b \geqslant (a/e)^b$: if this assertion fails anywhere it is at $b = a$, where it follows from Stirling's formula. Since $k' \leqslant n/2$ we have $(n-k')_{r(F)} \geqslant (n/2e)^{r(F)}$, and the lemma follows. □

From now on we consider the sum

$$T'_H = \sum_{F\subset H}{}' (p^{-1}-1)^{e(F)} \left(\frac{15\Delta}{n}\right)^{r(F)},$$

since we have from Lemmas 4.2, 4.3 and 4.4 that, under the conditions of Theorem 2.1,

$$f \lesssim e^{-\frac{1-p}{p}\alpha^2 N} S_H \lesssim S'_H \leqslant \sum_{F\subset H}{}' (p^{-1}-1)^{e(F)} \left(\frac{4e\Delta}{n}\right)^{r(F)} \leqslant T'_H.$$

Note that T'_H is just an evaluation of the rank-generating function of H, or, equivalently, of the Tutte polynomial of H, although we shall not make use of this fact.

We make one last simplification before directly bounding T'_H, reducing the sum to one over 'connected' graphs F. Here we mean graphs F on $[n]$ with exactly one nontrivial component, or $k'(F) = 1$ in the notation above. To do this, note that $\psi(F) = (p^{-1} - 1)^{e(F)}(15\Delta/n)^{r(F)}$ is *multiplicative*, in the sense that $\psi(F_1 \cup F_2) = \psi(F_1)\psi(F_2)$ when F_1 and F_2 are vertex-disjoint (ignoring isolated vertices). Writing $\sum''$ for summation over connected good graphs, since every good graph is a disjoint union of connected good graphs, we have

$$\sum_{F\subset H}{}' \psi(F) \leqslant 1 + \sum_{t=1}^{\infty} \frac{1}{t!} \left(\sum_{F\subset H}{}'' \psi(F)\right)^t,$$

that is,

$$T'_H \leqslant e^{T''_H}, \tag{4.6}$$

https://doi.org/10.1017/S0963548399004150 Published online by Cambridge University Press

where

$$T_H'' = \sum_{F\subset H}'' (p^{-1}-1)^{e(F)} \left(\frac{15\Delta}{n}\right)^{r(F)}.$$

To complete the proof of Theorem 2.1 we now bound T_H'' directly.

Lemma 4.5. *For every graph H on $[n]$ we have*

$$T_H'' \leqslant 50\Delta^2 \sum_{s=3}^{n} \left(\frac{50\Delta^2}{n}\right)^{s-2} p^{-e_H(s)}, \tag{4.7}$$

where Δ is the maximum degree of H.

Proof. We shall split the sum T_H'' according to $s = r+1$, the number of vertices in the unique nontrivial component F' of F, and according to $m = e(F)$. Note that an F with these parameters contributes $(p^{-1}-1)^m(15\Delta/n)^{s-1}$ to T_H''. For s and m given, how many such F are there? For each $V \subset [n]$ with $|V| = s$ we have at most $\binom{e_H(s)}{m}$ such F with $V(F') = V$. Thus

$$\begin{aligned} T_H'' &\leqslant \sum_{s=3}^{n} \left(\frac{15\Delta}{n}\right)^{s-1} \sum_{V}\sum_{m=0}^{e_H(s)} \binom{e_H(s)}{m} (p^{-1}-1)^m \\ &= \sum_{s=3}^{n} \left(\frac{15\Delta}{n}\right)^{s-1} \sum_{V} p^{-e_H(s)}, \end{aligned}$$

where the sums over V are restricted to $V \subset [n]$ with s elements such that the subgraph of H induced by V is connected. How many such V are there? Certainly at most the number $t_H(s)$ of subtrees of H with s vertices. In fact, we consider labelled trees.

Let $\mathscr{T}(s)$ be the set of trees with vertex set $[s]$, so $|\mathscr{T}(s)| = s^{s-2}$ by Cayley's formula. Let $\Phi(s)$ be the set of pairs (T,ϕ) with $T \in \mathscr{T}(s)$ and ϕ a 1–1 edge-preserving map from T to H. As H has maximum degree Δ, each $T \in \mathscr{T}(s)$ can be mapped into H in at most $n\Delta^{s-1}$ ways, so $|\Phi(s)| \leqslant n\Delta^{s-1}s^{s-2}$. Now each subtree of H with s vertices arises exactly $s!$ times as the image $\phi(T)$ with $(T,\phi) \in \Phi(s)$: once for each labelling of its vertices with $1,2,\dots,s$. Thus $t_H(s)$, and hence the number of choices for V, as above, is bounded by $n\Delta^{s-1}s^{s-2}/s!$. Since $s! \geqslant (s/e)^s$, this quantity is at most $nes^{-2}(e\Delta)^{s-1} < n(e\Delta)^{s-1}$, so we have

$$T_H'' \leqslant n \sum_{s=3}^{n} \left(\frac{15e\Delta^2}{n}\right)^{s-1} p^{-e_H(s)}.$$

Taking one factor of $15e\Delta^2/n$ out of the sum, and noting that $15e < 50$, we obtain (4.7), completing the proof. □

We have now done all the work required to prove Theorem 2.1. All that remains is some straightforward calculation.

Proof of Theorem 2.1. We assume, as we may, that (2.1) holds, and also that either (2.2) or (2.2′) holds. As noted in Section 2, conditions (2.3) and (2.4) are then also satisfied. We consider the quantity f defined in Lemma 3.1. By Lemma 3.1, it suffices to show that

https://doi.org/10.1017/S0963548399004150 Published online by Cambridge University Press

$f \lesssim 1$. Also, by Lemmas 4.2, 4.3 and 4.4 and from (4.6) we have

$$f \lesssim e^{-\frac{1-p}{p}\alpha^2 N} S_H \lesssim S'_H \leqslant T'_H \leqslant e^{T''_H},$$

so it suffices to show that $T''_H \to 0$.

From Lemma 4.5 and the definition of γ we have

$$T''_H \leqslant 50\Delta^2 \sum_{s=3}^{\infty} \left(\frac{50\Delta^2 p^{-\gamma}}{n}\right)^{s-2} \leqslant \sum_{s=3}^{\infty} \left(\frac{10^4\Delta^4 p^{-\gamma}}{n}\right)^{s-2}. \tag{4.8}$$

We now consider two cases, according to which one of (2.2) and (2.2′) holds.

If (2.2) holds then $\Delta^4 p^{-\gamma} n^{-1} \to 0$, so from (4.8) we have $T''_H \to 0$ as required.

If (2.2′) holds the argument is slightly more complicated. This time we have $p^{-1} \sim 1$. Thus the first $O(1)$ terms in (4.7) are of the form

$$O\left(\frac{\Delta^4}{n}\left(\frac{50\Delta^2}{n}\right)^{s-3}\right).$$

As $\Delta = o(n^{1/4})$ these terms, and their sum, tend to zero. When s is large we have $\Delta^2(50\Delta^2 p^{-\gamma} n^{-1})^{s-2} \leqslant (\Delta^{2+\epsilon} p^{-\gamma} n^{-1})^{s-2} = (o(1))^{s-2}$, so the sum of the remaining terms in (4.7) tends to zero. Thus $T''_H \to 0$ in this case also, completing the proof of Theorem 2.1. □

5. Proof of Theorem 2.2

In this section we prove Theorem 2.2. One might think that, to show the convergence in distribution, estimates of all the moments of $X = X_H(G)$ would be required. However, as in the context of [6], the variance estimate (2.9) turns out to be enough.

As before, let $X = X_H(G)$ be the number of copies of the given graph H in $G \in \mathcal{G}(n, M)$, $M = pN$. Let Y be the number of copies of P_2 (a pair of adjacent edges) in G. As noted in [6], if it should happen that

$$\mathrm{Cov}(X, Y)^2 \sim \mathrm{Var}(X)\mathrm{Var}(Y), \tag{5.1}$$

then X must be approximately a linear function of Y. More precisely, writing X^* for $(X - \mathbb{E}(X))/\sqrt{\mathrm{Var}(X)}$, Y^* similarly, and σ for the sign (± 1) of $\mathrm{Cov}(X, Y)$, we can rewrite (5.1) as

$$\mathbb{E}((X^* - \sigma Y^*)^2) \to 0.$$

Thus, if $Y^* \to^d N(0, 1)$ then $X^* \to^d N(0, 1)$ as claimed in Theorem 2.2. Now the conditions of Theorem 2.2 imply that $pn^{1/2} \to \infty$ (from (2.3)) and that $(1 - p)n \to \infty$. From [6] we know that under these conditions,

$$\frac{\mathrm{Var}(Y)}{\mathbb{E}(Y)^2} \sim \left(\frac{1-p}{p}\right)^2 \frac{2}{n^3},$$

and that $Y^* \to^d N(0, 1)$. Thus, to prove Theorem 2.2 it suffices to prove the variance estimate (2.9) together with

$$\frac{\mathrm{Cov}(X, Y)}{\mathbb{E}(X)\mathbb{E}(Y)} \sim (\alpha_2 - \alpha^2)\left(\frac{1-p}{p}\right)^2. \tag{5.2}$$

https://doi.org/10.1017/S0963548399004150 Published online by Cambridge University Press

To prove the variance estimate (2.9) it seems that we must repeat the calculations of the previous section to much greater accuracy. In [6] the corresponding calculations for much simpler graphs become rather messy, requiring a four-term expansion of the falling factorial $(k)_l$. However, the simple form of the answer suggests that a different approach should be possible, and this is indeed the case. We again start from relation (3.2) from Section 3. We use the results of Section 4 to show that all but a few 'typical' graphs contribute very little to the right-hand side of (3.2). We then recalculate the contribution from these typical graphs more accurately using a trick described below. We phrase the argument in terms of two graphs H and L so that we can apply the method to prove (5.2) as well as (2.9).

Let H and L be graphs on $[n]$ with $e(H) = \alpha N$, $e(L) = \beta N$. Writing $P_H(J)$ ($P_L(J)$) for the probability that a given graph J on $[n]$ is contained in a random copy of H (L) on $[n]$, we have from Lemma 3.2 that

$$f = \frac{\mathbb{E}(XY)}{\mathbb{E}(X)\mathbb{E}(Y)} = \sum_{J \subset K_n} \overline{f}(e(J)) P_H(J) P_L(J),$$

where, as before, $X = X_H(G)$, $Y = X_L(H)$, $G \in \mathcal{G}(n, pN)$ and $\overline{f}(m)$ is given by (3.3). Now $f \sim 1$, and we wish to estimate $f - 1$, so the trick is to subtract off 1 before we start. To do this we must express 1 in a suitable form.

Let X' count all subgraphs of G with αN edges, and Y' all subgraphs with βN edges. The argument giving (3.2) shows that

$$\frac{\mathbb{E}(X'Y')}{\mathbb{E}(X')\mathbb{E}(Y')} = \sum_{J \subset K_n} \overline{f}(e(J)) P_{\alpha N}(J) P_{\beta N}(J), \tag{5.3}$$

where $P_{\alpha N}(J)$ is the probability that J is contained in a random αN edge graph on $[n]$. However, $X' = \binom{pN}{\alpha N}$ and $Y' = \binom{pN}{\beta N}$ are constants, so the left-hand side of (5.3) is just 1, and

$$f - 1 = \sum_{J \subset K_n} \overline{f}(e(J))(P_H(J)P_L(J) - P_{\alpha N}(J)P_{\beta N}(J)). \tag{5.4}$$

Now, for any j, the graph H has exactly as many j-edge subgraphs as any other αN edge graph. Thus, as J varies over all j-edge graphs on $[n]$, the $P_H(J)$ average to the constant $P_{\alpha N}(J) = P_{\alpha N}(j)$, say, and similarly for the $P_L(J)$. Thus, as pointed out by Svante Janson, we may rewrite (5.4) as

$$f - 1 = \sum_{J \subset K_n} \overline{f}(e(J))(P_H(J) - P_{\alpha N}(e(J)))(P_L(J) - P_{\beta N}(e(J))).$$

In estimating the expression above it is convenient to rewrite the sum as an average: let $M_j = \binom{N}{j}$, the number of j-edge graphs on $[n]$, and let

$$W_j = \overline{f}(j) P_{\alpha N}(j) P_{\beta N}(j) M_j.$$

Also, let $\rho_H(J) = P_H(J)/P_{\alpha N}(e(J))$ and $\rho_L(J) = P_L(J)/P_{\beta N}(e(J))$. Assuming that $\beta \leqslant \alpha$, we have

$$f - 1 = \sum_{j=0}^{\beta N} W_j \mathbb{E}_{e(J)=j}[(\rho_H(J) - 1)(\rho_L(J) - 1)], \tag{5.5}$$

https://doi.org/10.1017/S0963548399004150 Published online by Cambridge University Press

where the expectation is over a uniform choice of J from all j-edge graphs on $[n]$. Also, for each j, $0 \leqslant j \leqslant \beta N$, we can rewrite the statement about the average over J of the $P_H(J)$ as

$$\mathbb{E}_{e(J)=j}(\rho_H(J)-1) = \mathbb{E}_{e(J)=j}(\rho_L(J)-1) = 0. \tag{5.6}$$

Thus, as $\rho_H(J)$ and $\rho_L(J)$ depend only on the isomorphism class of J, and there is only one isomorphism class for $e(J) = 0, 1$, the $j = 0$ and $j = 1$ terms of the sum in (5.5) are always zero.

Note that the W_j are nonnegative (from the formula (3.3) for $\overline{f}(j)$). Also, the sum of all the W_j is the right-hand side of (5.3), and is thus equal to 1. The final fact we shall use about the W_j is that

$$\frac{W_{j+1}}{W_j} = \frac{N-pN-j}{pN-\alpha N-\beta N+j+1}\frac{\alpha N-j}{N-j}\frac{\beta N-j}{N-j}\frac{N-j}{j+1}, \tag{5.7}$$

which follows from (3.3), $P_{\alpha N}(j) = \binom{N-j}{\alpha N-j}\binom{N}{\alpha N}^{-1}$ and $M = \binom{N}{j}$.

Having derived (5.5) to prove (2.9), we first use it to prove the covariance estimate, which on its own could probably be proved more simply by a more direct method.

Lemma 5.1. *Under the conditions of Theorem 2.2, with $X = X_H(G)$, $L = P_2$, $Y = Y_L(G)$ and $G \in \mathcal{G}(n, pN)$, the covariance estimate (5.2) holds.*

Proof. We shall only use the following consequences of the conditions of Theorem 2.2: that $\alpha N \to \infty$, $\alpha = o(p)$ and $(1-p)N \to \infty$, and the very weak condition that $\frac{\alpha_2}{\alpha_2-\alpha^2} = o(\alpha N)$. As $L = P_2$ we have $\beta N = 2$, so for $j = 0, 1$

$$\frac{W_{j+1}}{W_J} \sim \frac{1-p}{p}\alpha\frac{2-j}{j+1},$$

while $W_j = 0$ for $j > 2$. Thus $W_1/W_0 \sim 2\alpha\frac{1-p}{p} = o(1)$ and $W_2/W_1 \sim \frac{\alpha}{2}\frac{1-p}{p} = o(1)$. As $W_0 + W_1 + W_2 = 1$, this implies that $W_0 \sim 1$ and $W_2 \sim \alpha^2\left(\frac{1-p}{p}\right)^2$. Since the $j = 0$ and $j = 1$ terms of (5.5) are always zero, this gives

$$f_{X,Y} - 1 = \frac{\mathrm{Cov}(X,Y)}{\mathbb{E}(X)\mathbb{E}(Y)} \sim \alpha^2\left(\frac{1-p}{p}\right)^2 \mathbb{E}_{e(J)=2}[(\rho_H(J)-1)(\rho_L(J)-1)].$$

Now the fraction δ of 2-edge graphs on $[n]$ which are P_2s tends to zero. Together with (5.6) this immediately implies that the $J = P_2$ term dominates the expectation above, as the factors $(\rho_H(J)-1)$ and $(\rho_L(J)-1)$ are both $-\frac{1-\delta}{\delta} \sim -\delta^{-1}$ times as large for $J \cong P_2$ as for $J \cong 2K_2$, that is, J consisting of two isolated edges. Thus

$$f_{X,Y} - 1 \sim \alpha^2\left(\frac{1-p}{p}\right)^2 \delta(\rho_H(P_2)-1)(\rho_L(P_2)-1).$$

Now $\rho_L(2K_2) = 0$, as $L = P_2$, so $\rho_L(P_2) - 1 = \frac{1-\delta}{\delta} \sim \delta^{-1}$. Also, $P_H(P_2) = \alpha_2$, from the definition of α_2, while $P_{\alpha N}(2) = \binom{N-2}{\alpha N-2}\binom{N}{\alpha N}^{-1} = \alpha^2(1+O(\alpha^{-1}N^{-1}))$. Thus $\rho_H(P_2) = \frac{\alpha_2}{\alpha^2}(1+O(\alpha^{-1}N^{-1}))$, and

$$\rho_H(P_2) - 1 = \frac{\alpha_2-\alpha^2}{\alpha^2}\left(1 + O\left(\frac{\alpha_2}{\alpha_2-\alpha^2}\alpha^{-1}N^{-1}\right)\right).$$

https://doi.org/10.1017/S0963548399004150 Published online by Cambridge University Press

By assumption the $O()$ term above tends to zero, so $f_{X,Y} - 1 \sim (\alpha_2 - \alpha^2)\left(\frac{1-p}{p}\right)^2$, completing the proof of the lemma. □

We now turn to the variance estimate (2.9), whose proof will complete the proof of Theorem 2.2. The basic idea is simple: taking $L = H$ and $X = X_H(G)$ we can write $f - 1 = \mathrm{Var}(X)/\mathbb{E}(X)^2$ as

$$\sum_{j=0}^{\alpha N} W_j \mathbb{E}_{e(J)=j}[(\rho_H(J) - 1)^2]. \tag{5.8}$$

We shall show that only the terms with j near $j_0 = \frac{1-p}{p}\alpha^2 N$ matter, and that within these terms only the graphs $J_0 = jK_2$, consisting of j isolated edges, and $J_1 = P_2 \cup (j-2)K_2$, consisting of two adjacent edges and $j-2$ isolated ones, contribute significantly to the expectation. Then, as there are many fewer graphs J isomorphic to J_1 than to J_0, the J_1 term will dominate, and we only have to estimate $\rho_H(J_1) - 1$. Unfortunately, showing that the terms described as insignificant are indeed insignificant requires some work.

For the rest of this section we fix a function $\epsilon = \epsilon(n)$ which tends to zero very slowly – precise assumptions as to how slowly will be made later. We say that a number j of edges is *typical* if $|j - j_0| \leqslant \epsilon j_0$, and *atypical* otherwise. We say that a graph J is *typical* if it has a typical number of edges and is of the form J_0 or J_1 above, and *atypical* otherwise. Under the conditions of Theorem 2.2 we have from (2.3) that $\Delta^4 p^{-2} n^{-1} \to 0$. As $\alpha n \lesssim \Delta$, this gives $\alpha^4 n^3 p^{-2} \to 0$. Since $j_0 = \frac{1-p}{p}\alpha^2 N \leqslant \alpha^2 n^2 p^{-1}$ this implies that

$$j_0 = o(n^{1/2}), \tag{5.9}$$

and hence that $j_0 = o(N - pN)$. Also, as $\alpha = o(p)$ we have $j_0 = o(\alpha N)$. Thus, from (5.7),

$$\frac{W_{j+1}}{W_j} \sim \frac{1-p}{p}\alpha^2 \frac{N}{j+1} = \frac{j_0}{j+1}$$

uniformly in j, for $0 \leqslant j \leqslant 2j_0$, say, while $W_{j+1}/W_j \lesssim j_0/j$ for $j > 2j_0$. As $\sum W_j = 1$, it follows that if $\epsilon \to 0$ sufficiently slowly then

$$\sum_{|j-j_0|>\epsilon j_0} W_j \lesssim 2e^{-\epsilon^2 j_0/3} = o(n^{-2}),$$

using the condition (2.5) that $j_0/\log n \to \infty$. As the estimate $(\alpha_2 - \alpha^2)^2(p^{-1}-1)^2 n^3/2$ for $f-1$ we are trying to prove has magnitude at least some constant times $\Delta^2(p^{-1}-1)^2 n^{-1}$, we shall say that a quantity is *small* if it is $o(\Delta^2(p^{-1}-1)^2 n^{-1})$. In particular, $\sum_{|j-j_0|>\epsilon j_0} W_j$ is small.

For typical j the fraction of j-edge graphs that are atypical is of order $O(j_0^4/n^2)$, as two coincidences are required among the $O(j_0^2)$ endpoints of j edges. As $\alpha n \lesssim \Delta$,

$$j_0^4 n^{-2} = O((p^{-1}-1)^4 \alpha^8 n^6) = O(\Delta^6 n^{-1} p^{-2} \Delta^2 (p^{-1}-1)^2 n^{-1}).$$

Since $\Delta^6 n^{-1} p^{-2} \to 0$ from (2.6), this shows that $j_0^4 n^{-2}$ is small. Since the W_j sum to one, and the sum over atypical j of the W_j is small, this shows that

$$\sum_j W_j \mathbb{E}_{e(J)=j}[1_{\{J \text{ atypical}\}} 1^2] = o(\Delta^2(p^{-1}-1)^2 n^{-1}), \tag{5.10}$$

where 1_A is the indicator function of the event A.

https://doi.org/10.1017/S0963548399004150 Published online by Cambridge University Press

Given (5.10), to show that the contribution of atypical J to (5.8) is small, it suffices to prove

$$\sum_j W_j \mathbb{E}_{e(J)=j}[1_{\{J \text{ atypical}\}}\rho_H(J)^2] = o(\Delta^2(p^{-1}-1)^2 n^{-1}). \tag{5.11}$$

Now the left-hand side of (5.11) is just a different way of writing

$$\sum_{J\subset K_n,\ J \text{ atypical}} \overline{f}(e(J))\mathbb{P}_{ij}(J\subset H_i)^2,$$

that is, the contribution of atypical J to (3.2) (with $H{=}L$), so we can use the calculations from the previous sections. The reader who is prepared to take on trust that the conditions of Theorem 2.2 have been chosen to ensure that (5.11) follows from these calculations may wish to skip the next two lemmas and their proofs.

Lemma 5.2. *Under the conditions of Theorem 2.2 we have*

$$\Delta^2\sum_{s\geqslant 4}\left(\frac{50\Delta^2}{n}\right)^{s-2} p^{-e_H(s)} = o(\Delta^2(p^{-1}-1)^2n^{-1}).$$

Proof. From (2.6) we have $\Delta^4 n^{-1}p^{-2} = o((1-p)^2)$, so it suffices to show that

$$p^4n^2\Delta^{-4}\sum_{s\geqslant 4}\left(\frac{50\Delta^2}{n}\right)^{s-2} p^{-e_H(s)} = 2500\sum_{s\geqslant 4}\left(\frac{50\Delta^2}{n}\right)^{s-4} p^{4-e_H(s)}$$

is bounded. Now H is triangle-free, so $e_H(4)\leqslant 4$, and the $s=4$ term of the second sum above is at most one. Thus it suffices to show that

$$\sum_{s>4}\left(\frac{50\Delta^2}{n}p^{-\frac{e_H(s)-4}{s-4}}\right)^{s-4} = O(1). \tag{5.12}$$

If (2.8) holds then (5.12) follows easily, as, by definition $\frac{e_H(s)-4}{s-4}\leqslant\gamma'$, and $\Delta^2p^{-\gamma'}n^{-1}\to 0$. Suppose instead that (2.2′) holds. As $e_H(s)\leqslant s^2$, the difference $\frac{e_H(s)-4}{s-4}-\frac{e_H(s)}{s-2}$ is bounded, so $\frac{e_H(s)-4}{s-4}\leqslant\gamma+O(1)$. Thus, as $p\sim 1$, we have $p^{-\frac{e_H(s)-4}{s-4}}\lesssim p^{-\gamma}$. Combined with the consequence $\Delta^2p^{-\gamma}n^{-1}\to 0$ of (2.2′), this proves (5.12), completing the proof of the lemma. □

Using the above lemma, showing that the contribution of atypical graphs to (5.8) is small is now just a matter of tracing through the calculations of the previous section.

Lemma 5.3. *Under the conditions of Theorem 2.2,* (5.11) *holds.*

Proof. As noted earlier, the left-hand side of (5.11) is just another way of writing the contribution to (3.2) (with $L=H$) from atypical graphs F. We work backwards through the bounds given in the previous section for this quantity, using the word ‘small’ to mean $o(\Delta^2(p^{-1}-1)^2n^{-1})$, as before.

Let us say that a graph F with no isolated edges is *unusual* if F is neither the empty graph nor P_2. Thus a graph J is atypical if it has an unusual subgraph, or an atypical number of edges. From Lemma 5.2 the contribution to the right-hand side of (4.7) from terms with

https://doi.org/10.1017/S0963548399004150 Published online by Cambridge University Press

$s \geqslant 4$ is small. The proof of Lemma 4.5 gives (4.7) as a term-by-term bound for T''_H, so the contribution to T''_H from all connected $F \subset H$ with at least four vertices is small. As H contains no triangles, this is exactly the contribution from all unusual connected $F \subset H$.

From the argument giving (4.6), that $T'_H \leqslant e^{T''_H}$, and as $T''_H \to 0$, we can deduce that the contribution to T'_H from all F with an unusual component is small. For F consisting of two or more P_2s we need a separate calculation: the $s = 3$ term in (4.7) is $O(\Delta^4 p^{-2} n^{-1})$, which tends to zero, so such F contribute at most $O(\Delta^8 p^{-4} n^{-2})$ to T'_H. That this is small is exactly condition (2.6). Thus the total contribution to T'_H from all unusual F is small. By Lemma 4.4, T'_H gives a term-by-term bound for S'_H, so the same conclusion holds for S'_H.

Let us consider a connected graph F. In the notation of Lemma 4.3 the proof of that lemma gives that $S[F] \lesssim e^{\frac{1-p}{p}\alpha^2 N} S'[F] = e^{j_0} S'[F]$. Thus the contribution to S_H from all graphs formed by adding isolated edges to unusual graphs is $o(e^{j_0}\Delta^2(p^{-1}-1)^2 n^{-1})$. Also, with F empty or $F = P_2$, the proof of Lemma 4.3 gives that

$$\sum_{|t-j_0|\geqslant \epsilon j_0 - 2} S[F,t]/S'[F] \lesssim \sum_{|t-j_0|\geqslant \epsilon j_0 - 2} \left((p^{-1}-1)\alpha^2 N(1+o(1))\right)^t/t! = \sum_{|t-j_0|\geqslant \epsilon j_0 - 2} (j_0(1+o(1))^t/t!,$$

where we have included the -2 to allow for the edges of F. As for W_j, provided $\epsilon \to 0$ sufficiently slowly, this sum is at most $o(e^{j_0} n^{-2})$. Since $S'[F] \lesssim 1$ the contribution to S_H from all atypical graphs F_t is at most $o(e^{j_0} n^{-2})$, which is small compared to e^{j_0}.

Putting the above results together, the total contribution to $e^{-j_0} S_H$ from all atypical graphs F is small. However, from Lemma 4.2 this quantity is a term-by-term bound for $\sum_F \overline{f}(e(F)) P_H(F)^2$, that is, for $\sum_j W_j \mathbb{E}_{e(F)=j} \rho_H(F)^2$, completing the proof of the lemma. □

We have now done most of the work required to prove Theorem 2.2.

Proof of Theorem 2.2. Suppose that the conditions of the theorem are satisfied. We wish to estimate the ratio $\rho_H(J_1) = P_H(J_1)/P_{\alpha N}(J_1)$, where J_1 is the graph consisting of one P_2 and $j-2$ isolated edges, with $j \sim j_0$. Now $P_H(J_1) = X_{J_1}(H)/X_{J_1}(K_n)$, while

$$P_{\alpha N}(J_1) = \binom{N-j}{\alpha N - j}\binom{N}{\alpha N}^{-1} = \binom{\alpha N}{j}\binom{N}{j}^{-1},$$

so $\rho_H(J)$ can be written as $X_{J_1}(H)/\binom{\alpha N}{j}$ divided by $X_{J_1}(K_n)/\binom{N}{j}$, that is, the ratio of the proportion of j-edge subgraphs of H that are isomorphic to J_1 to the corresponding proportion for K_n.

Suppose two distinct edges of H are chosen at random. Then they are adjacent with probability

$$p_1 = X_{P_2}(H)\binom{e(H)}{2}^{-1} = \alpha_2 n \binom{n-1}{2}\binom{\alpha N}{2}^{-1} = \frac{4\alpha_2}{\alpha^2 n}(1+O(n^{-1})).$$

Since by assumption $\alpha_2 = O(\alpha^2)$, we have $p_1 = O(n^{-1})$. Now $j \sim j_0$, which is $o(n^{1/2})$ from (5.9). Thus $p_1 j^2 = o(1)$, and, if j distinct edges are chosen at random from H, the probability that the subgraph formed is isomorphic to J_1, that is, that exactly one pair of

https://doi.org/10.1017/S0963548399004150 Published online by Cambridge University Press

edges is adjacent, is $p_1\binom{j}{2}(1+O(p_1 j^2))$. In other words,

$$X_{J_1}(H)/\binom{\alpha N}{j} = p_1\binom{j}{2}(1+O(p_1 j^2)).$$

Similarly, as two edges of K_n are adjacent with probability $p_2 = n\binom{n-1}{2}\binom{N}{2}^{-1} = 4n^{-1}(1+O(n^{-1}))$, we have

$$X_{J_1}(K_n)/\binom{N}{j} = p_2\binom{j}{2}(1+O(p_2 j^2)).$$

Thus, as $p_1, p_2 = O(n^{-1})$ and $j \sim j_0$,

$$\rho_H(J_1) = \frac{p_1}{p_2}(1+O(j_0^2/n)) = \frac{\alpha_2}{\alpha^2}(1+O(j_0^2/n)).$$

In fact we wish to estimate $\rho_H(J_1)-1$. From the above we have

$$\rho_H(J_1)-1 = \frac{\alpha_2-\alpha^2}{\alpha^2}\left(1+O\left(\frac{\alpha_2}{\alpha_2-\alpha^2}\frac{j_0^2}{n}\right)\right).$$

Now by assumption $\alpha_2-\alpha^2 = \Omega(\Delta n^{-2})$, while $\alpha_2 = O(\alpha^2) = O(\Delta^2 n^{-2})$. Thus

$$\frac{\alpha_2}{\alpha_2-\alpha^2}\frac{j_0^2}{n} = O(\Delta j_0^2 n^{-1}) = O(\Delta p^{-2}\alpha^4 n^3) = O(\Delta^5 n^{-1} p^{-2}),$$

which tends to zero from (2.6). Thus, if $j \sim j_0$ then $\rho_H(J_1)-1 \sim \frac{\alpha_2-\alpha^2}{\alpha^2}$.

The proof is now essentially complete: from (5.8) we may write $f-1 = \mathrm{Var}(X)/\mathbb{E}(X)^2$ as $f-1 = A+B+C$, where

$$\begin{aligned} A &= \sum_{|j-j_0|\leqslant \epsilon j_0} W_j \mathbb{P}_{e(J)=j}(J=J_0)(\rho_H(J_0)-1)^2, \\ B &= \sum_{|j-j_0|\leqslant \epsilon j_0} W_j \mathbb{P}_{e(J)=j}(J=J_1)(\rho_H(J_1)-1)^2, \text{ and} \\ C &= \sum_j W_j \mathbb{E}_{e(J)=j}[1_{\{J \text{ atypical}\}}(\rho_H(J)-1)^2]. \end{aligned}$$

As for a fixed j we have $\mathbb{E}_{e(J)=j}(\rho_H(J)-1) = 0$, and as for typical j almost all j-edge graphs are isomorphic to J_0, we have $A = o(B+C)$: the argument is similar to that in Lemma 5.1. For typical j, $\mathbb{P}_{e(J)=j}(J=J_1) \sim \binom{j_0}{2}p_2 \sim 2j_0^2 n^{-1} \sim (p^{-1}-1)^2\alpha^4 n^3/2$, so

$$B \sim \left(\sum_{|j-j_0|\leqslant \epsilon j_0} W_j\right)\left(\frac{1-p}{p}\right)^2 \frac{\alpha^4 n^3}{2}\left(\frac{\alpha_2-\alpha^2}{\alpha^2}\right)^2 \sim (\alpha_2-\alpha^2)^2(p^{-1}-1)^2 n^3/2, \tag{5.13}$$

as the sum of the W_j over typical j is asymptotically one. Finally, from Lemma 5.3 and (5.10) we have $C = o(\Delta^2(p^{-1}-1)^2 n^{-1})$. From (2.7) we have $\alpha_2-\alpha^2 = \Omega(\Delta n^{-2})$. Together with (5.13) this gives $B = \Omega(\Delta^2(p^{-1}-1)^2 n^{-1})$, so $C = o(B)$.

Putting the above together, $f-1 \sim B \sim (\alpha_2-\alpha^2)^2(p^{-1}-1)^2 n^3/2$, proving (2.9). As shown earlier in this section, together with Lemma 5.1 this completes the proof of Theorem 2.2. □

https://doi.org/10.1017/S0963548399004150 Published online by Cambridge University Press

6. Conclusions

We start by justifying the claim that Theorem 2.1 can be applied to many graphs other than the two considered so far.

Corollary 6.1. *Let H be a fixed sequence of graphs with $n = |H| \to \infty$ and $e(H) \geqslant n$. Suppose that $2 \leqslant \Delta = o(n^{1/4})$, where $\Delta(= \Delta(n))$ is the maximum degree of H. Let*

$$p = \left(\frac{\omega\Delta^4}{n}\right)^{\frac{2(\Delta-1)}{\Delta(\Delta+1)}}, \tag{6.1}$$

where $\omega \to \infty$. Then almost every $G \in \mathcal{G}(n, pN)$ has a spanning subgraph isomorphic to H.

Proof. As before we have $e_H(v) \leqslant \min\{\binom{v}{2}, \frac{\Delta v}{2}\}$. After a little calculation this gives $\gamma(H) \leqslant \frac{\Delta(\Delta+1)}{2(\Delta-1)}$, so (2.2) holds. It is easy to check that (2.1) also holds, when ω tends to infinity sufficiently slowly, so the result follows from Theorem 2.1. □

Note that for $4 \leqslant \Delta = O(n^{1/7})$ the above result improves the bound given by Alon and Füredi [1], which is essentially $(\frac{\Delta^2 \log n}{n})^{1/\Delta}$. In fact, if $\Delta/\log n \to \infty$, we can replace the value for p in Corollary 6.1 by $p = (\omega\Delta^{2+\epsilon}n^{-1})^{\frac{2(\Delta-1)}{\Delta(\Delta+1)}}$ for any fixed $\epsilon > 0$. For this value of p condition (2.2) is not necessarily satisfied, but (2.2′) is, so Theorem 2.1 again applies.

For Δ not too small, Corollary 6.1 is close to best possible, as we shall now show.

Corollary 6.2. *Let H be a fixed sequence of Δ-regular graphs, where $n = |H| \to \infty$ and $c\sqrt{\log n} \leqslant \Delta = \Delta(n) = o(n^{1/4})$ for some constant $c > 0$. Let $p_0 = n^{-2/\Delta}$. Then almost no $G \in \mathcal{G}(n, \lfloor p_0 N \rfloor)$ contains a copy of H, while for some constant $C = C(c)$ almost every $G \in \mathcal{G}(n, \lceil C p_0 N \rceil)$ contains a copy of H.*

Furthermore, if $\Delta/\sqrt{\log n} \to \infty$, we may take for C any constant greater than 1.

Proof. For any H and any p_0 the expected number of copies of H in $G \in \mathcal{G}(n, \lfloor p_0 N \rfloor)$ is at most the expected number in $G_{p_0} \in \mathcal{G}(n, p_0)$, as edges 'correlate negatively' in $\mathcal{G}(n, M)$. In turn, this is at most $n! p_0^{e(H)}$. Under the conditions of the corollary, with $X = X_H(G)$ and $G \in \mathcal{G}(n, \lfloor p_0 N \rfloor)$, this gives $\mathbb{E}(X) \leqslant n!(n^{-2/\Delta})^{\Delta n/2} = n! n^{-n} = o(1)$, so $\mathbb{P}(X > 0) \to 0$. To complete the proof, let $p = C p_0$ where $C > 1$ is constant. Defining ω so that (6.1) holds, a little calculation shows that if C is large enough, or if $\Delta/\sqrt{\log n} \to \infty$, then $\omega \to \infty$. Corollary 6.1 can thus be applied to complete the proof. □

The above result shows that if $\Delta/\sqrt{\log n}$ is bounded below, then p_0 is a threshold function for the property of containing a copy of H, and that if $\Delta/\sqrt{\log n} \to \infty$, then this threshold is sharp. For smaller Δ the bound given by Corollary 6.1 is often not very good, as we have used a very simple estimate of $\gamma(H)$. We can improve this while still keeping a fair amount of generality in the following way.

We say that a graph H on $[n]$ is *d-decomposable* if there is a sequence $H_1 = K_1$, $H_2, \ldots, H_n = H$ of graphs each of which is obtained from the previous one by adding a single new vertex and joining it to at most d existing vertices.

https://doi.org/10.1017/S0963548399004150 Published online by Cambridge University Press

Corollary 6.3. *Let H be a fixed sequence of graphs with $n = |H| \to \infty$, $e(H) \geqslant n$ and and $\Delta = \Delta(H) = o(n^{1/4})$. Suppose that H is d-decomposable, where $d = d(n) = o(n^{1/4})$, and that either $d \geqslant 3$, or $d = 2$ and H is triangle-free. Let $p = (\frac{\omega\Delta^4}{n})^{1/d}$, where $\omega \to \infty$. Then almost every $G \in \mathcal{G}(n, pN)$ has a spanning subgraph isomorphic to H.*

Proof. For $v = 3$ we have in either case that $e_H(v) \leqslant d = d(v-2)$. As H is d-decomposable, $e_H(v+1) \leqslant e_H(v) + d$, so $e_H(v) \leqslant d(v-2)$ for all $v \geqslant 3$, and $\gamma(H) \leqslant d$. Hence (2.2) holds. Again, it is easy to check that (2.1) holds, for ω small enough, so the result follows from Theorem 2.1. □

Note that the lattice L_k is 2-decomposable, so the first part of Theorem 1.4 is just a special case of Corollary 6.3. Also, for $d \geqslant 3$, this result considerably strengthens the corresponding result (Theorem 2) in [5] – we do not need the remaining conditions, we have a (slightly) smaller value of p, and we find a copy of H in $G \in \mathcal{G}(n, pN)$, rather than in $G \in \mathcal{G}(3n, p)$. For example, Corollary 6.3 implies that almost every $G \in \mathcal{G}(n, (\omega n^{-1/3})N)$ has a spanning triangular lattice. On the other hand, we do not obtain such good results for the hexagonal lattice and similar graphs. For these Fernandez de la Vega and Manoussakis [5] give a result with $p = (\log n/n)^{2/3}$, but to apply Theorem 2.1 we need $pn^{1/2} \to \infty$.

For many specific graphs we can do better than the results above, since we can use additional properties of the graph to find better bounds on γ. For example, let $H = [k]^d$ be the *grid*, that is, the graph on $\{1, 2, \dots, k\}^d$ in which two points are adjacent if the Euclidean distance between them is exactly one. The results of Bollobás and Leader [4] give an implicit formula for $e_H(v)$ in this case, which we can use to calculate $\gamma(H)$. This time we omit the calculations and claim that, writing $p_0 = k^{-k/(k-1)}$, a lower bound given by the expectation, we obtain a value of p of the form $p = (1 + o(1))p_0$ if $d \to \infty$, and $p = \omega p_0$ otherwise, where $\omega \to \infty$ arbitrarily slowly.

So far we have presented consequences of Theorem 2.1. Due to the rather messy conditions, it is harder to give simple but general corollaries of Theorem 2.2, although the conditions are easy to verify in specific cases. However, we do have the following result for Δ-regular graphs.

Corollary 6.4. *Let H be a fixed sequence of triangle-free Δ-regular graphs, with $7 \leqslant \Delta = \Delta(n) = o(n^{1/6})$. Suppose that $p = p(n)$ is chosen such that pN is an integer, and for some $\omega \to \infty$*

$$p > \max\left\{(\omega\Delta^4 n^{-1})^{\frac{2(\Delta-1)}{\Delta(\Delta+1)}}, (\omega\Delta^2 n^{-1})^{\frac{2(\Delta-3)}{\Delta(\Delta+1)-8}}\right\}, \tag{6.2}$$

and

$$p < \min\left\{(1 + \omega \log n\Delta^{-2})^{-1}, (1 - \omega\Delta^3 n^{-1/2})\right\}. \tag{6.3}$$

Then for $X = X_H(G)$, $G \in \mathcal{G}(n, pN)$, we have that X^* converges in distribution to a standard normal distribution.

Proof. From $\Delta(H) = \Delta$ we have as before that $e_H(v) \leqslant \min\{\binom{v}{2}, \frac{\Delta v}{2}\}$. After some calculation this implies that $\gamma \leqslant \frac{\Delta(\Delta+1)}{2(\Delta-1)}$ and that $\gamma' \leqslant \frac{\Delta(\Delta+1)-8}{2(\Delta-3)}$. It is now straightforward to verify

https://doi.org/10.1017/S0963548399004150 Published online by Cambridge University Press

all the conditions of Theorem 2.2 apart from (2.6). For this note that $\Delta^6 n^{-1} = o((1-p)^2)$ from (6.3). Thus (2.6) is satisfied if p is bounded away from zero. If $p \to 0$, on the other hand, then from (6.2) we have $\Delta = o(\log n)$, so (2.6) follows from $p((\log n^6)/n)^{-1/2} \to \infty$, which follows from (6.2). □

We finish by considering where the bounds in Theorem 2.1 can be improved. There are two rather different parts to the proofs of Theorems 2.1 and 2.2 – estimating the contribution to (3.2) from *small* F, that is, F consisting of at most one P_2 together with some isolated edges, and estimating the contribution from large F. We consider small F first.

For F empty or $F = P_2$ the key estimate given by Lemma 4.4 is within a constant factor of the truth. However, we cannot deduce that Theorems 2.1 and 2.2 cannot be improved: suppose for example that $H = L_k$, $n = k^2$ and set $p = Cn^{-1/2}$ for C constant, so our results do not apply. Then the estimate $v = (\alpha_2 - \alpha^2)(p^{-1}-1)^2 n^3/2$ for $\mathrm{Var}(X)/\mathbb{E}(X)^2$ no longer tends to zero. One would thus expect that graphs F containing several P_2s would contribute significantly to $\mathrm{Var}(X)/\mathbb{E}(X)^2$, and perhaps that $\mathrm{Var}(X)/\mathbb{E}(X)^2 \sim e^v - 1$. However, the proofs of Theorems 2.1 and 2.2 break down in other places for this value of p. For example, as $\alpha^3 N p^{-2} \not\to 0$, one can check that (4.1) no longer holds. Thus, to estimate the variance in this case one would need either to redo all the calculations to greater accuracy, or, as suggested by Svante Janson, to find a method where the contributions from multiple edges and P_2s are counted 'automatically'. Such a method should avoid the 'coincidental' cancellation of the factors $e^{-\frac{1-p}{p}\alpha^2 N}$ and $e^{\frac{1-p}{p}\alpha^2 N}$ found in the current proof. In summary, it seems unlikely that $\mathrm{Var}(X)/\mathbb{E}(X)^2 \to 0$ when $v \not\to 0$, so Theorem 1.2 probably cannot be strengthened using the second moment method. This leaves open the question of whether Theorem 1.2 is in fact best possible.

Turning to large graphs F, the situation is rather different. We take $H = Q^d$ as an illustration. In this case the estimate in Lemma 4.4 is very weak: we expect that for almost all very large F the quantity $X_F(H)$ will be just the number of automorphisms of H. For such simple graphs as $H = Q^d$ it should be possible to improve the estimate, and to show that the contribution to $\mathrm{Var}(X)/\mathbb{E}(X)^2$ from large F is roughly $\mathbb{E}(X)$, the variance of a Poisson distribution. One could then use the same basic method to prove a result which would really be best possible, that is, to answer the following question, originally due to Bollobás.

Question. Let $n = 2^d$, $N = \binom{n}{2}$ and let $p = p(n)$ be such that the expected number of spanning cubes in $G \in \mathcal{G}(n, pN)$ tends to infinity. Is it true that almost every $G \in \mathcal{G}(n, pN)$ has at least one spanning cube?

Of course, one would also expect even more, namely that the distribution of the number of spanning cubes should be asymptotically Poisson in a suitable range of p. Note, however, that such results will not hold with $\mathcal{G}(n, pN)$ replaced by $\mathcal{G}(n, p)$. This is because there are values of p for which the expectation in $\mathcal{G}(n, p)$ tends to infinity, while that in $\mathcal{G}(n, pN)$ tends to zero. About half the $G \in \mathcal{G}(n, p)$ have at most pN edges, and are thus unlikely to have spanning cubes.

References

[1] Alon, N. and Füredi, Z. (1992) Spanning subgraphs of random graphs: Research problems. *Graphs and Combin.* **8** 91–94.

[2] Bollobás, B. (1985) *Random Graphs*, Academic Press, London.

[3] Bollobás, B. and Leader, I. (1990) Exact face-isoperimetric inequalities. *Europ. J. Combinatorics* **11** 335–340.

[4] Bollobás, B. and Leader, I. (1991) Edge-isoperimetric inequalities in the grid. *Combinatorica* **11** 299–314.

[5] Fernandez de la Vega, W. and Manoussakis, Y. (1994) Grids in random graphs. *Random Structures and Algorithms* **5** 329–336.

[6] Janson, S. (1994) The numbers of spanning trees, Hamilton cycles and perfect matchings in a random graph. *Combinatorics, Probability and Computing* **3** 97–126.

[7] Moon, J. W. (1972) The variance of the number of spanning cycles in a random graph. *Studia Sci. Math. Hungar.* **7** 281–283.

[8] Venkatesan, R. and Levin, L. (1988) Random instances of a graph coloring problem are hard. In *Proc. 20th Annual Symposium on the Theory of Computing*, ACM Press, pp. 217–222.