Preprint in the style of *Combinatorics, Probability and Computing*

Spanning subgraphs of random graphs

OLIVER RIORDAN

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

Department of Pure Mathematics and Mathematical Statistics
16 Mill Lane, Cambridge CB2 1SB, England

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 result which can be applied to many other graphs, improving also previous results for the lattice, i.e., 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 $[n] = \{1,2,\dots,n\}$ 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 which p does $G \in \mathcal{G}(2^d,p)$ 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. They also noted that for $p = \frac{1}{4}$ the expected number of d-cubes in G

tends to zero, so $p > \frac{1}{4}$ is definitely necessary. We shall prove the following result, which is thus close to best possible.

Theorem 1. *Let $p = p(d) = \frac{1}{4} + 6\log d/d$. Then with probability tending to 1 as $d \to \infty$ a random $G \in \mathcal{G}(2^d, p)$ 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. We shall show that the lower bound is again essentially correct, proving the following result.

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

Note that, together with the lower bound given by the expectation, Theorem 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]).

Theorem 2 also strengthens a result of Fernandez de la Vega and Manoussakis [5] concerning lattices L_k in $G \in \mathcal{G}(n, p)$ where k^2 is some constant fraction of n less than one, which we return to later.

We shall deduce Theorems 1 and 2 from a common generalisation 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 $\mathcal{G}(n, M)$ rather than $\mathcal{G}(n, p)$. We shall thus derive equivalents of Theorems 1 and 2 for this model, namely the following. As before we write N for $\binom{n}{2}$.

Theorem 3. *Let $p' = p'(d) = \frac{1}{4} + 5\log d/d$, and let $n = 2^d$. Then with probability tending to 1 as $d \to \infty$ a random $G \in \mathcal{G}(n, p'N)$ contains a copy of Q^d.*

Theorem 4. *Let $p' = p'(n) = \omega'(n)n^{-1/2}$, where $\omega'(n) \to \infty$ arbitrarily slowly, and let $n = k^2$. Then with probability tending to 1 as $k \to \infty$ a random $G \in \mathcal{G}(n, p'N)$ contains a copy of L_k.*

Note that these results immediately imply Theorems 1 and 2, as a random $G \in \mathcal{G}(n, p)$ almost certainly has at least $p'N$ edges, whenever $N(p - p')/\sqrt{Np(1-p)} \to \infty$. In the case of Theorem 1, this condition is satisfied by the given p and p'; for Theorem 2 we just take $\omega'(n) = \omega(n)/2$, for example.

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. Here, although we consider much more complicated graphs H, we do not find the whole distribution, and our calculations are much simpler. A simple example of the same expansion occurs in Moon [7], who uses it to find the variance of the number of Hamilton cycles in $G \in \mathcal{G}(n, \frac{1}{2})$.

2. Statement of the main result

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 $f \leq (1 + o(1))g$, i.e., if $\limsup f/g \leq 1$.

To state our result 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

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

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 the maximum is at $v = 3$ or $v = 4$.

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

Theorem 5. *Let $H (= H^{(i)})$ be a fixed sequence of graphs with $n = |H| \to \infty$. Let $e(H) = \alpha N$, and let $p = p(n) \in (0, 1)$. Suppose that the following conditions hold:*

$$\alpha N \geq n, \text{ and } (1 - 2\alpha)N,\ pN,\ (1 - p)\sqrt{n} \to \infty, \tag{1}$$

and that

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

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) is slightly weaker. When Δ is fairly small, there is very little difference. Since Theorem 5 is potentially applicable with Δ nearly as large as $n^{1/4}$, however, we note that we can instead use the conditions

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

where $\epsilon > 0$ is arbitrarily small.

Remark. The conditions in (1) are just technicalities, and can probably be weakened or dropped. However, they all hold in the range where Theorem 5 is likely to give good bounds. Also, we *do* make essential use of two consequences of the above conditions, namely

$$np^2/\Delta^4 \to \infty \tag{3}$$

and

$$\alpha^3 N p^{-2} \to 0. \tag{4}$$

Clearly, if (2') holds, then so does (3). Also, $\alpha N \geq n$ implies that $\Delta \geq 2$, and hence that $\gamma \geq 2$. Together with (2), this implies (3). Finally, since $\alpha \lesssim \Delta/n$, condition (3) implies (4). Thus (3) and (4) hold under the conditions of the theorem, whether we take (2) or (2').

We shall prove Theorem 5 in the next section. For the moment, we assume the result, and deduce Theorems 3 and 4, and hence Theorems 1 and 2.

Proof of Theorem 3. Let $H = Q^d$, $p = \frac{1}{4} + 5\log d/d$, and $n = 2^d$. As above we write $N = \binom{n}{2}$ and $e(H) = \alpha N$, so $\alpha = d/(n-1)$. It is easy to see that (1) holds, so to prove Theorem 3 we only need to verify (2). For this we use the fact from [3] that $e_H(v) \leq v \log_4 v$, 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 $\gamma = \gamma(H) = e(H)/(n-2) = dn/2(n-2)$. To check (2) we note that $1 + x \geq e^{x/2}$ for x sufficiently small. Thus, for d large enough,

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

As $n^{2/(n-2)} = o(d)$ we thus have $p^{-\gamma} = o(nd^{-4})$. As $\Delta = d$, we deduce that (2) holds, completing the proof. $\square$

Proof of Theorem 4. This time we take $H = L_k$, $n = k^2$ and $p = \omega(n)n^{-1/2}$, with $\omega \to \infty$. As $e(H) = 2n(1 - n^{-1/2})$ we have $\alpha \sim 4n^{-1}$, and it is easy to see that (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, $s \leq k$, 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 $1 \leq k \leq s$ 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) we use the observation that $e_H(v + 1) \leq e_H(v) + 2$, which follows from these formulae, or directly from the nature of the ordering. Thus, since $e_H(v)/(v - 2) = 2$ at $v = 3$, we have $\gamma(H) = 2$, and (2) holds. $\square$

In the next section we turn to the proof of Theorem 5.

3. Proof of the main result.

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.

Given a graph H on n vertices, let $H_1, \ldots, H_h$ be all copies of H with vertex set $[n]$. Thus h is just $n!$ divided by the number of automorphisms of H. For G a random graph, chosen from either $\mathcal{G}(n,p)$ or $\mathcal{G}(n,M)$, let X_i be the indicator function of the event that $H_i \subset G$. Let $X = \sum_{i=1}^h X_i$, so that, as before, X is the number of subgraphs of G isomorphic to H.

Lemma 6. *With the notation above, let* $f_{ij} = \mathbb{P}(X_i = 1 \mid X_j = 1)/\mathbb{P}(X_i = 1)$*, and let* $f = \frac{1}{h}\sum_{i=1}^h f_{i1}$*. Then*

$$\mathrm{Var}(X) = \mathbb{E}(X)^2(f-1),$$

and, if $f \lesssim 1$ *as* $n \to \infty$*,*

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

Proof. Let $\mu = \mathbb{E}(X)$. Since all the H_i are equivalent, we have

$$\begin{aligned} \mathbb{E}(X^2) &= \mathbb{E}(\sum_{i=1}^h X_i X) = \sum_{i=1}^h \mathbb{P}(X_i = 1)\mathbb{E}(X \mid X_i = 1) \\ &= h\mathbb{P}(X_1 = 1)\mathbb{E}(X \mid X_1 = 1) = \mu\mathbb{E}(X \mid X_1 = 1). \end{aligned}$$

Also,

$$\mathbb{E}(X \mid X_1 = 1) = \sum_{i=1}^h \mathbb{P}(X_i = 1 \mid X_1 = 1) = \sum_{i=1}^h f_{i1}\mathbb{P}(X_i = 1) = \frac{\mu}{h}\sum_{i=1}^h f_{i1} = \mu f,$$

so

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mu^2 f - \mu^2,$$

proving the first part of the lemma. For the second part, note that $f \geq 1$, as $\mathrm{Var}(X) \geq 0$, so $f \to 1$. Since $(X-\mu)^2$ has expectation $\mathrm{Var}(X)$, we have from Markov's inequality that $\mathbb{P}(X = 0) \leq \mathbb{P}((X-\mu)^2 \geq \mu^2) \leq \mathrm{Var}(X)/\mu^2 = f - 1$, which tends to zero. □

Remark. For the case when H is the d-dimensional cube Q^d, for example, we cannot hope to get reasonable bounds by applying Lemma 6 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_{i1} in this model, starting with a routine estimate of the probability that any given set of edges is present in G.

Lemma 7. *Let G be a random graph from $\mathcal{G}(n,pN)$, where $N = \binom{n}{2}$. Let $A(= A^{(n)})$ be a fixed graph (sequence of graphs) on $[n]$ with αN edges, and suppose that αN, $(1-\alpha)N$, $pN \to \infty$. Suppose also that $\alpha^3 N p^{-2} \to 0$. Then*

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

as $n \to \infty$.

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

$$\begin{aligned} \mathbb{P}(A \subset G) &= \binom{N-\alpha N}{pN-\alpha N}\binom{N}{pN}^{-1} = \frac{(N-\alpha N)!(pN)!}{(pN-\alpha N)!N!} \\ &\sim \sqrt{\frac{(1-\alpha)p}{p-\alpha}}\left[(1-\alpha)^{1-\alpha}p^p(p-\alpha)^{-(p-\alpha)}\right]^N. \end{aligned}$$

Since $\alpha^2 p^{-2} = \alpha^3 N p^{-2}/(\alpha N) \to 0$, we have $\alpha = o(p)$, so the factor under the square root is $1 + o(1)$. The lemma follows by expanding the logarithm of the factor in square brackets, neglecting terms of order $\alpha^3 p^{-2}$ or smaller. □

The above lemma immediately gives us the following asymptotic upper bound for f_{i1}.

Corollary 8. *Suppose that conditions (1) and (4) of the previous section are satisfied, and let H_i, H_j be two copies of H on $[n]$. Then for G a random graph from $\mathcal{G}(n,pN)$ we have*

$$f_{ij} = \mathbb{P}(H_i \subset G \mid H_j \subset G)/\mathbb{P}(H_i \subset G) \lesssim e^{-\frac{1-p}{p}\alpha^2 N} c^{e(H_i \cap H_j)},$$

where $c = p^{-1}e^{2\alpha\frac{1-p}{p}}$.

Proof. Let $e(H_i \cap H_j) = \beta N$, so $e(H_i) = e(H_j) = \alpha N$ and $e(H_i \cup H_j) = (2\alpha - \beta)N$. Since H_i, H_j and $H_i \cup H_j$ all satisfy the conditions of Lemma 7 we have

$$\begin{aligned} f_{ij} &= \mathbb{P}(H_i \cup H_j \subset G)\mathbb{P}(H_i \subset G)^{-1}\mathbb{P}(H_j \subset G)^{-1} \\ &\sim p^{(2\alpha-\beta)N} e^{-\frac{1-p}{2p}(2\alpha-\beta)^2 N} p^{-2\alpha N} e^{\frac{1-p}{p}\alpha^2 N} \\ &= p^{-\beta N} e^{-\frac{1-p}{2p}(2\alpha^2 - 4\alpha\beta + \beta^2)N} \\ &\leq e^{-\frac{1-p}{p}\alpha^2 N}\left(p^{-1}e^{2\alpha\frac{1-p}{p}}\right)^{\beta N}, \end{aligned}$$

as required. □

We now wish to express f as a sum over the subgraphs of H, rather than over the copies of H on $[n]$. It will be convenient to think of H_i as chosen uniformly at random from $H_1, \dots, H_h$, and we shall write $\mathbb{P}_i$, $\mathbb{E}_i$ for the probability and expectation in this probability space. In this notation we have that $f = \mathbb{E}_i f_{i1}$, by definition. Thus, from Corollary 8,

$$f \lesssim e^{-\frac{1-p}{p}\alpha^2 N}\mathbb{E}_i\left(c^{e(H_1\cap H_i)}\right), \tag{5}$$

where $c = p^{-1}e^{2\alpha\frac{1-p}{p}}$. The next lemma gives an explicit formula for the expectation above, in terms of the numbers of subgraphs of H of each type. As before, we write $X_F(H)$ for the number of subgraphs of H isomorphic to F. Also, from this point on all our graphs will have vertex set $[n]$, unless stated otherwise. Thus $F \subset H$ will mean that F is a spanning subgraph of H.

Lemma 9. *For any graph H on $[n]$ and any $c \in \mathbb{R}$ we have*

$$\mathbb{E}_i\left(c^{e(H_1\cap H_i)}\right) = \sum_{F\subset H}(c-1)^{e(F)}\frac{X_F(H)}{X_F(K_n)}, \tag{6}$$

where the sum is over all $2^{e(H)}$ spanning subgraphs of H.

Proof. By definition,

$$\mathbb{E}_i\left(c^{e(H_1\cap H_i)}\right) = \sum_{E\subset H_1} c^{e(E)}\mathbb{P}_i(H_1\cap H_i = E).$$

Now the probability that H_1 and H_i have exactly some fixed intersection E is much harder to calculate than the probability that the intersection of H_1 and H_i contains some fixed graph F. However, as the number of $F \subset E$ with r edges is just $\binom{e(E)}{r}$, and as $c^{e(E)} = (1+(c-1))^{e(E)}$, we have

$$\begin{aligned}
\mathbb{E}_i\left(c^{e(H_1\cap H_i)}\right) &= \sum_{E\subset H_1}\sum_{F\subset E}(c-1)^{e(F)}\mathbb{P}_i(H_1\cap H_i = E)\\
&= \sum_{F\subset H_1}(c-1)^{e(F)}\mathbb{P}_i(H_1\cap H_i \supset F)\\
&= \sum_{F\subset H_1}(c-1)^{e(F)}\mathbb{P}_i(F\subset H_i).
\end{aligned}$$

Finally, since each isomorphic copy of F on $[n]$ is equally likely as a subgraph of H_i, we have $\mathbb{P}_i(F \subset H_i) = X_F(H)/X_F(K_n)$. As it makes no difference whether we sum over $F \subset H_1$ or $F \subset H$, this completes the proof of the lemma. □

Together (5) and (6) give us a bound for f involving the numbers $X_F(H)$, $X_F(K_n)$ and $c = p^{-1}e^{2\alpha\frac{1-p}{p}}$. We would like to replace c by p^{-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 (5) and (6) by straightforward calculation.

Lemma 10. *Suppose that conditions (1), (3) and (4) of the previous section are satisfied. Then*

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

Proof. Writing $c = p^{-1}e^{2\alpha\frac{1-p}{p}}$ we have

$$\log\left(\frac{c-1}{p^{-1}-1}\right) = \log\left(\frac{pc-p}{1-p}\right) = \log\left(1 + \frac{e^{2\alpha\frac{1-p}{p}}-1}{1-p}\right) \le \frac{e^{2\alpha\frac{1-p}{p}}-1}{1-p}.$$

Under the given conditions we have $\alpha = o(p)$, so $e^{2\alpha\frac{1-p}{p}} - 1 \sim 2\alpha\frac{1-p}{p}$, and $\log(\frac{c-1}{p^{-1}-1}) \lesssim 2\alpha p^{-1}$. Now as $\alpha \le \Delta/(n-1)$ 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,

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

We thus have from (6) that $\mathbb{E}_i(c^{e(H_1\cap H_i)}) \le S_H$. Combined with (5) this completes the proof of the lemma. □

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 11. *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,

$$S_H = \sum_{[F]:F\subset H} (p^{-1}-1)^{e(F)}(1+n^{-\frac{1}{2}})^{r(F)} \frac{X_F(H)^2}{X_F(K_n)}, \tag{7}$$

and

$$S'_H = \sum_{[F]:F\subset H}{}' (p^{-1}-1)^{e(F)} 2^{r(F)} \frac{X_F(H)^2}{X_F(K_n)}, \tag{8}$$

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 $t \le \frac{v}{2}$ new edges, isolated in F_t. We shall write $S[F,t]$ for the contribution of $[F_t]$ to (7), and $S[F]$ for $\sum_{t\le v/2} S[F,t]$. We shall also write $S'[F]$ for the contribution of $[F]$ to (8). 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

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

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

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

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

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

Summing over t we obtain, for n large enough,

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

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 $e_H(w) \le \min\{\binom{w}{2}, \frac{w\Delta}{2}\}$, and hence that $\beta_w \le \min\{\binom{w}{2}, (\frac{w\Delta}{2})^2\binom{w}{2}^{-1}\}$. As the first term increases and the second decreases, and they are equal when $w = \Delta+1$, we have $\beta_w \le \binom{\Delta+1}{2} \le \Delta^2$ for all w. Hence $p^{-1}\beta_w \le p^{-1}\Delta^2 = o(\sqrt{n})$, and the error in truncating the sum above at $\sqrt{n}$ will tend to zero. Thus

$$\frac{S[F]}{S'[F]} \lesssim 1.5^{-r(F)} \sum_{t=0}^{\sqrt{n}} (p^{-1}-1)^t(1+n^{-\frac{1}{2}})^t \frac{\beta_v\beta_{v-2}\cdots\beta_{v-2t+2}}{t!}. \tag{9}$$

We now consider two cases, according to how close v is to n, noting that we expect the $v=n$ term of S'_H to dominate.

Suppose first that $v \le n-\sqrt{n}$. Using $\beta_w \le \Delta^2$ we have from (9) that

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

since $r(F) \ge \frac{1}{2}(n-v) \ge \frac{1}{2}\sqrt{n}$ and $\Delta^2 p^{-1} = o(\sqrt{n})$.

Suppose instead that $v \ge n-\sqrt{n}$. This time we use the bound $\beta_w \le e(H)^2/\binom{w}{2} = \alpha^2 N\binom{n}{2}/\binom{w}{2}$, which gives $\beta_w = \alpha^2 N(1+O(n^{-1/2}))$ for $w \ge n-3\sqrt{n}$. Ignoring, as we may, the factor $1.5^{-r(F)}$, we thus have from (9) that

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

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

Combining the two cases above we have that $S[F]/S'[F] \lesssim e^{\alpha^2 N\frac{1-p}{p}}$ for any F. This implies that $S_H/S'_H \lesssim e^{\alpha^2 N\frac{1-p}{p}}$, 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 12. *Let H be any graph on $[n]$ with maximum degree Δ, and $F \subset H$. Then*

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

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 non-trivial component C_i of F. Then go through the components C_i listing their remaining vertices such that each is preceeded 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. For v 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

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

Now for any $0 \le b \le a$ we have $(a)_b \ge (a/e)^b$—if this assertion fails anywhere it is at

$b = a$, where it follows from Stirling's formula. Since $k' \leq n/2$ we have $(n - k')_{r(F)} \geq (n/2e)^{r(F)}$, 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 10, 11 and 12 that, under the conditions of Theorem 5,

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

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 non-trivial 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

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

i.e.,

$$T'_H \leq e^{T''_H}, \tag{10}$$

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 5 we now bound T''_H directly.

Lemma 13. *For every graph H on $[n]$ we have*

$$T''_H \leq 50\Delta^2 \sum_{s=3}^{n} \left(\frac{50\Delta^2}{n}\right)^{s-2} p^{-e_H(s)}, \tag{11}$$

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 non-trivial 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

$$T''_H \le \sum_{s=3}^{n} \left(\frac{15\Delta}{n}\right)^{s-1} \sum_V \sum_{m=0}^{e_H(s)} \binom{e_H(s)}{m} (p^{-1}-1)^m$$
$$= \sum_{s=3}^{n} \left(\frac{15\Delta}{n}\right)^{s-1} \sum_V p^{-e_H(s)},$$

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 $\mathcal{T}(s)$ be the set of trees with vertex set $[s]$, so $|\mathcal{T}(s)| = s^{s-2}$ by Cayley's formula. Let $\Phi(s)$ be the set of pairs (T, ϕ) with $T \in \mathcal{T}(s)$ and ϕ a 1-1 edge preserving map from T to H. As H has maximum degree Δ, each $T \in \mathcal{T}(s)$ can be mapped into H in at most $n\Delta^{s-1}$ ways, so $|\Phi(s)| \le n\Delta^{s-1}s^{s-2}$. 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 $1, 2, \ldots, s$. 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 $s! \ge (s/e)^s$, this quantity is at most $nes^{-2}(e\Delta)^{s-1} < n(e\Delta)^{s-1}$, so we have

$$T''_H \le n \sum_{s=3}^{n} \left(\frac{15e\Delta^2}{n}\right)^{s-1} p^{-e_H(s)}.$$

Taking one factor of $15e\Delta^2/n$ out of the sum, and noting that $15e < 50$, we obtain (11), completing the proof. □

We have now done all the work required to prove Theorem 5. All that remains is some straightforward calculation.

Proof of Theorem 5. We assume, as we may, that (1) holds, and also that either (2) or (2') holds. As noted in the previous section, conditions (3) and (4) are then also satisfied. We consider the quantity f defined in Lemma 6. By Lemma 6, it suffices to show that $f \lesssim 1$. Also, by Lemmas 10, 11 and 12 and from (10) we have

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

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

From Lemma 13 and the definition of γ we have

$$T''_H \le 50\Delta^2 \sum_{s=3}^{\infty} \left(\frac{50\Delta^2 p^{-\gamma}}{n}\right)^{s-2} \le \sum_{s=3}^{\infty} \left(\frac{10^4\Delta^4 p^{-\gamma}}{n}\right)^{s-2}. \qquad (12)$$

We now consider two cases, according to which of (2) and (2') holds.

If (2) holds then $\Delta^4 p^{-\gamma} n^{-1} \to 0$, so from (12) we have $T''_H \to 0$ as required.

If (2') holds the argument is slightly more complicated. This time we have $p^{-1} \sim 1$. Thus the first $O(1)$ terms in (11) 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 $\Delta^2(50\Delta^2 p^{-\gamma} n^{-1})^{s-2} \leq (\Delta^{2+\epsilon} p^{-\gamma} n^{-1})^{s-2} = (o(1))^{s-2}$, so the sum of the remaining terms in (11) tends to zero. Thus $T''_H \to 0$ in this case also, completing the proof of Theorem 5. □

4. Conclusions

We start by justifying the claim that Theorem 5 can be applied to many graphs other than the two considered so far.

Corollary 14. *Let H be a fixed sequence of graphs with $n = |H| \to \infty$ and $e(H) \geq n$. Suppose that $2 \leq \Delta = o(n^{1/4})$, where $\Delta(= \Delta(n))$ is the maximum degree of H. Let*

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

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 $e_H(v) \leq \min\{\binom{v}{2}, \frac{\Delta v}{2}\}$. After a little calculation this gives $\gamma(H) \leq \frac{\Delta(\Delta+1)}{2(\Delta-1)}$, so (2) holds. It is easy to check that (1) also holds, when ω tends to infinity sufficiently slowly, so the result follows from Theorem 5. □

Note that for $4 \leq \Delta = O(n^{1/7})$ 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, for Δ much bigger than $\log n$ we are better off using the version of Theorem 5 where condition (2) is replaced by (2'), but we omit the details. Note also that when H is Δ-regular we need $p > n^{-2/\Delta}$ for the expected number of copies of H in G not to tend to zero. Thus, for Δ reasonably large, Corollary 14 is not far from best possible. On the other hand, for small Δ the bound given by Corollary 14 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.

Corollary 15. *Let H be a fixed sequence of graphs with $n = |H| \to \infty$, $e(H) \geq n$ and and $\Delta = \Delta(H) = o(n^{1/4})$. Suppose that H is d-decomposable and that either $d \geq 3$, 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 $e_H(v) \leq d = d(v-2)$. As H is d-decomposable, $e_H(v+1) \leq e_H(v) + d$, so $e_H(v) \leq d(v-2)$ for all $v \geq 3$, and $\gamma(H) \leq d$. Hence (2) holds. Again, it is easy to check that (1) holds, for ω small enough, so the result follows from Theorem 5. □

Note that the lattice L_k is 2-decomposable, so Theorem 4 is just a special case of Corollary 15. Also, for $d \geq 3$, 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 15 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 5 we need $pn^{1/2} \to \infty$. As we shall see below, this is an essential limitation of the method.

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*, i.e., the graph on $\{1, 2, \ldots, k\}^d$ 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.

We finish by considering where the bounds in Theorem 5 can be improved. In estimating the variance of X, the number of subgraphs of G isomorphic to H, we lose essentially nothing up to (5). In what follows, our estimates are within a constant factor when F is small, and in particular for $F = P_2$, a path with two edges. Thus the condition $np^2/\Delta^4 \to \infty$ is necessary for $\mathrm{Var}(X)/(\mathbb{E}(X)^2)$ to tend to zero. On the other hand, for almost all large F we expect $X_F(H)$ to be just the number of automorphisms of H. This will usually be much smaller than the bound $(\frac{\Delta}{n})^{r(F)}$ we used. For such simple graphs as the d-cube, it should be possible to use the same basic method to prove a result which would really be best possible, i.e., 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] N. Alon and Z. Füredi, Spanning subgraphs of random graphs, Research problems, *Graphs and Combin.* **8** (1992), 91–94.

[2] B. Bollobás, *Random Graphs,* Academic Press, London 1985, xvi+447pp.

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

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

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

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

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

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