How Many Graphs Are Unions Of k-Cliques?

Béla Bollobás[*]
Department of Mathematical Sciences
University of Memphis
Memphis TN 38152-6429, U.S.A.
and
Trinity College
Cambridge CB2 1TQ
U.K.

Graham R. Brightwell [†]
Department of Mathematics
London School of Economics
Houghton St.
London WC2A 2AE
U.K.

7 April 2003

Abstract

We study the number $F[n;k]$ of n-vertex graphs that can be written as the edge-union of k-vertex cliques. We obtain reasonably tight estimates for $F[n;k]$ in the cases (i) $k = n - o(n)$ and (ii) $k = o(n)$ but $k/\log n \to \infty$. We also show that $F[n;k]$ exhibits a phase transition around $k = \log_2 n$. We leave open several potentially interesting cases, and raise some other questions of a similar nature.

1 Introduction

An $[n;k]$-graph is a graph $G = (V, E)$ with vertex set $V = [n] = \{1, 2, \ldots, n\}$ such that E is the union of the edge-sets of copies of K_k, the complete graph on k vertices. Equivalently, an $[n;k]$-graph is a graph on $[n]$ such that every edge lies in some complete graph with at least k vertices. Note that, in this latter formulation, we need not assume that k is an integer. Clearly, a graph is an $[n;k]$-graph if it is an $[n;\lceil k \rceil]$-graph. We are interested in the number $F[n;k]$ of $[n;k]$-graphs. Putting it another way, we wish to estimate the probability that a random graph $G = G(n, 1/2)$ is such that every edge is in a k-clique.

The study of $F[n;k]$ seems a natural problem in its own right. It is also motivated by the work in [3] and [2], which deal with an analogous problem for the cube: how many subsets

[*] email: bollobas@msci.memphis.edu; partially supported by NSF grant DMS-9971788 and DARPA grant F33615-01-C-1900

[†] email: g.r.brightwell@lse.ac.uk; some of this research was carried out while this author was visiting the Isaac Newton Institute for Mathematical Sciences, Cambridge, U.K.

of the n-cube are unions of k-cubes? (This question has another natural equivalent version: how many Boolean functions on n variables can be expressed using a k-SAT formula?) The topic also suggests many other problems of a similar nature; we discuss this at the end of the paper.

The first observation we make is that there is some form of phase transition around $k = 2\lg n$, since below that level almost every graph is an $[n;k]$-graph, whereas above that level a random graph almost surely has no k-cliques. (Here and throughout the paper, lg denotes the binary logarithm.) To be more precise, given $k = k(n)$, define C_n by $k = 2\lg n - 2\lg\lg n + C_n$. If $C_n \geq C$, for some constant $C > 2\lg e - 1$, then a.a.s. (asymptotically almost surely) G contains no k-clique, while if $C_n \leq C'$, for some constant $C' < 2\lg e - 1$, then G a.a.s. contains a k-clique. See Section 11.1 of [1] for more details. The threshold for *every* edge of the random graph to be in a k-clique is effectively the same as that for a *specified* edge to lie in a k-clique. This threshold is lower than the first threshold by exactly 2: if $C_n \geq C$, for some constant $C > 2\lg e - 3$, then a.a.s. a fixed edge xy will not lie in a k-clique, while if $C_n \leq C'$, for some constant $C' < 2\lg e - 3$, then a.a.s. every edge is in a k-clique.

For us it is the last statement that gives us firm information, telling us that $F[n;k] = 2^{\binom{n}{2}}(1 - o(1))$ whenever $k \leq 2\lg n - 2\lg\lg n + 2\lg e - 3 - \varepsilon$ for any fixed $\varepsilon > 0$. One of our main aims in this paper is to exhibit a phase transition for our property: we shall show that, as soon as we have $k \geq 2\lg n - 2\lg\lg n + 2\lg e - 1 + \varepsilon$ ($\varepsilon > 0$), i.e., as soon as k increases beyond the threshold for the existence of a k-clique, $F[n;k]$ is already less than $2^{\binom{n}{2} - 24\lg^2 n(1+o(1))}$. We do not know what happens for the (typically two) values of k between the two thresholds. We discuss this range near the 'threshold' in Section 4.

If k is a little larger, so that $k/\log n \to \infty$ while $k = o(n)$, we are able to pin down the behaviour of $F[n;k]$ fairly precisely. Here we get a lower bound on $F[n;k]$ by considering graphs containing a fixed clique L on slightly more than $4k$ vertices; for most such graphs, every pair of vertices has at least $k - 2$ common neighbours in L, and so the graph is an $[n;k]$-graph. This shows that $F[n;k] \geq 2^{\binom{n}{2} - 8k^2 + o(k^2)}$; we prove an upper bound of the same form in Section 3.

At the other end of the spectrum, we have some tight bounds when $k = n - r$ and $r = o(n)$; here we have $F[n; n-r] \simeq n^{r^2/4}$. To see a lower bound of this form, take a clique T of size $\lceil r/2 \rceil$, and join each element x of T to some set S_x of $\lceil (r+1)/2 \rceil$ other vertices: the complement of such a graph is an $[n; n-r]$-graph, as it is the union of the cliques with vertex sets $V(G) \setminus (S_x \cup T \setminus \{x\})$, for each $x \in T$. We prove a matching upper bound in Section 5, along with some more precise results in the case where r is constant.

One range of interest we leave open is when $k = cn$, for some constant $c \in (0,1)$. Our belief is that there are two regimes: for c below some threshold value, most $[n;k]$-graphs resemble those constructed for $k = o(n)$ (every edge is in some k-clique with all the remaining vertices lying in one particular very large clique L), while above that threshold most $[n;k]$-graphs resemble those constructed for $k = n - o(n)$ (the vertex set can be partitioned into a clique L and an independent set T, with every vertex of T having at least $k - 1$ neighbours in L). We discuss this in a little more detail in Section 6.

2 Very small k

For values of k below the threshold, it is still reasonable to ask about the probability that a random graph $G_{n,1/2}$ is *not* an $[n;k]$-graph. For instance, it is straightforward to obtain sharp estimates of this probability when k is constant, and we discuss this briefly in this section, although this not a major concern for us.

For $k = 1, 2$, all graphs are $[n;k]$-graphs, so $F[n;k] = 2^{\binom{n}{2}}$. The first non-trivial case is $k = 3$.

For a fixed pair $\{x, y\}$ of vertices, the probability that xy is an edge of the random graph $G(n, \frac{1}{2})$ not in a triangle is equal to $\frac{1}{2}\left(\frac{3}{4}\right)^{n-2}$. The probability that two disjoint pairs $\{x, y\}$ and $\{u, v\}$ both form "bad" edges is at most $\left(\frac{3}{4}\right)^{2(n-4)}$, and the probability that $\{x, y\}$ and $\{y, z\}$ are both bad is at most $\left(\frac{5}{8}\right)^{n-3}$. Thus the probability of at least two bad edges is $O(n^3\left(\frac{5}{8}\right)^n)$, which is much smaller than the probability that any particular edge is bad. Therefore the probability that there is a bad edge is

$$\frac{1}{2}\binom{n}{2}\left(\frac{3}{4}\right)^{n-2} + O\left(n^3\left(\frac{5}{8}\right)^n\right).$$

For $k = 4$, the probability that xy is not in a K_4 is given by

$$\frac{1}{2}\sum_{j=0}^{n-2}\binom{n-2}{j}\left(\frac{1}{4}\right)^j\left(\frac{3}{4}\right)^{n-2-j}2^{-\binom{j}{2}},$$

since the j-term in the sum is the probability that the set of common neighbours of x and y is an independent set of size j. To a reasonable order of accuracy, the sum is equal to its largest term, which is a term with $|j - (\lg n - \lg\lg n)| = O(1)$, and the sum is

$$\left(\frac{3}{4}\right)^n 2^{\frac{1}{2}\lg^2 n - \lg n \lg\lg n + O(\log n)}.$$

As in the $k = 3$ case, the probability that there are two edges not lying in a K_4 is very roughly $\left(\frac{5}{8}\right)^n$, so the last expression is also the form of the probability that some edge does not lie in a K_4.

This calculation is very reminiscent of one from [2], and in fact one can continue for larger constant values of k in the same manner as in that paper, but shorn of the difficulties.

3 Quite Small k

Our results in this section cover the case when $k = o(n)$, but $k/\log n \to \infty$. Our aim is to prove that, in this range, $F[n;k] = 2^{\binom{n}{2} - 8k^2 + o(k^2)}$, or in other words that the probability that a random graph on $[n]$ is an $[n;k]$-graph is $2^{-8k^2+o(k^2)}$. The proofs we give here are also applicable for $k = C\lg n$, $C > 2$, but it would take some extra work to extract the best possible bounds and we choose not to go into any great detail for this range. In the next section, we shall develop the techniques further to deal with the threshold range.

We start with the lower bound. For a subset L of $[n]$, let $\mathcal{G}[L]$ be the set of all graphs on vertex set $[n]$ in which L is a clique, and make this into a probability space by making all graphs in $\mathcal{G}[L]$ equally likely. A random graph in $\mathcal{G}[L]$ can be constructed by taking each pair of vertices as an edge with probability $1/2$, each choice made independently, except that all pairs of vertices from L are automatically taken as edges.

For the deviation from the mean of a Binomial random variable, we shall make use of the following estimates, which are often referred to as the *Chernoff bounds.* See, for instance, [1].

Theorem 3.1 *Let X be a Binomial random variable with parameters (n,p), and set $\lambda = np = \mathbb{E}X$. Then,*

$$\begin{array}{rcll} \mathbb{P}(X \le \lambda - t) & \le & \exp(-t^2/2\lambda) & \text{for any } t; \\ \mathbb{P}(X \ge \lambda + t) & \le & \exp(-t^2/3\lambda) & \text{for } t \le \lambda. \end{array}$$

Lemma 3.2 *Suppose $k \ge \log n$. Let $\ell = \lceil 4k + 36\sqrt{k \log n} \rceil$, and set $L = [\ell] = \{1, 2, \ldots, \ell\}$. Asymptotically almost surely, every edge in a random graph G from $\mathcal{G}[L]$ is in a k-clique.*

Proof. We claim that, asymptotically almost surely, every pair of vertices in $[n] \setminus L$ has at least $\ell/4 - \sqrt{2\ell \log n}$ common neighbours in L. Note that $\ell \le 40k$, which implies that $9\sqrt{k} > \sqrt{2\ell}$, and so

$$\frac{\ell}{4} - \sqrt{2\ell \log n} \ge k + 9\sqrt{k \log n} - \sqrt{2\ell \log n} \ge k.$$

Thus our claim implies that every pair of vertices in $[n] \setminus L$ has at least k common neighbours in L, so every edge in the graph can be completed to a k-clique by adding vertices of L, as required.

To prove the claim, take any pair $\{x, y\}$ of vertices in $[n] \setminus L$; the number $N(x, y)$ of common neighbours of x and y in L is a Binomial random variable with parameters $(\ell, 1/4)$. The probability that $N(x, y)$ is less than $\ell/4 - \sqrt{2\ell \log n}$ is at most n^{-4}, by the Chernoff bound, so the expected number of pairs with fewer than $\ell/4 - \sqrt{2\ell \log n}$ common neighbours is at most $1/n^2 = o(1)$, as desired. □

Lemma 3.2 implies that, with $\ell = \lceil 4k + 36\sqrt{k \log n} \rceil$,

$$F[n;k] \ge (1 + o(1)) 2^{\binom{n}{2} - \binom{\ell}{2}} = 2^{\binom{n}{2} - 8k^2 + O(k^{3/2} \log^{1/2} n)},$$

which is a lower bound of the required form whenever $k/\log n \to \infty$.

For $k = c \log n$, we get a lower bound of the form $F[n;k] \ge 2^{\binom{n}{2} - \beta(c) \log^2 n}$, where $\beta(c)$ is a function of c – as it stands, our proof gives the bound $\beta(c) \le 8(c + 9\sqrt{c})^2$, but this could be improved by a more careful analysis. Of course this is only interesting when $c > 2/\log 2$.

For the upper bound, we start with a lemma stating that certain events in the space of (ordinary) random graphs occur with probability at most 2^{-8k^2} – so we can restrict attention to graphs where these events do not occur.

For the rest of this section and the next, we work in the probability space $\mathcal{G}(n,1/2)$ of random graphs G on $[n]$ in which each graph is equally likely. We set

$$s = \left\lceil \frac{k^{4/3}}{n^{1/3}} \right\rceil \quad \text{and} \quad \gamma = \frac{12k}{\sqrt{ns}} \le \min\left(12\left(\frac{k}{n}\right)^{1/3}, \frac{12k}{\sqrt{n}} \right).$$

Note that $s = 1$ whenever $k \le n^{1/4}$, and that $\gamma = o(1)$. Let $N(x)$ denote the set of neighbours of the vertex x in the random graph G.

Lemma 3.3 *Suppose $k \ge \log n$ and $k = o(n)$. Let E_1 be the event that there are sequences $c_1, \dots, c_s$, $d_1, \dots, d_s$ of distinct vertices such that, for each $i = 1, \dots, s$,*

$$|N(c_i) \cap N(d_i) \setminus \{c_1, \dots, c_{i-1}, d_1, \dots, d_{i-1}\}| \ge \frac{n}{4}(1+\gamma).$$

Let E_2 be the event that some set of $\left\lfloor \frac{n}{4}(1+\gamma) \right\rfloor$ vertices spans more than $\frac{n^2}{64}(1+4\gamma)$ edges, and let E_3 be the event that the total number of edges is less than $\frac{n^2}{4}(1-8k/n)$. Then each of E_1, E_2, E_3 has probability at most e^{-8k^2} for sufficiently large n.

Proof. We begin with E_1. Having chosen the sequences d_i and c_i, the sets $N(c_i) \cap N(d_i) \setminus \{c_1, \dots, c_{i-1}, d_1, \dots, d_{i-1}\}$ are independent, and their sizes are each dominated by Binomial random variables with parameters $(n, 1/4)$, so the probability that there are sequences where all are too big is at most

$$n^{2s}\left(e^{-\gamma^2 n/12}\right)^s.$$

Now $\gamma^2 ns/12 = 12k^2$, so we are done if $n^{2s} \le e^{4k^2}$, i.e., $s \log n \le 2k^2$, which is certainly true for sufficiently large n.

For E_2, there are at most 2^n possible "bad" sets of vertices, and the number of edges spanned by a given set of $r = \lfloor (1+\gamma)n/4 \rfloor$ vertices is a Binomial random variable with parameters $\left(\binom{r}{2}, 1/2\right)$; the probability that this exceeds $(1+\gamma)r^2/4$ is at most $e^{-\gamma^2 r^2/12} \le e^{-\gamma^2 n^2/192}$. For sufficiently large n,

$$(1+\gamma)\frac{r^2}{4} \le (1+\gamma)^3 \frac{n^2}{64} \le (1+4\gamma)\frac{n^2}{64},$$

and so the probability of E_2 is at most $2^n e^{-\gamma^2 n^2/192} = 2^n e^{-3k^2 n/4s} \le e^{-k^2 n/2s}$. This is at most e^{-8k^2} for sufficiently large n, since $n/s \to \infty$.

The fact that the probability of E_3 is suitably small is an immediate consequence of the Chernoff bounds. □

Lemma 3.4 *Let G be an $[n;k]$-graph such that none of the events E_1, E_2, E_3 occurs. Let F be a set of edges of G such that every k-clique in G has all but at most $k^{3/2}$ of its edges in F. Then, for sufficiently large n,*

$$|F| \geq 8k^2 \left(1 - \frac{3}{\sqrt{k}} - 60\left(\frac{k}{n}\right)^{1/3}\right).$$

Proof. We start by repeatedly extracting pairs (c_i, d_i) of vertices of G whose neighbourhood intersection is larger than $\frac{n}{4}(1+\gamma)$. Since E_1 does not occur, there is a set S of at most $2s-2$ "bad" vertices such that, in $G\setminus S$, every pair of vertices has common neighbourhood of size at most $\frac{n}{4}(1+\gamma)$ – since E_2 does not occur, this means that the common neighbourhood of any pair of vertices in $G \setminus S$ spans at most $\frac{n^2}{64}(1+4\gamma)$ edges .

We count, in two ways, the number N of *couplets* $(\{a,b\},\{c,d\})$ such that $\{a,b,c,d\}$ forms a clique in G that can be extended to a k-clique in G, and such that $\{c,d\}$ is in F, but neither c nor d is in the set S of bad vertices.

Every edge $\{a,b\}$ of G extends to a k-clique C: of the at least $\binom{k-2s}{2}$ edges between vertices of $C \setminus (S \cup \{a,b\})$, at most $k^{3/2}$ of them are not in F, so we have

$$N \geq |E(G)| \left(\binom{k-2s}{2} - k^{3/2}\right) \geq \frac{n^2}{4}\left(1 - \frac{8k}{n}\right)\frac{k^2}{2}\left(1 - \frac{4s+1}{k} - \frac{2}{\sqrt{k}}\right).$$

On the other hand, each edge of F not incident with a vertex in S only appears in at most $\frac{n^2}{64}(1+4\gamma)$ couplets, so

$$N \leq |F|\frac{n^2}{64}(1+4\gamma).$$

Combining these inequalities yields

$$\begin{aligned} |F| &\geq \tfrac{64}{n^2}(1-4\gamma)\tfrac{n^2}{4}\left(1-\tfrac{8k}{n}\right)\tfrac{k^2}{2}\left(1 - \tfrac{4s+1}{k} - \tfrac{2}{\sqrt{k}}\right) \\ &\geq 8k^2\left(1 - 48\left(\tfrac{k}{n}\right)^{1/3} - 8\tfrac{k}{n} - 4\left(\tfrac{k}{n}\right)^{1/3} - \tfrac{5}{k} - \tfrac{2}{\sqrt{k}}\right), \end{aligned}$$

from which the stated inequality follows for sufficiently large n. □

Theorem 3.5 *If $k = o(n)$ and $k/\log n \to \infty$, then*

$$F[n;k] \leq 2^{\binom{n}{2} - 8k^2 + O(k\log n) + O(k^{3/2}) + O(k^{7/3}n^{-1/3})}.$$

Proof. Let G be any $[n;k]$-graph, and suppose that events E_1, E_2 and E_3 do not occur. Consider the following process to construct a set F of edges of G. Start with F empty. If there is a k-clique C with at least $k^{3/2}$ edges not yet in F, put all the edges of C into F. Stop as soon as $|F| \geq 8k^2(1-\delta)$, where $\delta = \frac{3}{\sqrt{k}} + 60(k/n)^{1/3}$. By Lemma 3.4, there is always some suitable clique C available until F reaches this size. When the process stops, certainly $|F| < 9k^2$.

As at least $k^{3/2}$ new edges are taken into F at each stage, the number of cliques taken before the process stops is at most $8\sqrt{k}$. Also, as each vertex incident with an edge of F is incident with at least $k-1$ edges in F, the total number of such vertices taken is at most $2|F|/(k-1)$.

In summary, if G is a graph such that E_1, E_2 and E_3 do not occur, then there is a set F of edges of size between $8k^2(1-\delta)$ and $9k^2$, with ends in a set of at most $2|F|/(k-1)$ vertices, which can be covered by at most $8\sqrt{k}$ cliques of order k.

The probability that a random graph G contains such a set F is at most

$$\sum_{f=8k^2(1-\delta)}^{9k^2} \binom{n}{2f/(k-1)} 2^{2f/(k-1)\cdot 8\sqrt{k}}\, 2^{-f},$$

where the middle term is an overestimate for the number of ways in which the cliques can be arranged: this estimate allows us a free choice of whether each vertex belongs to each clique. The f-term of this sum is at most

$$n^{2f/(k-1)}\, 2^{2f/(k-1)\cdot 8\sqrt{k}}\, 2^{-f} = 2^{f[2\lg n/(k-1)+16\sqrt{k}/(k-1)-1]}.$$

For $k \geq (2+\varepsilon)\lg n$ and sufficiently large n, the exponent is negative and the quantity is decreasing in f, so the probability that G contains a suitable set F is at most $2k^2 2^{-8k^2(1-\delta)+16k\lg n+128k^{3/2}}$. (Note that $(1-\delta)/(k-1) \leq 1/k$.)

Hence, for sufficiently large n, the probability that a random graph is an $[n;k]$-graph is at most

$$\Pr(E_1 \vee E_2 \vee E_3) + 2^{-8k^2(1-\delta)+128k^{3/2}+16k\lg n} \leq 3e^{-8k^2} + 2^{-8k^2(1-20/\sqrt{k}-40(k/n)^{1/3}-2\lg n/k)},$$

which is an upper bound of the required form. □

Combining the lower and upper bounds gives

$$F[n;k] = 2^{\binom{n}{2}-8k^2+O(k^{3/2}\log^{1/2} n)+O(k^{7/3}n^{-1/3})}$$

whenever $k = o(n)$ and $k/\log n \to \infty$.

Now that we know this much about the structure of “most” $[n;k]$-graphs, we can go on to deduce more, in particular we can show that the set F constructed in the previous proof must be “close to” being the edge-set of a clique of size about $4k$, as in the example showing the lower bound.

For $k = c\lg n$, $c > 2$, the probability of a single k-clique in the random graph is already as small as $2^{-((c-2)/2c+o(1))k^2}$. The proof of Theorem 3.5 gives that the probability that every edge is in a k-clique is at most $2^{-8k^2((c-2)/c+o(1))}$. The methods of the next section give better bounds for values of c just above 2.

4 k just above the threshold

In this section, we are interested in values of k which are just above the threshold for a random graph to contain a k-clique a.a.s. For most values of n, we shall show that, as long as k is large enough to ensure that a random graph a.a.s. contains no k-cliques, the probability that every edge is in a k-clique is already about as small as 2^{-6k^2}.

We shall prove the following theorem.

Theorem 4.1 *Suppose that $k \le 10 \lg n$ and*

$$\frac{k-1}{2} \ge \lg n - \lg\lg n + \lg e - 1 + 10/\lg\lg n.$$

Then $F[n;k] \le 2^{\binom{n}{2} - 6k^2 + o(k^2)}$.

The method used will be an extension of that from the previous section. Throughout what follows, we set $t = \lceil \lg^3 n \rceil$ and $\delta = 1/\lg\lg n$.

Consider the following process, which generates a set F of edges in our graph G that is the union of edge-sets of k-cliques. We start with F empty. While there is a k-clique of G with at least δk^2 edges outside F, and we have so far taken fewer than t cliques, we put the entire edge-set of the clique into F. We stop either because there is no suitable clique – meaning that every k-clique in G has all but at most δk^2 edges in F – or after taking exactly t cliques.

Given any set F of edges of G, made up as the union of edge-sets of k-cliques, let $W = W(F)$ be the set of endpoints of F. Set $f = |F|$ and $w = |W|$. For $x \in W$, let $N_F(x)$ denote the set of vertices joined to x by edges in F, and let $d_F(x) = |N_F(x)|$ denote the number of edges in F incident with x. Note that $d_F(x) \ge k-1$ for all $x \in W$. Let the *excess* of F be $\operatorname{exc}(F) = \sum_{x\in W}(d_F(x) - (k-1))$.

The plan of the proof is as follows. We can bound from above the probability that G contains a suitable set F by an expression of the form $\binom{n}{w} N 2^{-(k-1)w/2 - \operatorname{exc}(F)/2}$, where N is the number of possibilities for arranging the cliques. The terms $\binom{n}{w}$ and $2^{-(k-1)w/2}$ will roughly cancel in the range of interest, so we need to show that either the excess is large, and $2^{-\operatorname{exc}(F)/2}$ drowns out N, or N is sufficiently small to be drowned out by the residual term arising from $\binom{n}{w} 2^{-(k-1)w/2}$. Accordingly, we aim to prove lower bounds on the excess in various circumstances.

First we need a simple technical lemma.

Lemma 4.2 *Let E_4 be the event that there are sets A and X, of sizes $4k - 4k\sqrt{\delta}$ and n/k^4 respectively, such that every element of X has at least $2k - k\sqrt{\delta}$ neighbours in A.*

Let E_5 be the event that there are sets B and Y, of sizes $2k - k\sqrt{\delta}$ and n/k^4 respectively, such that every element of Y has at least $k - k\sqrt{\delta}/4$ neighbours in B.

Then, for n sufficiently large, E_4 and E_5 each have probability at most $e^{-\sqrt{n}}$.

Proof. We give the details for E_4: the proof for E_5 is obviously similar.

For a fixed set A of size $4k(1-\sqrt{\delta})$, and a fixed vertex x, the number of neighbours of x in A is a Binomial random variable with parameters $(4k(1-\sqrt{\delta}), 1/2)$, so, by the Chernoff bound,

$$\Pr(|N(x)\cap A| \geq 2k - 2k\sqrt{\delta} + k\sqrt{\delta}) \leq \exp\left(-\frac{k^2\delta}{3\cdot 2k}\right) = e^{-k\delta/6}.$$

Now for a fixed A, the events E_x that $|N(x)\cap A| \geq 2k - k\sqrt{\delta}$ are independent, so the probability that there is a set X of $K = n/k^4$ such "bad" vertices x is at most

$$\binom{n}{K}\left(e^{-k\delta/6}\right)^K \leq \left(\frac{en}{K}e^{-k\delta/6}\right)^K = \left(ek^4e^{-k\delta/6}\right)^{n/k^4},$$

which is much smaller than $e^{-2\sqrt{n}}$. As there are at most $n^k < e^{-\sqrt{n}}$ choices for A, we are done. □

Besides the events E_4 and E_5, we recall the events E_1, E_2, E_3 of the previous section. Note that $s = 1$ in this range, so that the non-occurrence of E_1 means that there is no pair of vertices in G with common neighbourhood of size at least $(1+\gamma)n/4$.

Lemma 4.3 *Suppose that G is a graph such that none of the properties $E_1, \ldots, E_5$ occurs, and that every edge of G is in a k-clique. Suppose also that the process described above terminates before t cliques are taken, arriving at a set F of edges spanning a set W of w vertices of G. Then $w \geq 4k - o(k)$ and $\mathrm{exc}(F) \geq 12k^2 - o(k^2)$.*

Proof. Let us first see what we can deduce from the fact that E_4 and E_5 do not occur. Fix a set A of $4k(1-\sqrt{\delta})$ vertices of G, and let $J(A)$ be the set of pairs (u,v) of vertices such that $|N(u)\cap N(v)\cap A| \geq k - k\sqrt{\delta}/4$. As E_4 does not occur, there are at most $(n/k^4)n$ pairs (u,v) in $J(A)$ with $|N(u)\cap A| \geq 2k - k\sqrt{\delta}$. Also, as E_5 does not occur, each vertex u with $|N(u)\cap A| < 2k - k\sqrt{\delta}$ gives rise to at most n/k^4 pairs $(u,v)\in J(A)$. Therefore, for any A, $|J(A)| \leq 2n^2/k^4$.

As the process terminates before t cliques are taken, each k-clique of G has at most δk^2 edges outside F. Also, the total number of vertices in W is at most tk.

For a vertex x of a k-clique K, we say that x is *central* to K if $xz \in F$ for all but at most $\sqrt{\delta}k/4$ vertices z of K. Note that, as there are at most δk^2 edges of K that are not in F, the k-clique K has at most $4\sqrt{\delta}k$ non-central vertices.

We develop the method of proof introduced in Lemma 3.4. We call a pair $(\{x,y\},\{u,v\})$ a *central couplet* if:

- $uv \in E(G)$,
- $xy \in F$,
- there is a k-clique K containing u and v such that x and y are central to K.

For each edge uv of $E(G)$, there is a k-clique K containing u and v, and at most $\delta k^2 + 4\sqrt{\delta}k^2$ of the edges xy of K fails one of the conditions above. Thus $\{u,v\}$ appears in at least $k^2/2 - 6\sqrt{\delta}k^2$ central couplets. As E_3 does not occur, the total number of central couplets is at least

$$\frac{n^2}{4}\frac{k^2}{2}(1-o(1)) = \frac{n^2k^2}{8}(1-o(1)).$$

Let W_H denote the set of vertices x of W with $d_F(x) \geq 4k(1-\sqrt{\delta})$, and set $W_L = W \backslash W_H$.

Our plan is to show that the vertices of W_L appear in few central couplets. Indeed, if $(\{x,y\},\{u,v\})$ is a central couplet, witnessed by the k-clique K, then $y \in N_F(x)$, and $N(u) \cap N(v) \cap N_F(x) \supset K \cap N_F(x)$, which has size at least $k - \sqrt{\delta}k/4$, as x is central to K. So $(u,v) \in J(N_F(x))$. If $x \in W_L$, then $|N_F(x)| \leq 4k(1-\sqrt{\delta})$, and $|J(N_F(x))| \leq 2n^2/k^4$. Therefore each $x \in W_L$ appears in at most $|N_F(x)| \cdot |J(N_F(x))| \leq 8n^2/k^3$ central couplets.

Hence the total number of central couplets $(\{x,y\},\{u,v\})$ in which either x or y is in W_L is at most $|W|8n^2/k^3 \leq 8tn^2/k^2 \leq 8n^2k$, which is much less than the total number of central couplets.

Now let F_H denote the set of edges of F between two vertices of W_H. As in the proof of Lemma 3.4, we see that every edge $\{x,y\}$ of F appears in at most $n^2/64(1-o(1))$ central couplets. Therefore we must have

$$|F_H|\frac{n^2}{64}(1-o(1)) \geq \frac{n^2k^2}{8}(1-o(1)),$$

implying $|F_H| \geq 8k^2(1-o(1))$. To span this many edges, we must have $|W_H| \geq 4k(1-o(1))$. Hence certainly $w \geq 4k(1-o(1))$, and also each vertex in W_H contributes at least $3k(1-o(1))$ to $\mathrm{exc}(F)$, so $\mathrm{exc}(F) \geq 12k^2(1-o(1))$, as claimed. □

Given sets W of vertices and F of edges generated by running the process as described above, let B denote the set of vertices of W that are in more than one of the cliques produced during the process, and set $b = |B|$. Also, let q denote the number of times during the process that a vertex already in W is chosen for a clique to be placed into F; note that $q \geq b$.

Lemma 4.4 *The excess* $\mathrm{exc}(F)$ *of* F *is at least* $q\delta k/2$.

Proof. We consider how q and $\mathrm{exc}(F)$ are increased on taking each new clique K into F. There are two cases.

If K contains at least $\delta k/2$ "new" vertices (not already in W), then each "old" vertex in W has its excess raised by at least $\delta k/2$.

On the other hand, if K contains fewer than $\delta k/2$ new vertices, then it contains fewer than $\delta k^2/2$ edges incident with new vertices. By the definition of the process, K contains at least δk^2 edges not currently in F, and each of these edges incident with two old vertices increases the total excess by 2. Thus the total excess increases by at least δk^2 on the inclusion of K, while q increases by at most k.

In either case, if q increases by r on the addition of K, $\mathrm{exc}(F)$ increases by at least $r\delta k/2$. This implies the result. □

Clearly there is something to spare in the argument above, but this result suffices.

In the case when we terminate the process before taking t cliques, we now have two lower bounds on $\mathrm{exc}(F)$. Taking a convex combination, we have that

$$\mathrm{exc}(F) \geq 4\delta\frac{q\delta k}{2} + (1-4\delta)(12k^2 - o(k^2)) = 2q\delta^2 k + 12k^2 - o(k^2).$$

Thus we have that, at the end of our process, either $|F| \geq (k-1)w/2 + q\delta^2 k + 6k^2 - o(k^2)$, or both $|F| \geq (k-1)w/2 + q\delta k/4$ and $s = t$.

We are now ready to prove Theorem 4.1.

Proof. Recall that

$$\frac{k-1}{2} \geq \lg n - \lg\lg n + \lg e - 1 + 10\delta,$$

so that $n2^{-(k-1)/2} \leq \lg n(2/e)(1-5\delta)$. We need to show that the probability that a random graph is an $[n;k]$-graph is at most $2^{-6k^2+o(k^2)}$. In proving this, we may suppose that none of the events $E_1, \ldots, E_5$ occurs, as their probabilities are suitably small.

We consider running our process, terminating after taking s cliques, and distinguish two cases for the value of q at termination.

(a) Suppose $q \geq \delta ks$. Then, for some value of s at most t, our graph G contains a set W of $w \leq ks$ vertices, spanning s k-cliques with an edge-union F of size at least $(k-1)w/2 + \delta^3 k^2 s + 6k^2 - o(k^2)$ if $s < t$, and at least $(k-1)w/2 + \delta^2 k^2 t/4$ if $s = t$.

Consider first the case $s = t$. The probability that G contains such a set W in this case is at most

$$\sum_w \binom{n}{w}\binom{w}{k}^t 2^{-(k-1)w/2-\delta^2 k^2 t/4}.$$

Here $\binom{w}{k}^t$ is a crude bound on the number of ways of choosing the t k-cliques with vertices in W. Again crudely, this probability is at most

$$\sum_w \left(n2^{-(k-1)/2}\right)^w \left(w^k 2^{-\delta^2 k^2/4}\right)^t.$$

Now we use that $w \leq kt$, $kt \leq \lg^4 n$, and $n2^{-(k-1)/2} \leq \lg n$ to bound this above by

$$\sum_w \left(\lg n \lg^4 n 2^{-\delta^2 k/4}\right)^{kt}.$$

The term in parentheses is at most $1/2$ for sufficiently large n (recall that $\delta = 1/\lg\lg n$), and the number of choices for w is at most $\lg^4 n$, which is negligible, so the probability that

G contains a subgraph (W, F) of this form is at most $2^{-kt} = 2^{-O(\lg^4 n)}$, which is even smaller than we require.

In the case where $s < t$, the calculation is effectively the same with the alternative estimate for $|F|$ being used, and we see that the probability that G contains a subgraph (W, F) of the required form is at most

$$\sum_s \left(\lg^5 n 2^{-\delta^3 k}\right)^{ks} 2^{-6k^2+o(k^2)} \leq 2^{-6k^2+o(k^2)}.$$

as required.

(b) Suppose $q \leq \delta ks$.

Let $A = W \setminus B$ be the set of vertices of W that are in exactly one of the cliques taken during the process. Set $a = |A|$, and observe that $a = ks - b - q \geq (1 - 2\delta)ks$.

Given sets A and B of appropriate sizes, we need an upper bound on the number of ways to arrange them into s k-sets $S_1, \ldots, S_s$ so that every element of A occurs exactly once, and the total number of occurrences of elements of B is $q + b$. There are just $\binom{bs}{q+b}$ ways of choosing the $b + q$ instances of occurrences of elements of B in a set S_i. Having chosen these instances, we have to top each set S_i up by some known number a_i, $(i = 1, \ldots, s)$, where the a_i sum to a. The number of ways to do this is just the number of ways to partition A into sets of sizes $a_1, \ldots, a_s$, which is $a!/(a_1! \cdots a_s!)$; this is greatest when all the a_i are as nearly equal as possible, and is at most $a!/((1 - \delta)a/se)^{(a/s)s} \leq a!/((1 - 3\delta)k/e)^a$: here we used a crude version of Stirling's formula and the bound $a \geq (1 - 2\delta)ks$.

Again let us start with the case where we terminate with $s = t$, so that our upper bound on $|F|$ is simply $(k - 1)(a + b)/2 + q\delta k/4$. In this case, the probability that G contains sets A and B spanning cliques as necessary is at most

$$\sum_{a,b,q} \binom{n}{a}\binom{n}{b}\binom{bt}{q+b} \frac{a!}{((1 - 3\delta)k/e)^a} 2^{-a(k-1)/2-b(k-1)/2-q\delta k/4}.$$

Collecting terms with powers a, b and q, and using standard estimates, shows that this sum is at most

$$\sum_{a,b,q} \left(\frac{n2^{-(k-1)/2}}{(1 - 3\delta)k/e}\right)^a \left(nbt2^{-(k-1)/2}\right)^b \left(bt2^{-\delta k/4}\right)^q.$$

Note that $k \geq 2\lg n(1 - \delta)$, that $b \leq q$, and that $bt \leq \delta kt^2 \leq \lg^7 n$, so the probability is at most

$$\sum_{a,b,q} \left(\frac{\lg n(2/e)(1 - 5\delta)}{(1 - 4\delta)2\lg n/e}\right)^a \left(\lg^{15} n2^{-\delta k/4}\right)^q.$$

The second term here is at most 1, and the number of terms in the sum is at most $\lg^{12} n$, which is not significant. Hence the probability is at most

$$\lg^{12} n(1 - \delta)^{(1-2\delta)kt} \leq 2^{-\delta kt/2},$$

which is suitably small.

Finally we have to repeat the last calculation in the case where we terminate before taking t cliques, in which case we have a stronger lower bound on $|F|$. We obtain, much as above, the following upper bound on the probability of a subgraph (W, F) of the necessary form:

$$\sum_{a,b,q,s} (1-\delta)^{(1-2\delta)ks} \left(\lg^{15} n 2^{-\delta^2 k}\right)^q 2^{-6k^2+o(k^2)} \leq 2^{-6k^2+o(k^2)},$$

as required. □

5 Large k

We now work from the opposite end of the spectrum, working down from the largest values of k.

Clearly $F[n;n] = 2$ and $F[n;n-1] = \binom{n}{2} + n + 2$, since a union of $(n-1)$-cliques either misses exactly one edge (union of two cliques), or is a single clique plus an isolated vertex, or is empty or complete. We can similarly calculate $F[n;n-2]$ exactly.

Based on the examples above, it looks as though, for r fixed, there are only finitely many isomorphism classes of $[n;n-r]$-graphs, and $F[n;n-r]$ is a polynomial in n whose degree increases with r. We show that this is true, and find exactly the coefficient of the leading term of $F[n;n-r]$.

Lemma 5.1 *Suppose G is an $[m;m-r]$-graph, having no vertex of degree* 0 *or* $m-1$*. Then* $m \leq \lfloor (r+2)^2/4 \rfloor$*.*

Furthermore, if equality holds and $r \geq 4$, then the complement G^c of G consists of a clique T of size t, and a copy of $K_{1,s}$ rooted at each vertex of T, where: $t = s+1 = \frac{r+2}{2}$ if r is even and $\{t, s+1\} = \{\frac{r+1}{2}, \frac{r+3}{2}\}$ if r is odd.

Proof. Since G^c has no isolated vertex, we can take a spanning subgraph H of G^c consisting of stars (i.e., copies of some $K_{1,s}$ with $s \geq 1$). Let $x_1, \ldots, x_t$ be the roots of these stars, and let s_i be the number of leaves adjacent to x_i in H. Without loss of generality $s_1 = s$ is the largest of the s_i, so $m \leq t(s+1)$. Now consider an $(m-r)$-clique C in G containing x_1: this does not contain any of the s adjacent leaves, and also, for each $j \neq 1$, misses either x_j or all of its associated leaves. So C misses at least $s+t-1$ vertices, and therefore $s+t-1 \leq r$, or $t+(s+1) \leq r+2$. From this and $m \leq t(s+1)$, we conclude that $m \leq \lfloor (r+2)^2/4 \rfloor$.

Furthermore, if we have equality, then t and $s+1$ must both be as close as possible to $(r+2)/2$. Also, all the s_i must be equal to s. Moreover, provided $s \geq 2$, the only $(m-r)$-clique in G containing x_i misses exactly its associated leaves and the other x_j: therefore the x_i form a clique in G^c. Provided $t \geq 3$, the union of the t cliques containing the x_i contains all edges between leaves, and so is the graph stated.

Finally we observe that the graphs described in the theorem are $[m; m-r]$-graphs. □

Theorem 5.2 *For fixed $r \geq 1$, $F[n; n-r]$ is a polynomial of degree $\lfloor (r+2)^2/4 \rfloor$ in n. The leading coefficient $L(r)$ is 1/2 for $r = 1$, 3/4 for $r = 2$ and 17/9 for $r = 3$. Furthermore, for $r \geq 4$,*

$$L(r) = \begin{cases} \left((r/2)!^{r/2+1}(r/2+1)!\right)^{-1} & \text{if } r \text{ is even,} \\ \left(\frac{r+1}{2}\right)!^{-(r+3)/2} + \left(\left(\frac{r-1}{2}\right)!^{(r+3)/2}\left(\frac{r+3}{2}\right)!\right)^{-1} & \text{if } r \text{ is odd.} \end{cases}$$

Proof. Set $m_0(r) = \lfloor (r+2)^2/4 \rfloor$. Observe that, for $r \geq 1$, $m_0(r+1) > m_0(r) + 1$.

Let $F'[n; n-r]$ denote the number of $[n; n-r]$-graphs with no isolated vertices. We shall prove that $F'[n; n-r]$ is a polynomial of degree $m_0(r)$, with leading coefficient $L(r)$ as stated in the theorem. The result will then follow, since

$$F[n; n-r] = 1 + \binom{n}{r} + \sum_{q=0}^{r-1} \binom{n}{q} F'[n-q; n-r].$$

Here, the initial 1 accounts for the empty graph, and the next term counts the single $(n-r)$-cliques; the q-term in the sum counts the $[n; n-r]$-graphs with exactly q isolated vertices. The q-term in the sum is a polynomial of degree $q + m_0(r-q)$, and the unique largest of these is when $q = 0$.

By Lemma 5.1, for any $[n; n-r]$-graph G with no isolated vertices, all but at most $m_0 = m_0(r)$ of the vertices of G have degree $n-1$. For $m \leq m_0$, let $C(m, r)$ be the number of $[m; m-r]$-graphs with no vertex of degree 0 or $m-1$; then

$$F'[n; n-r] = \sum_{m=0}^{m_0} \binom{n}{m} C(m, r),$$

so $F'[n; n-r]$ is indeed a polynomial of degree $m_0(r)$ in n, and its leading coefficient $L(r)$ is $C(m_0, r)/m_0!$.

Also by Lemma 5.1, if $r \geq 4$ is even then $C(m_0, r)$ is the number of ways of partitioning m_0 labelled vertices into $r/2 + 1$ stars $K_{1,r/2}$, which is $m_0!/\left((r/2)!^{r/2+1}(r/2+1)!\right)$. If $r \geq 5$ is odd, then similarly

$$C(m_0, r) = m_0! \left(\frac{r+1}{2}\right)!^{-(r+3)/2} + m_0! \left(\frac{r-1}{2}\right)!^{-(r+3)/2} \left(\frac{r+3}{2}\right)!^{-1},$$

as claimed.

The cases $r = 2$ and $r = 3$ need to be handled separately. Investigation of the various possibilities shows that $C(4, 2) = 18$ and $C(6, 3) = 1360$. □

Another way of looking at what we have proved is that, for fixed r and n sufficiently large (at least $m_0(r)$), the number of isomorphism classes of $[n; n-r]$-graphs is fixed.

Note that, for r fixed, the leading term of the polynomial $F[n;n-r]$ dominates as $n \to \infty$, so we have

$$F[n;n-r] = L(r)n^{\lfloor (r+2)^2/4 \rfloor}(1+o(1)).$$

The basic behaviour $F[n;n-r] \simeq n^{r^2/4}$ is actually valid whenever $r = o(n)$, as we now show.

Theorem 5.3 *For $r = o(n)$,*

$$F[n;n-r] = \exp\left(\frac{r^2}{4}\log(n/r) + O(r\log n) + O(r^2)\right) = \exp\left(\frac{r^2}{4}\log n(1+o(1))\right).$$

Proof. For the lower bound, consider the family of all graphs with vertex set $V = [n]$ of the following form: there is a designated set T of $t = \lceil r/2 \rceil$ vertices – T forms a clique, and each vertex x of T is also adjacent to a set $S(x)$ of $r-t+1 = \lceil (r+1)/2 \rceil$ vertices from $V \setminus T$. For any such graph H, the complement G is an $[n;n-r]$-graph: for each vertex $x \in T$, take a clique on $V \setminus (S(x) \cup T \setminus \{x\})$, and take also any other cliques inside $V \setminus T$ required so that $V \setminus T$ is a clique in G.

The number of graphs H constructed as above is $\binom{n}{t}\binom{n-t}{r-t+1}^t$, which is of the form stated in the theorem: note that, given H, we can always recover T as all other vertices have lower degree.

We now turn to the upper bound. Suppose that G is a non-empty $[n;n-r]$-graph, and let H be the complement G^c. Note that H has no matching of size greater than r (otherwise there is no $(n-r)$-clique in G). Thus by the defect form of Tutte's 1-factor Theorem [5], there is a set T of $t \le r$ vertices such that $H \setminus T$ has at least $n-2r+t$ odd-order components; furthermore, taking T minimal with this property ensures that there is a matching M of t edges in H, each containing exactly one vertex of T: say $V(M) = T \cup T'$.

Suppose that $n-t-u$ of the odd-order components of $H \setminus T$ are single vertices, so there are exactly u vertices in non-trivial components of $H \setminus T$. However, there are at least $u+2t-2r$ non-trivial odd-order components, so at least $3(u+2t-2r)$ vertices in these components. It follows that $u \ge 3(u+2t-2r)$, so that $u \le 3(r-t)$.

Now observe that each vertex x of T either has degree $n-1$ in H, or is adjacent to at most $r-t$ vertices outside $T \cup T'$ in H; otherwise there is no clique in G containing x of size $n-r$, as any such clique must miss one vertex on each edge in M, as well as all the other neighbours of x in H.

To recap, there is a set T of $t \le r$ vertices in H, and a set W (the set of vertices in non-trivial components of $H \setminus T$, together with T') of order at most $3r-2t$, such that all neighbours of vertices in $Z = V(H) \setminus (T \cup W)$ are in T, and each vertex of T either has degree $n-1$ or has at most $r-t$ neighbours in Z. (We can say more, but this is all we need.)

The number of graphs H with this structure, for given t, is at most

$$\binom{n}{t}\binom{n-t}{3r-2t}2^{\binom{3r-t}{2}}\left[\binom{n-3r+t}{\le r-t}+1\right]^t = \exp\left(t(r-t)\log(n/r) + O(r\log n) + O(r^2)\right).$$

This is maximized when $t = r/2$, and the result follows. □

6 Middling k

We have found reasonably good estimates for $F[n;k]$ if $k = o(n)$ and also if $n - k = o(n)$; it is natural to ask what happens if $k = cn$, for some fixed c with $0 < c < 1$. In this case the probability that a random graph contains a single k-clique is $2^{-c^2n^2/2+O(n)}$, and we shall present lower bounds of the same form, thus showing that

$$2^{\alpha(c)n^2} \leq F[n;cn] \leq 2^{\beta(c)n^2},$$

for some $\alpha(c)$, $\beta(c)$ with $0 < \alpha < \beta < 1/2$. In fact, we conjecture that the lower bounds implicit from the examples we present below give the correct answer.

Our examples will be variants of the families we know to be "optimal" at either end of the range.

For c small, consider graphs G of the following type. The vertex set is split into a clique A of size an, and an arbitrary graph on the set B containing the remaining $(1-a)n$ vertices. Between A and B is a graph with edge-density $q = (1+\epsilon)\sqrt{c/a}$ (ϵ can be taken to be $n^{-1/4}$, for instance). Such a graph a.a.s. has the property that every two vertices of B have at least $q^2an(1-\epsilon) > cn$ common neighbours in A, and therefore is an $[n;cn]$-graph. Therefore

$$F[n;cn] \geq 2^{(1-a)^2n^2/2}\binom{g}{qg}; \quad g = a(1-a)n^2,$$

so

$$\frac{\lg F[n;cn]}{n^2} \geq \frac{(1-a)^2}{2} + a(1-a)H(q)(1-o(1)),$$

where $H(x) = -x\lg x - (1-x)\lg(1-x)$ is the entropy function. Given c, one can maximize this expression over a (setting $\epsilon = 0$ for the purposes of the calculation, so $q = \sqrt{c/a}$). However, there seems to be no particularly pleasant way of expressing the outcome of this calculation.

For c large, consider graphs G of the following type. The vertex set is split into a clique A of size an, and an independent set B of size $(1-a)n$. Between A and B is a graph with edge-density $q = (1+\epsilon)c/a$. This time all we need is that each single vertex of B has at least cn neighbours in A, and this is indeed the case a.a.s. So

$$F[n;cn] \geq \binom{g}{qg}; \quad g = a(1-a)n^2,$$

and therefore

$$\frac{\lg F[n;cn]}{n^2} \geq a(1-a)H(c/a)(1-o(1)).$$

Again, one can maximize this over a, and again there seems to be no straightforward way to express the result.

Our calculations suggest that the first family is larger for $c \leq 0.51$, and the second is larger for $c \geq 0.52$.

7 Related questions

As we mentioned at the beginning of the paper, our interest in this problem originated from the study of a closely related problem: how many subsets of the n-dimensional cube can be written as unions of k-dimensional subcubes? Indeed, our problem is a natural translation of this from the cube to the complete graph, with an equally natural choice of specified substructure: clique rather than subcube.

There are some other combinatorial structures where similar questions might be of some interest. For example, the same framework can be translated to the setting of hypergraphs, of bipartite graphs, or of grids. To be precise, here are a number of questions, or rather families of questions.

(1) How many r-uniform hypergraphs on n vertices can be written as the union of complete r-uniform hypergraphs on k vertices?

(2) How many bipartite graphs with specified vertex classes A and B of size n can be written as the edge-union of complete bipartite graphs $K_{m,m}$? Or $K_{s,t}$? This can also be interpreted geometrically: for an $n \times n$ piece X of the rectangular grid, how many subsets of X can be written as the union of $m \times m$ subgrids? Here a subgrid is defined by any choice of m vertical co-ordinates and m horizontal co-ordinates, not necessarily consecutive. The question can also be asked in higher dimensions.

(3) A similar question in a slightly different setting has recently been asked by Verstraëte and studied by Green and Ruzsa [4]: how many subsets of $\{1, \dots, n\}$ can be written as $A + A = \{a + b : a, b \in A\}$ for some A? Here too one could investigate extensions: what about $A + A + A$, for instance?

(4) Returning to our setting of cliques in graphs, it is not hard to think up variations on our problem. For instance, one could ask the same question in the space $G(n, p)$ of random graphs for non-constant $p = p(n)$. (Presumably there are no surprises for other constant values of p.)

(5) Or one could ask for the number of graphs with vertex set $[n]$ that are the edge-union of *disjoint* k-cliques. Even the case $k = 3$ is not trivial. Or one could restrict the number of cliques, or ask about the number of *unlabelled* graphs.

Returning finally to our threshold result, Theorem 4.1, it would be of great interest to discover exactly how $F[n; k]/2^{\binom{n}{2}}$ behaves very near the threshold. Our belief is that the proper way to view this is to treat k as the independent parameter, and look at how $F[n; k]/2^{\binom{n}{2}}$ varies with $n = n(k)$.

References

[1] Béla Bollobás, *Random Graphs*, Second edition, Cambridge University Press, 2001, xviii + 498pp.

[2] Béla Bollobás and Graham Brightwell, The number of k-SAT functions, to appear in *Random Structures and Algorithms*.

[3] Béla Bollobás, Graham Brightwell and Imre Leader, The number of 2-SAT functions, *Israel J. Math.* **133** (2003) 45–60.

[4] Ben Green and Imre Ruzsa, Counting sumsets and sumfree sets I, submitted.

[5] W.T. Tutte, A factorization of linear graphs, *J. London Math. Soc.* **22** (1947) 107–111.