FACTORS IN RANDOM GRAPHS

ANDERS JOHANSSON, JEFF KAHN, AND VAN VU

ABSTRACT. Let H be a fixed graph on v vertices. For an n-vertex graph G with n divisible by v, an *H-factor* of G is a collection of n/v copies of H whose vertex sets partition $V(G)$.

In this paper we consider the threshold $th_H(n)$ of the property that an Erdős-Rényi random graph (on n points) contains an H-factor. Our results determine $th_H(n)$ for all strictly balanced H.

The method here extends with no difficulty to hypergraphs. As a corollary, we obtain the threshold for a perfect matching in random k-uniform hypergraph, solving the well-known "Shamir's problem."

1. INTRODUCTION

Let H be a fixed graph on v vertices. For an n-vertex graph G with n divisible by v, an *H-factor* of G is a collection of n/v copies of H whose vertex sets partition $V(G)$. (For the purposes of this introduction, a copy of H in G is a (not necessarily induced) subgraph of G isomorphic to H; but we will later find it convenient to deal with *labeled* copies.) *We assume throughout this paper that v divides n.*

Let $G(n,p)$ be the Erdős-Rényi random graph with edge density p. Recall that a function $f(n)$ is said to be a *threshold* for an increasing graph property Q if

$$\begin{array}{ll} \Pr(G(n,p) \text{ satisfies } Q) \to 1 & \text{if } p = p(n) = \omega(f(n)), \text{ and} \\ \Pr(G(n,p) \text{ satisfies } Q) \to 0 & \text{if } p = p(n) = o(f(n)). \end{array} \tag{1}$$

Of course if $f(n)$ is a threshold, then so is $cf(n)$ for any positive constant c; nonetheless, following common practice, we will sometimes say "*the* threshold for Q."

Here we are interested in the (increasing) property that $G(n,p)$ has an H-factor. We denote by $th_H(n)$ a threshold for this property (where, to be precise, we really mean (1) holds for $n \equiv 0 \pmod{v}$). Determination of the thresholds th_H is a central problem in the theory of random graphs, particular cases of which have been considered e.g. in [2, 13, 16, 19].

In this paper we come close to a complete solution to this problem, determining $th_H(n)$ for strictly balanced graphs (defined below), and getting to within a factor $n^{o(1)}$ of optimal in general. Our results generalize without difficulty to hypergraphs. Here the simplest case—namely, where H consists of a single hyperedge—settles

J. Kahn is supported by an NSF Grant.
V. Vu is supported by an NSF Career Grant.

the much-studied "Shamir's problem" on the threshold for perfect matchings in random hypergraphs of a fixed edge size. (We will say a little more about this below, but for simplicity have chosen to give detailed arguments only for strongly balanced graphs and just comment at the end of the paper on how these arguments extend.)

To get some feeling for the magnitude of $th_H(n)$, let us first consider lower bounds. If G contains an H-factor, then each vertex of G is covered by some copy of H. So, if we denote by $th_H^{[1]}(n)$ the threshold for the property that in $G(n,p)$ every vertex is covered by a copy of H, then

$$th_H^{[1]}(n) \le th_H(n). \tag{2}$$

For strictly balanced graphs (2) will turn out to give the correct value for th_H. But for graphs in which some regions are denser than (or at least as dense as) the whole, we can run into the problem that, though the copies of H cover $V(G)$, there is at least one vertex x of H for which only a few vertices of G play the role of x (in the obvious sense) in these copies. Formally we have

$$(th_H^{[1]}(n) \le)\ \ th_H^{[2]}(n) \le th_H(n), \tag{3}$$

where $th_H^{[2]}(n)$ is the threshold for the property:

- every vertex of G is covered by at least one copy of H, and
- for each $x \in V(H)$, there are at least n/v vertices x' of G for which some isomorphism of H into G takes x to x'.

(We could replace n/v by $\Omega(n)$ without affecting this threshold.)

We believe $th_H^{[2]}$ tells the whole story:

Conjecture 1.1. *For every H,*

$$th_H^{[2]}(n) = th_H(n). \tag{4}$$

Though Conjecture 1.1 may already be a little optimistic, it does not seem impossible that even a "stopping time" version is true, *viz.*: if we start with n isolated vertices and add random (uniform) edges $e_1, \dots,$ then with probability $1 - o(1)$ we have an H-factor as soon as we have the property described above in connection with $th_H^{[2]}$. See [10] and [15, 4] for the analogous statements for matchings and Hamiltonian cycles respectively.

We next want to say something about the behavior of $th_H^{[1]}$ and $th_H^{[2]}$. For the issues we are considering the following notion of density is natural. Throughout the paper we use $v(G)$ and $e(G)$ for the numbers of vertices and edges of a graph G.

Definition 1.2. *For a graph H on at least two vertices,*

$$d(H) = \frac{e(H)}{v(H) - 1}$$

and

$$d^*(H) = \max_{H' \subseteq H} d(H').$$

Definition 1.3. *A graph H is* strictly balanced *if for any proper subgraph H' of H with at least two vertices, $d(H') < d(H)$.*

(When the weaker condition $d^*(H) = d(H)$ holds H is said to be *balanced.*)

Strictly balanced graphs are the primary concern of the present paper and are of considerable importance for the theory of random graphs in general; see for example [3, 12]. Basic graphs such as cliques and cycles are strictly balanced, as is a typical random graph.

The thresholds $th_H^{[1]}$ are known; see [12, Theorem 3.22]. In particular, for any H one has $th_H^{[1]}(n) = \Omega(n^{-1/d(H)}(\log n)^{1/m})$, where m is the number of edges in H. To see why this is natural, notice that the expected number of copies of H covering a fixed $x \in V(G(n,p))$ is $\Theta(n^{v-1}p^m)$. We may then guess that (i) if this expectation is much less than $\log n$ then the number is zero with probability much more than $1/n$, and (ii) in this case it is likely that there are x's for which the number *is* zero.

To specify $th_H^{[2]}$, we need a little notation. For $v \in V(H)$ let

$$d^*(v,H) = \max\{d(H') : H' \subseteq H, v \in V(H')\}$$

(the "local density" at v). Then clearly $d^*(v,H) \le d^*(H)$ for all v, with equality for some v. Let $s_v = \min\{e(H') : H' \subseteq H, v \in V(H'), d(H') = d^*(v,H)\}$ and let s be the maximum of the s_v's. (Similar notions enter into the determination of $th_H^{[1]}$; again see [12], noting that our uses of s_v and s are not the same as theirs.) A proof of the following assertion will appear separately.

Lemma 1.4. *If $d^*(v,H) = d^*$ $\forall v \in V(H)$ then*

$$th_H^{[2]}(n) = n^{-1/d^*} \log^{1/s} n.$$

Otherwise

$$th_H^{[2]}(n) = n^{-1/d^*}.$$

In particular we have

$$th_H^{[1]}(n) = th_H^{[2]}(n) = n^{-1/d(H)}(\log n)^{1/m} \tag{5}$$

for a strictly balanced H with with m edges (see [22, 19]), and

$$th_H^{[2]}(n) \ge n^{-1/d^*(H)} \tag{6}$$

for a general H.

Conjecture 1.1 says that the values for $th_H^{[2]}$ given in Lemma 1.4 are also the values for th_H. Some cases of the conjecture were known earlier: when H consists of a single edge, it is just the classic result of Erdős and Rényi [10] giving $\log n/n$ as the threshold for a perfect matching in $G(n,p)$; and in [19, 2] (see also [12, Section 4.2]) it is proved whenever $d^*(H) > \delta(H)$ (the minimum degree of H).

The most studied case of the problem is that when H is a triangle, for which the conjectured value is $\Theta(n^{-2/3}(\log n)^{1/3})$. This problem was apparently first suggested by Ruciński [20]; see [12, Section 4.3] for further discussion. Partial results of the form $th_H(n) \le O(n^{-2/3+\varepsilon})$ were obtained by Krivelevich [16] ($\varepsilon = 1/15$) and Kim [13] ($\varepsilon = 1/18$).

2. New results

The main purpose of this paper is establishing Conjecture 1.1 for strictly balanced graphs:

Theorem 2.1. *Let H be a strictly balanced graph with m edges. Then*
$$th_H(n) = \Theta(n^{-1/d(H)}(\log n)^{1/m}).$$

For general graphs, we have

Theorem 2.2. *For an arbitrary H,*
$$th_H(n) = O(n^{-1/d^*(H)+o(1)}).$$

This is Conjecture 3.1 of [2]. Note that, in view of (6), the bound is sharp up to the $o(1)$ term.

We will actually prove more than Theorems 2.1 and 2.2, giving (lower) bounds on the *number* of H-factors. Here, for simplicity, we confine the discussion to strictly balanced H. From this point through the statement of Theorem 2.4 we fix a strictly balanced graph H on v vertices and m edges (thus $d(H) = m/(v-1)$).

We use $\mathcal{F}(G) = \mathcal{F}_H(G)$ for the set of H-factors in G and $\Phi(G) = \Phi_H(G) = |\mathcal{F}(G)|$. For the complete graph K_n this number is
$$\frac{n!}{(n/v)!}\left(\frac{1}{|Aut(H)|}\right)^{n/v} = n^{\frac{v-1}{v}n}e^{-O(n)} \tag{7}$$

where $Aut(H)$ is the automorphism group of H. (Note: in general the expressions $O(n)$, $o(n)$ can be positive or negative, so the minus sign is not really necessary here.) Thus, recalling E denotes expectation, we have
$$\mathsf{E}\Phi(G(n,p)) = n^{\frac{v-1}{v}n}e^{-O(n)}p^{mn/v} = e^{-O(n)}(n^{v-1}p^m)^{n/v}. \tag{8}$$
Strengthening Theorem 2.1, we will show that, for a suitably large constant C and
$$p > Cn^{-1/d(H)}(\log n)^{1/m},$$
$\Phi(G(n,p))$ is close to its expectation:

Theorem 2.3. *For any C_1 there is a C_2 such that for any $p > C_2 n^{-1/d(H)}(\log n)^{1/m}$,*
$$\Phi(G(n,p)) = e^{-O(n)}(n^{v-1}p^m)^{n/v}$$
with probability at least $1 - n^{-C_1}$.

Again note $O(n)$ can be positive or negative. The upper bound follows from (8) via Markov's inequality. Our task is to prove the lower bound, for which it will be convenient to work with the following alternate version.

We say that an event holds *with very high probability (w.v.h.p.)* if the probability that it fails is $n^{-\omega(1)}$. We will prove the following equivalent form of Theorem 2.3. (The equivalence, though presumably well-known, is proved in an appendix at the end of the paper.)

Theorem 2.4. *For $p = \omega(n^{-1/d(H)}(\log n)^{1/m})$, the number of H-factors in $G(n,p)$ is, with very high probability, at least $e^{-O(n)}(n^{v-1}p^m)^{n/v}$.*

As suggested above, the analogous extension of Theorem 2.2 also holds.

Finally, we should say something about hypergraphs. Recall that a *k-uniform hypergraph on vertex set V* is simply a collection of k-subsets, called *edges*, of V. Write $\mathcal{H}_k(n,p)$ for the random k-uniform hypergraph on vertex set $[n]$; that is, each k-set is an edge with probability p, independent of other choices.

The preceding results extend essentially *verbatim* to k-uniform hypergraphs with a fixed k. In particular, as the simplest case (when H is one hyperedge), we have a resolution of the question, first studied by Schmidt and Shamir [21] and sometimes referred to as "Shamir's problem": for fixed k and n ranging over multiples of k, what is the threshold for $\mathcal{H}_k(n,p)$ to contain a perfect matching (meaning, of course, a collection of edges partitioning the vertex set)? The problem seems to have first appeared in [9], where Erdős says that he heard it from E. Shamir; this perhaps explains the name.

Shamir's problem has been one of the most studied questions in probabilistic combinatorics over the last 25 years. The natural conjecture, perhaps first explicitly proposed in [7], is that the threshold is $n^{-k+1}\log n$, reflecting the idea that, as for graphs, isolated vertices are the primary obstruction to existence of a perfect matching. Progress in the direction of this conjecture and related results may be found in, for example, [21, 11, 13, 7, 17]. Again, some discussion of the problem may be found in [12, Section 4.3].

We just state the threshold results for hypergraphs, the relevant definitions and notation extending without modification to this context.

Theorem 2.5. *For a strictly balanced k-uniform hypergraph H with m edges,*

$$th_H(n) = \Theta(n^{-1/d(H)}(\log n)^{1/m}).$$

Corollary 2.6. *The threshold for perfect matching in a k-uniform random hypergraph is*

$$\Theta(n^{-k+1}\log n).$$

Theorem 2.7. *For an arbitrary k-uniform hypergraph H,*

$$th_H(n) = O(n^{-1/d^*(H)+o(1)}).$$

Again, the counting versions of these statements also hold (and are what's actually obtained by following the proof of Theorem 2.3).

In the next section we give an overview of the proof of Theorem 2.4 and (at the end) an outline of the rest of the paper.

Notation and conventions. We use asymptotic notation under the assumption that $n \to \infty$, and assume throughout that n is large enough to support our assertions. We will often pretend that large numbers are integers, preferring this common abuse to cluttering the paper with irrelevant "floor" and "ceiling" symbols.

For a graph G, $V(G)$ and $E(G)$ denote the vertex and edge sets of G, respectively, and, as earlier, we use $v(G)$ and $e(G)$ for the cardinalities of these sets. As earlier, for our fixed H we always take $v(H) = v$ and $e(H) = m$.

Throughout the paper V is $[n] := \{1, \dots, n\}$ and, as usual, K_n is the complete graph on this vertex set. We also use K_W for the complete graph on a general vertex set W.

We use log for natural logarithm, $\mathbf{1}_{\mathcal{E}}$ for the indicator of event $\mathcal{E}$, and Pr and E for probability and expectation.

As noted earlier, we will henceforth use *copy of H in G* to mean an injection $\varphi : V(H) \to V(G)$ that takes edges to edges (i.e. $\varphi(E(H)) \subseteq E(G)$ with the natural interpretation of $\varphi(E(H))$). This avoids irrelevant issues involving $|\mathrm{Aut}(H)|$ and will later make our lives somewhat easier in other ways as well. Note that since this change multiplies $\Phi(G)$ by $|Aut(H)|^{n/v} = e^{O(n)}$, it does not affect the statements of Theorems 2.3 and 2.4. Note also that we will still sometimes use (e.g.) K for a copy of H, in which case we think of a labeled copy in the usual sense.

We use $\mathcal{H}(G)$ for the set of copies of H in G, $\mathcal{H}(x, G) = \{K \in \mathcal{H}(G) : x \in V(K)\}$, (where $x \in V(G)$), and $D(x, G) = |\mathcal{H}(x, G)|$.

3. Outline

In this section we sketch the proof of Theorem 2.4, or, more precisely, give the proof modulo various assertions (and a few definitions) that will be covered in later sections. For the moment we actually work with the model $G(n, M)$—that is the graph chosen uniformly from M-edge graphs on V—though in proving the assertions made here it will turn out to be more convenient to return to $G(n, p)$. The relevant connection between the two models is given by Lemma 4.1, which in particular implies that Theorem 2.4 is equivalent to

Theorem 3.1. *For $p = p(n) = \omega(n^{-1/d(H)}(\log n)^{1/m})$ and $M = M(n) = \binom{n}{2}p$,*

$$\Pr(\Phi(G(n, M) \geq (n^{v-1}p^m)^{n/v}e^{-O(n)}) \geq 1 - n^{-\omega(1)}.$$

For the proof of this, let $M = M(n)$ be as in the statement and $T = \binom{n}{2} - M$. Let $e_1, \ldots, e_{\binom{n}{2}}$ be a random (uniform) ordering of $E(K_n)$ and set $G_i = K_n - \{e_1, \ldots, e_i\}$ (so $G_0 = K_n$) and $\mathcal{F}_i = \mathcal{F}(G_i)$ (recall this is the set of H-factors in G_i). Let ξ_i be the fraction of members of $\mathcal{F}_{i-1}$ containing e_i (where we think of an H-factor as a subgraph of G in the obvious way). Then (for any t)

$$|\mathcal{F}_t| = |\mathcal{F}_0|\frac{|\mathcal{F}_1|}{|\mathcal{F}_0|}\cdots\frac{|\mathcal{F}_t|}{|\mathcal{F}_{t-1}|} = |\mathcal{F}_0|(1-\xi_1)\cdots(1-\xi_t),$$

or

$$\log|\mathcal{F}_t| = \log|\mathcal{F}_0| + \sum_{i=1}^{t}\log(1-\xi_i). \tag{9}$$

Now

$$\log|\mathcal{F}_0| = \log\frac{n!}{(n/v)!|Aut(H)|^{n/v}} = \frac{v-1}{v}n\log n - O(n). \tag{10}$$

We also have

$$\mathsf{E}\xi_i = \frac{mn/v}{\binom{n}{2} - i + 1} =: \gamma_i, \tag{11}$$ (*1)

since in fact

$$\mathsf{E}[\xi_i|e_1, \ldots, e_{i-1}] = \gamma_i \qquad -\frac{mn}{v}\log p \tag{12}$$

for any choice of $e_1, \ldots, e_{i-1}$. Thus

$$\sum_{i=1}^{t}\mathsf{E}\xi_i = \sum_{i=1}^{t}\gamma_i = \frac{mn}{v}\log\frac{\binom{n}{2}}{\binom{n}{2}-t} + o(1), \tag{13}$$

$\sim \frac{mn}{dv}\log n$

provided $\binom{n}{2} - t > \omega(n)$. Let $\mathcal{A}_t$ $(= \mathcal{A}_t(n))$ be the event

$\{-\sum_{i=1}^{t}\xi_i \cdots > -\sum_{i=1}^{t}\gamma_i - O(n)\}$

$$\{\log|\mathcal{F}_t| > \log|\mathcal{F}_0| - \sum_{i=1}^{t}\gamma_i - O(n)\}.$$

$\equiv$

$|X_t| = O(n)$

Our basic goal is to show that failure of $\mathcal{A}_t$ is unlikely. Precisely, we want to show

$$\text{for } t \le T, \quad \Pr(\bar{\mathcal{A}}_t) = n^{-\omega(1)}. \tag{14}$$

$d = \frac{M}{v-1}$

This implies Theorem 3.1, since if $\mathcal{A}_T$ holds then (10) and (13) (the latter with $t = T$) give

$$\log\Phi(G_T) = \log|\mathcal{F}_T| > \frac{v-1}{v}n\log n + \frac{mn}{v}\log p - O(n). > 0$$

$\frac{mn}{v}\left[\frac{1}{d}\log n - \log w - \frac{1}{m}\log\log n\right]$

The proof of (14) uses the method of martingales with bounded differences (Azuma's inequality). Here it is natural to consider the martingale

$$X_t = \sum_{i=1}^{t}(\xi_i - \gamma_i)$$

(it is a martingale by (12)), with associated difference sequence

$$Z_i = \xi_i - \gamma_i.$$

$N = \binom{n}{2}$

7a

(*1) $$E\xi_i = \frac{1}{|\mathcal{F}_i|} \cdot \frac{1}{N-i+1} \cdot \sum_e \sum_F 1_{F \ni e}$$

$$= \frac{1}{|\mathcal{F}_i|} \cdot \frac{1}{N-i+1} \cdot \sum_F \sum_e 1_{e \in F}$$

$$= \frac{1}{|\mathcal{F}_i|} \cdot \frac{1}{N-i+1} \cdot |\mathcal{F}_i| \cdot \frac{mn}{v}.$$

Establishing concentration of X_t depends on maintaining some control over the $|Z_i|$'s, for which purpose we will keep track of two sequences of auxiliary events, $\mathcal{B}_i$ and $\mathcal{R}_i$ $(1 \le i \le T-1)$. Informally, $\mathcal{B}_i$ says that no copy of H is in much more than its proper share of the members of $\mathcal{F}_i$, while $\mathcal{R}_i$ includes regularity properties of G_i, together with some limits on the numbers of copies of fragments of H in G_i. The actual definitions of $\mathcal{B}_i$ and $\mathcal{R}_i$ are given in Section 8.

For $i \le T$ it will follow easily from $\mathcal{B}_{i-1}$ and $\mathcal{R}_{i-1}$ (see Lemma 8.2) that

$$\xi_i = o(\log^{-1} n). \tag{15}$$

(The actual bound on ξ_i implied by $\mathcal{B}_{i-1}$ and $\mathcal{R}_{i-1}$ is better when i is not close to T, but for our purposes the difference is unimportant.)

The bound (15), while (more than) sufficient for the concentration we need, can occasionally fail, since the auxiliary properties $\mathcal{B}_{i-1}$ and $\mathcal{R}_{i-1}$ may fail. To handle this problem we slightly modify the preceding X's and Z's by setting

$$Z_i = \begin{cases} \xi_i - \gamma_i & \text{if } \mathcal{B}_j \text{ and } \mathcal{R}_j \text{ hold for all } j < i \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

(and $X_t = \sum_{i=1}^t Z_i$).

As shown in Section 7, a martingale analysis along the lines of Azuma's Inequality then gives (for example)

$$\Pr(|X_t| > n) < n^{-\omega(1)}. \tag{17}$$

Notice that if we do have $\mathcal{B}_i$ and $\mathcal{R}_i$ for $i < t \le T$ (so that $X_t = \sum_{i=1}^t (\xi_i - \gamma_i)$) and $|X_t| \le O(n)$ (it will actually be smaller) then we have $\mathcal{A}_t$, since:

$$\sum_{i=1}^t \xi_i < \sum_{i=1}^t \gamma_i + O(n) < O(n \log n),$$

whence (noting we have (15) for $i \le t$)

$$\sum_{i=1}^t \xi_i^2 < o(\log^{-1} n) \sum_{i=1}^t \xi_i < o(n),$$

so that (using (9))

$$\log |\mathcal{F}_t| > \log |\mathcal{F}_0| - \sum_{i=1}^t (\xi_i + \xi_i^2) > \log |\mathcal{F}_0| - \sum_{i=1}^t \gamma_i - O(n).$$

Thus the first failure (if any) of an $\mathcal{A}_t$ (with $t \le T$) must occur either because X_t is too large or because one of the properties $\mathcal{B}$, $\mathcal{R}$ fails even earlier (that is, $\mathcal{B}_i$ or $\mathcal{R}_i$ fails for some $i < t$). Formally we may write

$$\Pr(\bar{\mathcal{A}}_t) < \sum_{i<t} \Pr(\bar{\mathcal{R}}_i) + \sum_{i \le t} \Pr(\wedge_{j<i}(\mathcal{B}_j \mathcal{R}_j) \wedge \bar{\mathcal{A}}_i) + \sum_{i<t} \Pr(\mathcal{A}_i \mathcal{R}_i \bar{\mathcal{B}}_i).$$

(18) $\qquad A_t \Leftarrow |X_t| < n \qquad$ (19)

Here the second sum is, in view of the preceding discussion and (17), at most $n^{-\omega(1)}$, and we will show, for $i \leq T$,

$$\Pr(\bar{\mathcal{R}}_i) < n^{-\omega(1)} \tag{18}$$

and

$$\Pr(\mathcal{A}_i\mathcal{R}_i\bar{\mathcal{B}}_i) < n^{-\omega(1)}. \tag{19}$$

Combining these three bounds gives (14) and Theorem 3.1. □

The rest of the paper is organized as follows. The next few sections (4-6) fill in some general results, and, as already mentioned, Section 7 gives the calculation needed for (17). In Section 8 we finally define the properties $\mathcal{B}_i$ and $\mathcal{R}_i$, and slightly restate our main inequalities, (18) and (19). Proofs of these—the heart of our argument—are given in Sections 9-11. Extensions to general graphs and hypergraphs are discussed in Section 12. Finally, Section 13 gives the equivalence of Theorems 2.3 and 2.4.

4. Models of random graphs

As noted earlier, we go back and forth between the models $G(n,p)$ and $G(n,M)$.

We recall the definitions, which were already used above: $G(n,p)$ is the random graph on vertex set $[n]$ gotten by taking each pair of vertices to be an edge with probability p, independently of other choices; and $G(n,M)$ is the random graph on $[n]$ gotten by choosing M edges uniformly from $E(K_n)$.

The following observation allows us to switch between the two models. (Note that $\binom{n}{2}p$ is the expected number of edges in $G(n,p)$.)

Lemma 4.1. *Let $n^{\Omega(1)} = M \leq \binom{n}{2}$ be an integer, and $p = M/\binom{n}{2}$.*

(a) *If an event $\mathcal{E}$ holds with probability at least $1-\varepsilon$ in $G(n,p)$, then it holds with probability at least $1-O(n\varepsilon)$ in $G(n,M)$.*

(b) *If an event $\mathcal{E}$ holds with probability at least $1-\varepsilon$ in $G(n,M')$, for all $M/2 \leq M' \leq 2M$, then it holds with probability at least $1-\varepsilon-n^{-\omega(1)}$ in $G(n,p)$.*

Notice that the error terms in the conclusions are not the same as those in the hypotheses. For our purposes the extra factors will not matter, since we will always be dealing with failure probabilities of order $n^{-\omega(1)}$.

Proof of Lemma 4.1. The number of edges in $G(n,p)$ has the binomial distribution $\mathrm{Bin}(\binom{n}{2},p)$. Claim (a) follows from the fact that this number is equal to M with probability $\Omega(1/n)$, and claim (b) from the fact that it is between $M/2$ and $2M$ with probability $1-n^{-\omega(1)}$.

Both these facts are simple properties of the distribution $\mathrm{Bin}(\binom{n}{2},p)$ for p as in the lemma.

5. Concentration

Let $f = f(t_1, \dots, t_n)$ be a polynomial of degree d with real coefficients. We say that f is *normal* if its coefficients are positive numbers and the maximum coefficient is 1. We will be interested in the situation where the t_i's are independent Bernoulli random variables (which will be i.i.d. in our applications, but need not be for the results stated here). Then f is itself a r.v., and we will want to say that, under appropriate conditions, it is likely to be close to its mean.

All polynomials needed for the proof of Theorem 2.4 are multilinear, that is, of the form $f(t) = \sum \alpha_U t_U$, where U ranges over subsets of $[n]$ and $t_U := \prod_{u \in U} t_u$. So for simplicity we confine the present discussion to such polynomials, though the results hold in greater generality.

For a (multilinear) polynomial f as above and $L \subseteq [n]$, the partial derivative (of order $|L|$) with respect to the variables indexed by L is $\sum_{U \supseteq L} \alpha_U t_{U \setminus L}$, and its expectation, denoted E_L or $\mathsf{E}_L f$, is $\sum\{\alpha_U \prod_{i \in U\setminus L} p_i : U \supseteq L\}$, where $t_i \sim \mathrm{Ber}(p_i)$. In particular, when $p_i = p\ \forall i$ we have $\sum\{\alpha_U p^{|U \setminus L|} : U \supseteq L\}$. Set $\mathsf{E}_j f = \max_{|L|=j} \mathsf{E}_L f$.

It may be helpful to observe that when, as here, f is multilinear, we don't really need the polynomial language: We have a weight function α on subsets of $[n]$ and are choosing a random subset, say T, of $[n]$, with membership in T determined by independent coin flips. Then t is the indicator of T; $f(t)$ is the (total) weight of the surviving sets (i.e. those contained in T); and $\mathsf{E}_L f$ is the expected weight of the surviving sets containing L, given that L itself survives.

We begin with a special case of the main result of [14].

Theorem 5.1. *The following holds for any fixed positive integer d and positive constant ε. Let f be a multilinear, homogeneous, normal polynomial of degree d such that $\mathsf{E}f \ge n^{\varepsilon} \max_{1 \le j \le d} \mathsf{E}_j f$. Then*

$$\Pr(|f - \mathsf{E}f| > \varepsilon \mathsf{E}f) = n^{-\omega(1)}.$$

Note that (here and below) d and ε are fixed, and expressions such as $\omega(1)$ are interpreted as $n \to \infty$.

Remark 5.2. Though we have stated the results in this section for normal polynomials, they clearly remain true if we instead assume some fixed bound on coefficients—in this case we say a polynomial is $O(1)$-*normal*—and we will sometimes use them in this form.

When f is normal, $\mathsf{E}_d f = 1$; so in order to use Theorem 5.1 we must have $\mathsf{E}f = \Omega(n^{\varepsilon})$ (if one follows [14] more closely, then $\mathsf{E}f = \omega((\log n)^{2d})$ is sufficient). Thus, the theorem is not applicable when, for example, $\mathsf{E}f$ is on the order of $\log n$, a range that will turn out to be crucial for us. In such ranges the following theorem, which is essentially (the multilinear case of) Theorem 1.2 of [23], is useful.

Theorem 5.3. *The following holds for any fixed positive integer d and positive constant ε. Let f be a multilinear, normal, homogenous polynomial of degree d such that $\mathsf{E}f = \omega(\log n)$ and $\max_{1\le j\le d-1} \mathsf{E}_j f \le n^{-\varepsilon}$. Then*

$$\Pr(|f - \mathsf{E}f| > \varepsilon \mathsf{E}f) = n^{-\omega(1)}.$$

(Note: Theorem 1.2 of [23] asks that $\mathsf{E}f$ be $o(n)$, but this requirement is superfluous: since f is normal, $\mathsf{E}f \le n^d$; so we can add, say, n^{d+1} dummy variables that don't appear in f to guarantee $\mathsf{E}f = o(n)$, and then apply the theorem as stated. Also, to be very picky, the statement in [23] is in finitary form—for any given β the probability is less than $n^{-\beta}$ if $n/Q > \mathsf{E}f > Q\log n$ for a large enough $Q = Q_\beta$—but this clearly implies the present version.)

The key difference between Theorem 5.3 and Theorem 5.1 is that here we only need to consider partial derivatives of order up to $d-1$. While this may seem to be a minor point, the proof of Theorem 5.3 is considerably more involved than that of Theorem 5.1.

Our basic concentration statement is the following combination of Theorems 5.1 and 5.3.

Theorem 5.4. *The following holds for any fixed positive integer d and positive constant ε. Let f be a multilinear, homogeneous, normal polynomial of degree d such that $\mathsf{E}f = \omega(\log n)$ and $\max_{1\le j\le d-1} \mathsf{E}_j f \le n^{-\varepsilon}\mathsf{E}f$. Then*

$$\Pr(|f - \mathsf{E}f| > \varepsilon \mathsf{E}f) = n^{-\omega(1)}.$$

Corollary 5.5. *The following holds for any fixed positive integer d and positive constant ε. Let f be a multilinear, homogeneous, normal polynomial of degree d with $\mathsf{E}f \le A$, where A $(= A(n))$ satisfies*

$$A \ge \omega(\log n) + n^{\varepsilon} \max_{0<j<d} \mathsf{E}_j f,$$

then

$$\Pr(f > (1+\varepsilon)A) \le n^{-\omega(1)}.$$

$Ef \geq \frac{\omega(\log n)}{2}$ ✓
$\max E_j f \leq 2n^{-\epsilon} Ef$ ✓

Proof (sketch). If $\mathsf{E}f \ge A/2$ then this is immediate from Theorem 5.4. Otherwise we can augment f to some $g \ge f$ for which we still have $\mathsf{E}h \le A$ and g satisfies the hypotheses of Theorem 5.4. (For instance, we can do this with $g = g(t,s) = f(t) + h(s)$, where $h(s) = h(s_1,\dots,s_n) = \gamma\sum\{s_U : U \subseteq [n], |U| = d\}$ with the s_i's all Ber(p) (and different from the t_i's), and γ chosen so that $\mathsf{E}h = A/2$.) □

We also need an inhomogeneous version of Theorem 5.4. Write $\mathsf{E}'_L = E'_L f$ for the expectation of the *nonconstant* part of the partial derivative of f with respect to L. Of course for f homogeneous of degree d and $0 < |L| < d$ as in Theorem 5.4, we have $E'_L f = E_L f$.

Theorem 5.6. *The following holds for any fixed positive integer d and positive constant ε. Let f be a multilinear, normal polynomial of degree at most d, with $\mathsf{E}f = \omega(\log n)$ and $\max_{L\ne\emptyset} \mathsf{E}'_L f \le n^{-\varepsilon}\mathsf{E}f$. Then*

$$\Pr(|f - \mathsf{E}f| > \varepsilon \mathsf{E}f) \le n^{-\omega(1)}.$$

5.1: $\max\{1, m_{d+1}\} \leq n^{-\epsilon}\, Ef$

5.3: $m_{d-1} \leq n^{-\epsilon}$ & $Ef = \omega \log n$

5.4: $m_{d-1} \leq n^{-\epsilon}\, Ef$ & $Ef = \omega \log n$

Suppose $Ef = n^{\theta}$

(i) $\theta = \Omega(1)$

$\epsilon' < \min\{\epsilon, \theta\}$

5.1 is true for ϵ' $\Rightarrow$ true (in some sense) $\forall\, \epsilon > \epsilon'$.

(ii) $\theta = o(1)$

$\epsilon' < \epsilon/2$

5.3 is true for ϵ'.

Proof (sketch). We apply Theorem 5.4 to each homogeneous part of f. Say the kth homogeneous part is f_k. If $\mathsf{E}f_k > \mathsf{E}f/2$ then Theorem 5.4 gives $\Pr(|f_k - \mathsf{E}f_k| > (\varepsilon/d)\mathsf{E}f) = n^{-\omega(1)}$ (since $\mathsf{E}f_k = \Theta(\mathsf{E}f)$). Otherwise we can, for example, introduce dummy variables z_{ij} for $1 \le i \le \lceil \mathsf{E}f \rceil$ and $j \in [k]$, each equal to 1 with probability 1, and let $g_k = \sum_i \prod_j z_{ij}$ and $h_k = f_k + g_k$. Theorem 5.4 (applied to h_k, for which we have $\mathsf{E}h_k = \Theta(\mathsf{E}f)$) then gives

$$\Pr(|f_k - \mathsf{E}f_k| > \frac{\varepsilon}{d}\mathsf{E}f) = \Pr(|h_k - \mathsf{E}h_k| > \frac{\varepsilon}{d}\mathsf{E}f) = n^{-\omega(1)}.$$

(The number of variables used for h_k is not exactly n, but cannot be big enough to make any difference.) Consequently, we have

$$\Pr(|f - \mathsf{E}f| > \varepsilon \mathsf{E}f) \le \sum_{k=1}^{d} \Pr(|f_k - \mathsf{E}f_k| > \frac{\varepsilon}{d}\mathsf{E}f) = n^{-\omega(1)}.$$

□

The corresponding generalization of Corollary 5.5 is

Corollary 5.7. *The following holds for any fixed positive integer d and positive constant ε. Let f be a multilinear, normal polynomial of degree at most d with $\mathsf{E}f \le A$, where $A = A(n)$ satisfies*

$$A \ge \omega(\log n) + n^{\varepsilon} \max_{L \ne \emptyset} \mathsf{E}'_L f.$$

Then

$$\Pr(f > (1+\varepsilon)A) = n^{-\omega(1)}.$$

Finally, we will sometimes need to know something when $\mathsf{E}f$ is smaller. Here we cannot expect that f is (w.v.h.p.) close to its mean, but will be able to get by with the following weaker statement, which is (in slightly different language) Corollary 4.9 of [24]. Note that in this case our hypothesis includes $\mathsf{E}_\emptyset f$ $(= \mathsf{E}f)$.

Theorem 5.8. *The following holds for any fixed positive integer d and positive constant ε. Let f be a multilinear, normal polynomial of degree at most d with $\max_L \mathsf{E}'_L f \le n^{-\varepsilon}$. Then for any $\beta(n) = \omega(1)$,*

$$\Pr(f > \beta(n)) = n^{-\omega(1)}.$$

For a random graph $G(n,p)$, we use t_e for the Bernoulli variable representing the appearance of the edge e; thus t_e is 1 with probability p and 0 with probability $1-p$. Recall that for $S \subseteq E(K_n)$, $t_S := \prod_{e \in S} t_e$.

To get a feel for where this is headed, let us briefly discuss a typical use of the preceding theorems. Consider a strictly balanced graph H with v vertices and m edges, and suppose we are interested in the number of copies of H in $G(n,p)$ containing a fixed vertex x_0. This number is naturally expressed as a multilinear, normal, homogeneous polynomial of degree m in the t_e's: $f = \sum_U t_U$, where U runs over edge sets of copies of H in K_n containing x_0. Clearly $\mathsf{E}f = \Theta(n^{v-1}p^m)$.

Now consider a non-empty subset L of $E(K_n)$. The partial derivative $\mathsf{E}_L f$ is $\sum_{U \supseteq L} t_{U \setminus L}$. Let H' be the subgraph of K_n consisting of L and those vertices incident with edges of L, and let v' and m' ($= |L|$) be the numbers of vertices and edges of H'. Then $\mathsf{E}_L f$ is $O(n^{v-v'}p^{m-m'})$ if L covers x_0 and $O(n^{v-v'-1}p^{m-m'})$ if it does not. In either case

$$\mathsf{E}f/\mathsf{E}_L f = \Omega(n^{v'-1}p^{m'}).$$

Since H is strictly balanced, $(v'-1)/m' > (v-1)/m$. This implies that if $\mathsf{E}f = \Omega(1)$ (that is, if $p = \Omega(n^{-(v-1)/m})$ then $\mathsf{E}f/\mathsf{E}_L f \geq n^{\Omega(1)}$, a hypothesis for most of the theorems of this section. (This discussion applies without modification to hypergraphs, since it makes no use of the fact that edge are of size two.)

6. Entropy

We use $H(X) = H_e(X)$ for the base e entropy of a discrete r.v. X; that is,

$$H(X) = \sum_x p(x) \log \frac{1}{p(x)},$$

where $p(x) = \Pr(X = x)$. (For entropy basics see e.g. [8] or [18].)

Given a graph G and $y \in V(G)$, we use $X(y, G)$ for the copy of H containing y in a uniformly chosen H-factor of G, and $h(y, G)$ for $H(X(y, G))$. (We will not need to worry about G's without H-factors.) The next lemma is a special case of a fundamental observation of Shearer [6].

Lemma 6.1. *For any G,*

$$\log \Phi(G) \leq \frac{1}{v} \sum_{y \in V(G)} h(y, G).$$

(To be precise, the statement we are using is more general than what's usually referred to as "Shearer's Lemma," but is what Shearer's proof actually gives; it may be stated as follows. Suppose $Y = (Y_i : i \in I)$ is a random vector and $\mathcal{S}$ a collection of subsets of I (repeats allowed) such that each $i \in I$ belongs to at least t members of $\mathcal{S}$. Then $H(Y) \leq t^{-1} \sum_{S \in \mathcal{S}} H(Y_S)$, where Y_S is the random vector $(Y_i : i \in S)$. To get Lemma 6.1 from this, let Y be the indicator of the random H-factor (so I is the set of copies of H in K_n) and $\mathcal{S} = (S_v : v \in V)$, where S_v is the set of copies of H (in K_n) containing v.)

$\longleftarrow I \longrightarrow$
$Y = (0, 1, 0, \cdots 1, \cdots)$
$H \in Y$

$Y_{S_v} = (0, 0, 1, 0, 0, 0)$
$\downarrow$
$x \in H \in Y$

$H \in S_x$ for $x \in H$

For the next lemma, S is a finite set, $\mathsf{w} : S \to \Re^+$, and X is the random variable taking values in S according to

$$\Pr(X = x) = \mathsf{w}(x)/\mathsf{w}(S)$$

(where for $A \subseteq S$, $\mathsf{w}(A) = \sum_{x \in A} \mathsf{w}(x)$). For perspective recall that for any r.v. X taking values in S, one has $H(X) \leq \log |S|$ (with equality iff X is uniform from S).

Lemma 6.2. *If $H(X) > \log |S| - O(1)$, then there are $a, b \in \text{range}(\mathsf{w})$ with*

$$a \leq b < O(a) \tag{20}$$

$$H(X) = \frac{\omega(L)}{\omega(S)} \sum_{x \in L} \frac{\omega(x)}{\omega(L)} \left[\log \frac{\omega(L)}{\omega(x)} + \log \frac{\omega(S)}{\omega(L)} \right] + \cdots$$
$$\leq \left(\frac{\omega(L)}{\omega(S)} \log |L| + \frac{\omega(L)}{\omega(S)} \log \frac{\omega(S)}{\omega(L)} \right) + \cdots$$

such that for $J = \mathsf{w}^{-1}[a, b]$ *we have*

$$|J| = \Omega(|S|)$$

and

$$\mathsf{w}(J) > .7\mathsf{w}(S).$$

Proof. Let

$$H(X) = \log |S| - K \tag{21}$$

and define C by $\log C = 4(K + \log 3)$. With $\overline{\mathsf{w}} = \mathsf{w}(S)/|S|$, let $a = \overline{\mathsf{w}}/C$, $b = C\overline{\mathsf{w}}$, $L = \mathsf{w}^{-1}([0, a))$, $U = \mathsf{w}^{-1}((b, \infty])$, and $J = S \setminus (L \cup U)$. We have

$$H(X) \leq \log 3 + \frac{\mathsf{w}(L)}{\mathsf{w}(S)} \log |L| + \frac{\mathsf{w}(J)}{\mathsf{w}(S)} \log |J| + \frac{\mathsf{w}(U)}{\mathsf{w}(S)} \log |U|. \tag{22}$$

Then we have a few observations. First, $|U| < |S|/C$ implies that the r.h.s. of (22) is less than

← [redacted]

$$(H(X) = \log|S| - K <) \log 3 + \log |S| - \frac{\mathsf{w}(U)}{\mathsf{w}(S)} \log C,$$

which with (21) implies

$$\mathsf{w}(U) < \frac{K + \log 3}{\log C} \mathsf{w}(S) = \mathsf{w}(S)/4. \tag{23}$$

Of course this also implies $|U| < |S|/4$.

$C \geq 81$

$\omega(L) < \frac{\overline{\omega}}{C}|S|$

$\uparrow$

$\omega(L) = a|S|$

$\Rightarrow \omega(x) = a, \forall a \in S$

$\Rightarrow b = a$ works.

Second, combining (23) with the trivial $\mathsf{w}(L) < \mathsf{w}(S)/C$, we have (say) $\mathsf{w}(J) > .7\mathsf{w}(S)$. But then (third) since the r.h.s. of (22) is at most

$$\log|S| - K \leq \log 3 + \log |S| + \frac{\mathsf{w}(J)}{\mathsf{w}(S)} \log \frac{|J|}{|S|} < \log 3 + \log |S| + \text{[redacted]} \log \frac{|J|}{|S|},$$

we have

$$|J| \geq \exp[-\text{[redacted]}\ 1\ (K + \log 3)]|S| \quad (= \Omega(|S|)).$$

*

7. Martingale

Here we give the proof of (17). As in Section 3, let

$$X_t = Z_1 + \cdots + Z_t,$$

where we use the modified Z_i's of (16). We first prove that

$$\Pr(X_t \geq n) < n^{-\omega(1)}. \tag{24}$$

Notice that Z_i is a function of the random sequence $e_1, \ldots, e_i$, with $\mathsf{E}(Z_i | e_1, \ldots, e_{i-1}) = 0$ for any $e_1, \ldots, e_{i-1}$ and (relaxing (15) a little) $|Z_i| < \varepsilon := \log^{-1} n$. By Markov's inequality, we have, for any positive h,

14a

$$\frac{\omega(L)}{\omega(S)}\log|L| + \frac{\omega(\mathcal{J})}{\omega(S)}\log|\mathcal{J}| + \frac{\omega(U)}{\omega(S)}\log|S| \leqslant \log|S|$$

$$\Pr(X_t \geq n) = \Pr(e^{h(Z_1+\cdots+Z_t)} \geq e^{hn}) \leq \mathsf{E}(e^{h(Z_1+\cdots+Z_t)})e^{-hn}. \qquad (25)$$

Next we bound $\mathsf{E}(e^{h(Z_1+\cdots+Z_t)})$. Since $Z_i = \xi_i - \gamma_i$, $\mathsf{E}(\xi_i|e_1,\dots,e_{i-1}) = \gamma_i$ and $0 \leq \xi_i \leq \varepsilon$, we have (using convexity of e^x),

$$\mathsf{E}(e^{hZ_i}|e_1,\dots,e_{i-1}) \leq e^{-h\gamma_i}((1-\frac{\gamma_i}{\varepsilon}) + \frac{\gamma_i}{\varepsilon}e^{h\varepsilon}).$$ $-\gamma_i \leq Z_i \leq \varepsilon - \gamma_i$

A simple Taylor series calculation shows that the right hand side is at most $e^{h^2\varepsilon\gamma_i}$, for any $0 \leq h \leq 1$. Thus, for such h

$$\mathsf{E}(e^{hZ_i}|e_1,\dots,e_{i-1}) \leq e^{h^2\varepsilon\gamma_i},$$

and induction on t gives

$$\begin{aligned}
\mathsf{E}(e^{h(Z_1+\cdots+Z_t)}) &= \mathsf{E}[\mathsf{E}(e^{h(Z_1+\cdots+Z_t)}|e_1,,\dots,,e_{t-1})] \\
&= \mathsf{E}[e^{h(Z_1+\cdots+Z_{t-1})}\mathsf{E}(e^{hZ_t}|e_1,,\dots,,e_{t-1})] \\
&\leq \mathsf{E}[e^{h(Z_1+\cdots+Z_{t-1})}e^{h^2\varepsilon\gamma_t}] \\
&\leq e^{h^2\varepsilon\sum_{i=1}^t\gamma_i}.
\end{aligned}$$

Inserting this in (25) we have

$$\Pr(X_t \geq n) \leq e^{h^2\varepsilon\sum_{i=1}^t\gamma_i - hn}.$$

Since $\sum_{i=1}^t \gamma_i = O(n\log n)$ and $\varepsilon = \log^{-1} n$, if we set h to be a sufficiently small positive constant, the right hand side is $e^{-\Omega(n)} = n^{-\omega(1)}$, proving (24). (Of course we could do much better—for $\Pr(|X| > \lambda) < n^{-\omega(1)}$, $\lambda > \Omega(\sqrt{n\log n})$ is enough—but this makes no difference for our purposes.) That $\Pr(X_n \leq -n) < n^{-\omega(1)}$ is proved similarly, using $-X_t$ in place of X_t. (Actually we only use the bound on $\Pr(X_n \geq n)$.)

8. The properties $\mathcal{B}$ and $\mathcal{R}$

In this section we define the properties B_i and $\mathcal{R}_i$ and observe that in dealing with (18) and (19) we can work with $G(n,p_i)$ rather than G_i, where

$$p_i = 1 - \frac{i}{\binom{n}{2}}.$$

We will actually define properties $\mathcal{B}$ and $\mathcal{R}(p)$ ($p \in [0,1]$); the events $\mathcal{B}_i$ and $\mathcal{R}_i$ are then

$$\{G_i \text{ satisfies } \mathcal{B}\}$$

and

$$\{G_i \text{ satisfies } \mathcal{R}(p_i)\}.$$

We need some notation. For a finite set A and $\mathsf{w} : A \to \Re^+$ $(:= [0, \infty))$, set

$$\overline{\mathsf{w}}(A) = |A|^{-1} \sum_{a \in A} \mathsf{w}(a),$$

$$\max \mathsf{w}(A) = \max_{a \in A} \mathsf{w}(a),$$

and

$$\operatorname{maxr} \mathsf{w}(A) = \overline{\mathsf{w}}(A)^{-1} \max \mathsf{w}(A)$$

(the maximum normalized value of w), and write $\operatorname{med} \mathsf{w}(A)$ for the median of w.

For a graph G on vertex set V and $Z \in \binom{V}{v}$, let $\mathsf{w}_G(Z) = \Phi(G - Z)$. Thus, $\mathsf{w}_G(Z)$ denotes the number of H-factors in the graph induced by G on the vertex set $V \backslash Z$.

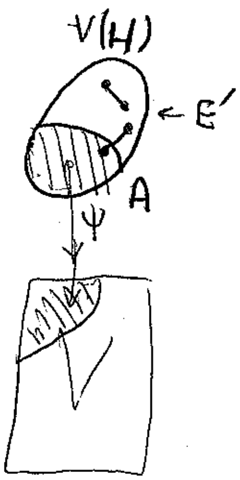
K= copy of H

We also use $\mathsf{w}_G(K) = \mathsf{w}_G(V(K))$ for $K \in \mathcal{H}(G)$.

The property $\mathcal{B}$ for a graph G is

$$\mathcal{B}(G) = \{\operatorname{maxr} \mathsf{w}_G(\mathcal{H}(G)) = O(1)\}. \tag{26}$$

Thus, $\mathcal{B}(G)$ asserts that for any copy K of H in G, the number of H-factors containing K is not much more than the average.



There are two parts to the definition of $\mathcal{R}(p)$. The first refers to the following setup, in which ($V = [n]$ and) expectations refer to $G(n, p)$.

Given $A \subseteq V(H)$, $E' \subseteq E(H) \backslash E(H[A])$, ψ an injection from A to V, and $G \subseteq K_n$, let $X(G)$ be the number of injections $\varphi : V(H) \to V$ with

$$\varphi \equiv \psi \text{ on } A \tag{27}$$

and

$$xy \in E' \Rightarrow \varphi(x)\varphi(y) \in E(G). \tag{28}$$

Write $X(G(n,p))$ in the obvious way as a polynomial in variables $t_e = \mathbf{1}_{\{e \in E(G(n,p))\}}$, $e \in E(K_n)$:

$$X(G(n,p)) = h(t) = \sum_{\varphi} t_{\varphi(E')}, \tag{29}$$

where $t = (t_e : e \in E(K_n))$ and the sum is over injections $\varphi : V(H) \to V$ satisfying (27). Then h is multilinear, $O(1)$-normal and homogeneous of degree $d = |E'|$, and we set

and (28)

$$\mathsf{E}^* = \max\{\mathsf{E}_L h : |L| < d\}. \tag{30}$$

(Note this includes $L = \emptyset$, corresponding to $\mathsf{E}h$.)

For the second part of the definition, we use $D(p)$ for the expected number of copies of H in $G(n, p)$ using a given $x \in V$; that is,

$$D(p) = v(n-1)_{v-1} p^m = \Theta(n^{v-1} p^m).$$

Definition 8.1. *We say $G \subseteq K_n$ satisfies $\mathcal{R}(p)$ if the following two properties hold.*

(a) *For A, E' and ψ (and associated notation) as above: if $\mathsf{E}^* = n^{-\Omega(1)}$, then for any $\beta(n) = \omega(1)$, $X(G) < \beta(n)$ for large enough n; if $\mathsf{E}^* \geq n^{-o(1)}$, then for any fixed $\varepsilon > 0$ and large enough n, $X(G) < n^{\varepsilon}\mathsf{E}^*$.*

(b) *For each $x \in V$, $|D(x,G) - D(p)| = o(D(p))$*

Recall $D(x, G)$ is the number of copies of H (in G) containing x. Thus (b) says that for any x, the number of copies of H containing x is close to its expectation.

We pause for the promised

Lemma 8.2. *For $i \leq T$ (T as in Section 3), $\mathcal{B}_{i-1}$ and $\mathcal{R}_{i-1}$ (that is, $\mathcal{B}$ and $\mathcal{R}(p_{i-1})$ for G_{i-1}) imply* (15).

Proof. Write w for $\mathrm{w}_{G_{i-1}}$. We first observe that $\mathcal{B}$ and $\mathcal{R}_{i-1}$ imply that, for any $K \in \mathcal{H}(G_{i-1})$,

$$\mathrm{w}(K)/\Phi(G_{i-1}) = O(1/D(p_{i-1})) \tag{31}$$

(the left side is the fraction of H-factors in G_{i-1} that use K), since

$$\begin{aligned}\Phi(G_{i-1}) &= \frac{v}{n}\,\mathrm{w}(\mathcal{H}(G_{i-1})) \\ &= \frac{v}{n}\,\Omega(|\mathcal{H}(G_{i-1})| \max \mathrm{w}(\mathcal{H}(G_{i-1}))) \\ &= \Omega(D(p_{i-1})\mathrm{w}(K))\end{aligned}$$

$$\omega(\mathcal{H}(G)) = \sum_{H \in \mathcal{H}(G)} \sum_{F \in \mathcal{F}(G)} 1_{H \in F} = \sum_{F \in \mathcal{F}(G)} \frac{n}{v}$$

(using $\mathcal{B}$ in the second line and part (b) of $\mathcal{R}_{i-1}$ in the third).

Kay

On the other hand, a simple application of part (a) of $\mathcal{R}_{i-1}$ shows that, for any $e \in E(G_{i-1})$, the number of $K \in \mathcal{H}(G_{i-1})$ containing e is at most β for some $\beta = \beta(n)$ satisfying $\beta^{-1}D(p_{i-1}) = \omega(\log n)$. (Here we need the observation that for $i \leq T$, $D(p_{i-1}) = \omega(\log n)$.) Combining this with (31) we have (15). $\square$

$$\xi = \sum_{K \ni e} \frac{w(K)}{\Phi} \overset{o}{\leq} \sum_{K \ni e} \frac{1}{D} \leq \frac{X(e)}{D} \leq \frac{\beta}{D} = o(1/\log n)$$

We also need the "p-version" of $\mathcal{A}_t$:

$$\mathcal{A}(p) = \{\log |\mathcal{F}(G)| > \log |\mathcal{F}_0| - \sum_{i=1}^{t} \gamma_i - O(n)\}, \tag{32}$$

where $t = \lceil (1-p)\binom{n}{2} \rceil$ (and γ_i is as in (11)).

According to Lemma 4.1, (18) and (19)—and thus, as noted at the end of Section 3, Theorem 3.1—will follow from the next two lemmas.

Lemma 8.3. *For $p > \omega(n^{-1/d(H)}(\log n)^{1/m})$,*

$$\Pr(G(n,p) \textit{ satisfies } \mathcal{R}(p)) = 1 - n^{-\omega(1)}.$$

(The assumption on p is only needed for (b) of Definition 8.1.)

17a

(i) $E^* = n^{-\Omega(1)}$: $\chi(G) < \overset{\text{choose}}{\beta} = \omega(1)$ & $\beta^{-1} \overbrace{\omega(\log n)}^{D(P_{i-1})} = \omega(\log n)$

(ii) $E^* = n^{-o(1)}$: $\underbrace{\overbrace{E^* \leq n^{-c} D(P_{i-1})}}_{\text{Simple Random Graph Computation}}$ $\Rightarrow \beta^{-1} D(P_{i-1}) = n^{c} = \omega(\log n)$

Lemma 8.4. *For* $p > \omega(n^{-1/d(H)}(\log n)^{1/m})$,

$$\Pr(G(n,p) \text{ satisfies } \mathcal{A}(p)\mathcal{R}(p)\bar{\mathcal{B}}) = n^{-\omega(1)}.$$

These are proved in Section 9 and Sections 10 and 11 respectively.

9. Regularity

Here we verify Lemma 8.3; that is, we show that w.v.h.p. $G = G(n,p)$ satisfies (a) and (b) of Definition 8.1.

For (a), since there are only $n^{O(1)}$ possibilities for A, E', ψ, it's enough to show that for any one of these the probability of violating (a) is $n^{-\omega(1)}$. This is immediate from the results of Section 5: recalling that each (A, E', ψ) corresponds to a homogeneous, multilinear, $O(1)$-normal polynomial h as in (29) and E^* given by (30), we find that the first part of (a) is given by Theorem 5.8, and the second by Corollary 5.5 with $A = \frac{1}{2}n^{\varepsilon}\mathsf{E}^*$ (and any ε smaller than the present one). in Cor. 5.5

For (b), write $D(x,G)$ in the natural way as a polynomial of degree m (recall $m = |E(H)|$) in the variables $t_e = \mathbf{1}_{\{e \in E(G)\}}$ ($e \in E(K_n)$); namely (with $t = (t_e : e \in E(K_n))$),

$$D(x,G) = f(t) := \sum \{t_K : K \in \mathcal{H}_0(x)\},$$

where $\mathcal{H}_0(x)$ is the set of copies of H in K_n containing x (and, as in Section 5, $t_K = \prod_{e \in K} t_e$).

We have

$$\mathsf{E}f = \Theta(n^{v-1}p^m) = \omega(\log n), \tag{33}$$

and intend to show concentration of f about its mean using Theorem 5.4; thus we need to say something about the expectations $\mathsf{E}_L f$.

For $L \subseteq E(K_n)$ with $1 \le |L| = l < m$ we have

$$\mathsf{E}_L f = p^{m-l}N(L),$$

where $N(L)$ is the number of $K \in \mathcal{H}_0(x)$ with $L \subseteq E(K)$. Let $W = V(L) \cup \{x\}$, with $V(L) \subseteq V$ the set of vertices incident with edges of L, and $v' = |W|$. Then $N(L) = \Theta(n^{v-v'})$ if the graph $H' := (W, L)$ is (isomorphic to) a subgraph of H, and zero otherwise.

Thus, in view of (33), we have

$$\begin{aligned} \mathsf{E}f/\mathsf{E}_L f &= \Omega(n^{v'-1}p^l) = \Omega(n^{[(v'-1)/l-(v-1)/m]l}) \\ &= \Omega(n^{[1/d(H')-1/d(H)]l}) = n^{\Omega(1)} \end{aligned}$$

(where we used strict balance of H for the final equality).

Combining this with (33), we have the hypotheses of Theorem 5.4, so also its conclusion, which is (b).

10. Proof of Lemma 8.4

We now assume p is as in Lemma 8.4 and slightly simplify our notation, using G, $\mathcal{R}$ and $\mathcal{A}$ for $G(n,p)$, $\mathcal{R}(p)$ and $\mathcal{A}(p)$. Events and the notation "Pr" now refer to G; so for instance $\Pr(\mathcal{A})$ is the probability that G satisfies $\mathcal{A}$.

We will get at $\mathcal{B}$ via an auxiliary event $\mathcal{C}$. Write $\mathcal{H}_0$ for the collection of v-subsets of V, and, for $Y \subseteq V$ with $|Y| \le v$, $\mathcal{H}_0(Y)$ for $\{Z \in \mathcal{H}_0 : Z \supseteq Y\}$. We extend the weight function $\mathsf{w} = \mathsf{w}_G$ to such Y by setting

$$\mathsf{w}(Y) = \sum\{\mathsf{w}(Z) : Z \in \mathcal{H}_0(Y)\}$$

Thus $\mathsf{w}(Y)$ is the number of "partial H-factors" of size $\frac{n}{v}-1$ in $G-Y$. The property $\mathcal{C}$ for G is

$$\text{for any } Y \in \tbinom{V}{v-1},\ \underbrace{\max \mathsf{w}(\mathcal{H}_0(Y))}_{\psi(Y)} \le \max\{n^{-2(v-1)}\Phi(G), 2\operatorname{med}\mathsf{w}(\mathcal{H}_0(Y))\} \tag{34}$$

$B \geq n^{-2(v-1)}\Phi$

(where, recall, $\Phi(G)$ is the number of H-factors in G).

Lemma 8.4 follows from the next two lemmas.

$\mathcal{A}\mathcal{R}\bar{\mathcal{B}} \subseteq \mathcal{A}\mathcal{R}\bar{\mathcal{C}} \cup \mathcal{R}\mathcal{C}\bar{\mathcal{B}}$

Lemma 10.1. $\Pr(\mathcal{A}\mathcal{R}\bar{\mathcal{C}}) = n^{-\omega(1)}$

Lemma 10.2. $\Pr(\mathcal{R}\mathcal{C}\bar{\mathcal{B}}) = n^{-\omega(1)}$

The proof of Lemma 10.1, which is really the heart of the matter, is deferred to the next section. Here we deal with Lemma 10.2. We first need to say that $\mathcal{C}$ implies that maxr $\mathsf{w}(\mathcal{H}_0) = O(1)$, or, equivalently, that for some positive constant δ,

$$|\{K \in \mathcal{H}_0 : \mathsf{w}(K) \ge \delta \max \mathsf{w}(\mathcal{H}_0)\}| = \Omega(|\mathcal{H}_0|)\ (= \Omega(n^v)). \tag{35}$$

This is an instance of a simple and general deterministic statement that we may formulate (more precisely than necessary) as follows. Let V be any set of size n and $\mathsf{w} : \binom{V}{v} \to \Re^+$. For $X \subseteq V$ of size at most v write $\mathcal{H}_0(X)$ for the collection of v-subsets of V containing X, and for simplicity set $\psi(X) = \max \mathsf{w}(\mathcal{H}_0(X))$. Let B be a positive number.

Lemma 10.3. *Suppose that for each $Y \subseteq V$ satisfying $|Y| = v-1$ and $\psi(Y) \ge B$ we have*

$$|\{Z \in \mathcal{H}_0(Y) : \mathsf{w}(Z) \ge \frac{1}{2}\psi(Y)\}| \ge \frac{n-v}{2}.$$

Then for any $X \subseteq V$ with $|X| = v-i$ and $\psi(X) \ge 2^{i-1}B$ we have

$$|\{Z \in \mathcal{H}_0(X) : \mathsf{w}(Z) \ge \frac{1}{2^i}\psi(X)\}| \ge \left(\frac{n-v}{2}\right)^i \frac{1}{(i-1)!}. \tag{36}$$

Proof. Write N_i for the r.h.s. of (36). We proceed by induction on i, with the case $i = 1$ given. Assume X is as in the statement and choose $Z \in \mathcal{H}_0(X)$ with $\mathsf{w}(Z)$ maximum (i.e. $\mathsf{w}(Z) = \psi(X)$). Let $y \in Z \setminus X$ and $Y = X \cup \{y\}$. Then $|Y| = v-(i-1)$ and $\psi(Y) = \psi(X) \ge 2^{i-1}B$ $(\ge 2^{i-2}B)$; so by our induction hypothesis there are at least N_{i-1} sets $Z' \in \mathcal{H}_0(Y)$ with $\mathsf{w}(Z') \ge 2^{-(i-1)}\psi(Y)$ $(= 2^{-(i-1)}\psi(X))$. For

$\geq B$

each such Z', $Z' \setminus \{y\}$ is a $(v-1)$-subset of V with $\psi(Z' \setminus \{y\}) \geq \mathsf{w}(Z') \geq B$. So (again, for each such Z') there are at least $(n-v)/2$ sets $Z'' \in \mathcal{H}_0(Z' \setminus \{y\})$ with

$$\mathsf{w}(Z'') \geq \psi(Z' \setminus \{y\})/2 \geq 2^{-i}\psi(X).$$

The number of these pairs (Z', Z'') is thus at least $N_{i-1}(n-v)/2$. On the other hand, each Z' associated with a given Z'' is $Z'' \setminus \{u\} \cup \{y\}$ for some $u \in Z'' \setminus (X \cup \{y\})$; so the number of such Z' is at most $i-1$ and the lemma follows. □

Proof of Lemma 10.2. Set $\gamma = [2^{v+1}(v-1)!]^{-1}$ and $\delta = 2^{-v}$. Now $\mathcal{C}$ implies the hypothesis of Lemma 10.3 with $B = (2n)^{-(v-1)}\Phi(G)$ (actually with any B greater than $n^{-2(v-1)}$), and we have (trivially)

$\psi(Y) > n^{-2(v-1)}\Phi \Rightarrow$ med $(\mathsf{w}(\mathcal{H}_0(Y))) \geq \frac{\psi(Y)}{2}$

$$\psi(\emptyset) \geq n^{-(v-1)}\Phi(G) = 2^{v-1}B.$$

Thus (36) applies, yielding (35), now with constants specified:

$$|\{K \in \mathcal{H}_0 : \mathsf{w}(K) \geq \delta \underbrace{\max \mathsf{w}(\mathcal{H}_0)}_{\psi(\emptyset)}\}| > \gamma n^v.$$

Let J be the largest power of 2 not exceeding $\max \mathsf{w}$ and

$$\mathcal{Z} = \{Z \in \mathcal{H}_0 : \mathsf{w}(Z) > \delta J\}.$$

For $X \subseteq V$ with $|X| \leq v$ let $\mathcal{Z}(X) = \{Z \in \mathcal{Z} : X \subseteq Z\}$, and say $X \subseteq V$ with $|X| \leq v$ is *good* if $|\mathcal{Z}(X)| > \gamma n^{v-|X|}$. Thus in particular,

$$\textit{the empty set is good} \tag{37}$$

(as are all singletons, but we don't need this).

Fix an ordering $a_1, \ldots, a_v$ of $V(H)$. For distinct $x_1, \ldots, x_r \in V$, write $S(x_1, \ldots, x_r)$ for the collection of copies φ of H in K_n for which

$a_1, a_2 \cdots a_{r-1} \; a_r \cdots a_v$
$x_1 \; x_2 \; x_{r-1}$
$\in \mathcal{Z}$

$$\varphi(a_i) = x_i \text{ for } i \in [r],$$

$$\varphi(a_i)\varphi(a_j) \in E(G) \text{ whenever } i, j \geq r \text{ and } a_ia_j \in E(H),$$

and $\varphi(V(H)) \in \mathcal{Z}$.

For $r \in \{0, \ldots, v\}$ let $N_r = N(a_r) \cap \{a_{r+1}, \ldots, a_v\}$ and $d_r = |N_r|$. (In particular $d_v = 0$.) Let $\mathcal{Y}(x_1, \ldots, x_r)$ be the event

$$\{|S(x_1, \ldots, x_r)| = \Omega(p^{d_r + \cdots + d_{v-1}}n^{v-r})\}. \quad \Theta[E(\#\text{ of such } \phi)]$$

In particular

$$S(\emptyset) = \{\varphi \in \mathcal{H}(G) : \mathsf{w}(\varphi(V(H))) > \delta J\}$$

and $\mathcal{Y}(\emptyset)$ is the event

$$\{|S(\emptyset)| = \Omega(p^m n^v)\}. \quad = \Omega(|\mathcal{H}(G)|) \text{ whp}$$

For (distinct) $x_1, \ldots, x_r \in V$ let $Q(x_1, \ldots, x_r)$ be the event

$$\{\{x_1, \ldots, x_r\} \text{ is good}\} \wedge \overline{\mathcal{Y}}(x_1, \ldots, x_r).$$

Since $\bar{\mathcal{B}}\mathcal{C} \subseteq Q(\emptyset)$ (by (37)), Lemma 10.2 will follow if we can show

$$\Pr(\mathcal{R}Q(\emptyset)) = n^{-\omega(1)}.$$

20a

$$\psi(\emptyset) = \max_{|Y|=v} \Phi(V-Y)$$

$$\frac{n}{v}\Phi = \#[F, Y\in F] = \sum_{Y} \Phi(V-Y)$$

$$\le \binom{n}{v}\psi(\emptyset)$$

$$\mathcal{B}_\epsilon = |\{K\in \mathcal{H}(G), w(K) > \epsilon \max w\}| = \Omega(|\mathcal{H}(G)|)$$

$$\mathcal{B} = \exists \epsilon : \mathcal{B}_\epsilon$$

$$B\in\mathcal{B}_{\delta/2} \Leftarrow \underbrace{|\mathcal{H}(G)| = \Omega(n^v p^n)}_{\mathcal{E}} \quad \cancel{A} \quad \mathcal{Y}(\emptyset)$$

$$\overline{\mathcal{B}}\,\mathcal{C} \subseteq [\overline{\mathcal{Y}} \vee \overline{\mathcal{E}}]\,\mathcal{C} \subseteq [\overline{\mathcal{Y}} \vee \overline{\mathcal{E}}]\,\mathcal{G}(\emptyset)$$

$$\subseteq \mathcal{Q}(\emptyset) \vee \overline{\mathcal{E}}$$

For inductive purposes we will actually prove the more general statement that for any r and $x_1, \dots, x_r$,

$$\Pr(\mathcal{R}Q(x_1, \dots, x_r)) = n^{-\omega(1)}. \tag{38}$$

Proof of (38). We proceed by induction on $v-r$. The case $r = v$ being trivial (since for v-subsets of V being good is the same as belonging to $\mathcal{Z}$), we consider $r < v$, letting $X = \{x_1, \dots, x_r\}$. $K = \{x_1, \dots, x_n\}$ good $\Rightarrow K \in \mathcal{Z} \Rightarrow S(K) \neq \emptyset \Rightarrow \mathcal{Y}(K)$

Let $\mathcal{P}$ be the event

$$\{y \in V \setminus X, X \cup \{y\} \text{ good } \Rightarrow \mathcal{Y}(x_1, \dots, x_r, y)\}.$$

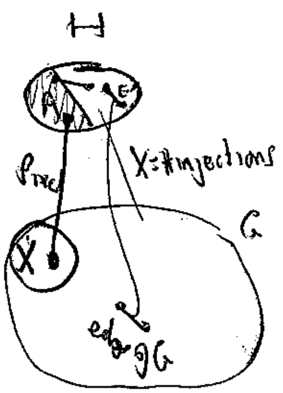
By inductive hypothesis it's enough to show

$$\Pr(\mathcal{R}\mathcal{P}Q(x_1, \dots, x_r)) < n^{-\omega(1)} \tag{39}$$

(since $\Pr(\mathcal{R}Q) \le \Pr(\mathcal{R}\mathcal{P}Q) + \Pr(\mathcal{R}\overline{\mathcal{P}})$ and, by induction, $\Pr(\mathcal{R}\overline{\mathcal{P}}) = n^{-\omega(1)}$). $\leftarrow \overline{\mathcal{P}} = \exists y$: $X+y$ good $\neg\mathcal{Y}(X+y)$

Note that if X is good then

$$|\{y : X \cup \{y\} \text{ good}\}| = \Omega(n). \tag{40}$$



We need to slightly relax $\mathcal{R}$ to arrange that the edges between x_r and $V \setminus X$ are independent of our conditioning. Say G satisfies $\mathcal{R}_X$ if it satisfies (a) in the definition of $\mathcal{R}$ $(= \mathcal{R}(p))$ whenever $A = \{a_1, \dots, a_r\}$, $\psi(a_i) = x_i$ $(i \in [r])$ and $E' \subseteq E(H - A)$. [as opposed to $E' \subseteq E / E[H[A]]$]

Q

If $\mathcal{R}\mathcal{P} \wedge \overbrace{\{X \text{ good}\} \text{ holds but } \mathcal{Y}(x_1, \dots, x_r)}$ does not, then we have the following situation. There is some $J = 2^k$ with k an integer not exceeding $n \log n$ (see (10)) so that (with $\mathcal{Z}$, "good" and other quantities as in the preceding discussion, but now defined in terms of this J)

(i) $\mathcal{R}_X$ holds;

(ii) there are at least $\Omega(n)$ y's in $V \setminus X$ for which we have $\mathcal{Y}(x_1, \dots, x_r, y)$ (by (40)); but

(iii) $\mathcal{Y}(x_1, \dots, x_r)$ does not hold.

Note that, for a given J, properties (i) and (ii) depend only on $G' := G - X$. Since the number of possibilities for J is at most $n \log n$, it is thus enough to show that for any J and G' satisfying (i) and (ii) (with respect to J),

$$\Pr(\overline{\mathcal{Y}}(x_1, \dots, x_r) | G') = n^{-\omega(1)}. \tag{41}$$

Given such a G' (actually, any G'), $|S(x_1, \dots, x_r)|$ is naturally expressed as a multilinear polynomial in the variables

$$t_u := \mathbf{1}_{\{x_r u \in E(G)\}} \quad u \in V \setminus X;$$

namely

$$|S(x_1, \dots, x_r)| = g(t) := \sum_U \alpha_U t_U,$$

G' fixed
Only edges
$x_r u$ are
undetermined at
this point.
(edges that make
edges $\subseteq X$ do not
influence W)

where U ranges over d_r-subsets of $V \setminus X$, and α_U is the number of copies ψ of $K := H - \{a_1, \ldots, a_r\}$ in G' with

$$\psi(N_r) = U$$

and

$$\psi(\{a_{r+1}, \ldots, a_v\}) \cup X \in \mathcal{Z}. \tag{42}$$

In order to apply Theorem 5.4 we normalize and consider

$$f(t) = \alpha^{-1} g(t),$$

with α the maximum of the α_U's. We then need the hypotheses of Theorem 5.4; these will follow (eventually) from $\mathcal{R}_X$.

We may rewrite

$$g(t) = \sum_{y \in V \setminus X} \sum \{t_{\varphi(N_r)} : \varphi \in S(x_1, \ldots, x_r, y)\}.$$

Noting again that the variables t_u ($u \in V \setminus X$) are independent of G' (which determines the sets $S(x_1, \ldots, x_r, y)$), and using property (ii), we have

Now we have removed influence of x_r

$$\mathsf{E}g = p^{d_r} \sum_{y \in V \setminus X} |S(x_1, \ldots, x_r, y)| = \Omega(p^{d_r + \cdots + d_{v-1}} n^{v-r}). \tag{43}$$

Note that if $d_r = 0$ then there is nothing random at this stage and we have

$$|S(x_1, \ldots, x_r)| = \sum_{y \in V \setminus X} |S(x_1, \ldots, x_r, y)| = \Omega(p^{d_r + \cdots + d_{v-1}} n^{v-r}),$$

which is what we want; so we may assume from now on that $d_r > 0$.

If we set $H' = H - \{a_1, \ldots, a_{r-1}\}$ then $d_r + \cdots + d_{v-1} = e(H')$ and $v - r = v(H') - 1$, so that the r.h.s. of (43) is $\Omega(p^{e(H')} n^{v(H')-1})$. Thus, since H is strictly balanced, we have

$$\mathsf{E}g = \begin{cases} \omega(\log n) & \text{if } r = 1 \\ n^{\Omega(1)} & \text{if } r > 1. \end{cases} \tag{44}$$

We are going to prove that

$$\mathsf{E}f = \omega(\log n) \tag{45}$$

f

(in most cases it will be $n^{\Omega(1)}$) and

$$\max\{\mathsf{E}_T f : T \subseteq V \setminus X, 0 < |T| < d_r\} = n^{-\Omega(1)} \mathsf{E}f. \tag{46}$$

If we have these then Theorem 5.4 says that w.v.h.p. f and (therefore) g are close to their expectations, which in view of (43) is what we want.

For the proofs of (45) and (46), it will be more convenient to work with the partial derivatives of g than with those of f. As in Section 5 we use $t_e = \mathbf{1}_{\{e \in E(G)\}}$, $t_S = \prod_{e \in S} t_e$ and

$$t = (t_e : e \in E(K_{V \setminus X})) \tag{47}$$

(where, recall, $K_{V \setminus X}$ is the complete graph on $V \setminus X$).

Let $T \subseteq V \setminus X$ with $0 < l := |T| \leq d_r$. (Note we now include $l = d_r$, in which case $\mathsf{E}_T g$ is just α_T.)

Since we are only interested in upper bounds on the partial derivatives of g, we may now disregard the requirement (42); thus we use

$$p^{-(d_r - l)}\mathsf{E}_T g \leq h(t) := \sum_{\varphi} t_{\varphi(E(K))}, \tag{48}$$

where the sum is over injections

$K = H - \{a_1, \dots, a_r\}$

$$\varphi : V(K) \to V \setminus X \text{ with } \varphi(N_r) \supseteq T \tag{49}$$

(and, again, with the obvious meaning for $\varphi(E(K))$).

Set $\mathsf{E}^* = \max\{\mathsf{E}_L h : L \subseteq E(K_{V\setminus X}), |L| < |E(K)|\}$. We assert that there is a positive constant ε (depending only on H) so that (for large enough n),

$$p^{d_r - l}\mathsf{E}^* < n^{-\varepsilon}\mathsf{E}g. \tag{50}$$

(Of course we have already chosen G', but we now need to reexamine the randomization that produced it to see what $\mathcal{R}_X$ says about the $\mathsf{E}_T g$'s.) Removing conditioning on G'

Before proving (50), we show that it gives (45) and (46). For (45) we need

$$\alpha^{-1}\mathsf{E}g = \omega(\log n).$$

We apply(50) with $T = U$, a d_r-subset of $V \setminus X$ (in which case, as noted above, α_U is just $\mathsf{E}_U g$). We consider two possibilities. If $\mathsf{E}g \geq n^{\varepsilon/2}$ then $\mathcal{R}_X$ gives (note here $d_r - l = 0$)

$\alpha_U \equiv X(G)$ of Def: 8.1

$$\alpha_U < n^{\varepsilon/4}\max\{1, \mathsf{E}^*\} \leq n^{-\varepsilon/4}\mathsf{E}g.$$

$\exists \epsilon$

$\boxed{X(G) \leq n^{\epsilon} E^* \ \forall \epsilon}$

Otherwise we have $\mathsf{E}^* < n^{-\varepsilon/2}$. In this case, recalling that $\mathsf{E}g = \omega(\log n)$ (see (44)), we can choose $\beta(n) = \omega(1)$ with $\beta(n)^{-1}\mathsf{E}g = \omega(\log n)$, and $\mathcal{R}_X$ guarantees that $\alpha_U \leq \beta(n)$. So in either case we have $\alpha^{-1}\mathsf{E}g = \omega(\log n)$. Since none of these bounds depended on the choice of U, this gives (45).

For (46) we need $\mathsf{E}_T g = n^{-\Omega(1)}\mathsf{E}g$ for any T as in (46). This follows from $\mathcal{R}_X$ if we assume, as we may, that the constant ε in (50) is less than $1/d(H)$, since we then have, e.g.,

$$E_T g < \underbrace{p^{d_r - l} n^{\varepsilon/2}\max\{1, \mathsf{E}^*\}}_{\mathcal{R}_X} \overset{(50)}{\leq} n^{-\varepsilon/2}\mathsf{E}g.$$

For the proof of (50), fix $L \subseteq E(K_{V\setminus X})$ and let $k = |E(K)| - |L|$. We have $\mathsf{E}_L h = p^k N_L$, where N_L is the number of φ as in (49) with $\varphi(E(K)) \supseteq L$. In particular each such φ satisfies $\varphi(V(K)) \supseteq W := T \cup V(L)$, where, as earlier, $V(L) \subseteq V \setminus X$ is the set of vertices incident with edges of L. Letting $W = \{w_1, \dots, w_s\}$, we have $N_L = \sum N_L(b_1, \dots, b_s)$, where $(b_1, \dots, b_s)$ ranges over s-tuples of distinct elements of $V(K)$ and the summand is the number of φ's as above with $\varphi(b_i) = w_i$ for $i \in [s]$. Since there are only $O(1)$ choices for the b_i's, we will have (50) if we show (for any choice of b_i's)

$$p^{d_r - l + k} N_L(b_1, \dots, b_s) = n^{-\Omega(1)}\mathsf{E}g. \tag{51}$$

Let (given b_i's) $H'' = H[\{a_r, b_1, \ldots, b_s\}]$. Then

$$N_L(b_1, \ldots, b_s) < n^{v-r-s} = n^{v-r-(v(H'')-1)},$$

while $k \geq l + d_{r+1} + \cdots + d_{v-1} - e(H'')$ (since $|E(K)| = d_{r+1} + \cdots + d_{v-1}$ and $E(H'')$ contains $\varphi^{-1}(L)$ and at least l edges joining a_r to $V(K)$). Thus

$$p^{d_r - l + k} N_L(b_1, \ldots, b_s) < p^{d_r + \cdots + d_{v-1}} n^{v-r} [n^{v(H'')-1} p^{e(H'')}]^{-1}.$$

Since H is strictly balanced (and since $v(H'') - 1 \geq l > 0$), the expression in square brackets is $n^{\Omega(1)}$, so that, in view of (43), we have (51).

This completes the proof of Lemma 10.2.

11. Proof of Lemma 10.1

We continue to use G, $\mathcal{R}$ and $\mathcal{A}$ as in the preceding section.

Assume that we have $\mathcal{A}$ and $\mathcal{R}$ and that $\mathcal{C}$ fails at Y (see (34)). Let $R = Y \cup \{x\} \in \mathcal{H}_0(Y)$ satisfy $\mathsf{w}(R) = \max \mathsf{w}(\mathcal{H}_0(Y))$. Note that $|Y| = v-1$

$$\mathsf{w}(R) > n^{-2(v-1)} \Phi(G). \tag{52}$$

$h(y,G) = H(X(y,G))$ entropy of copy of H ∋ y in random H-factor

Choose $y \in V \setminus Y$ with $\mathsf{w}(Y \cup \{y\}) \leq \operatorname{med} \mathsf{w}(\mathcal{H}_0(Y))$, and with $h(y, G - R)$ maximum subject to this restriction (see Section 6 for h), and set $S = Y \cup \{y\}$. Thus in particular

$$\mathsf{w}(R) > 2\mathsf{w}(S). \quad \geq 2\operatorname{med}(S)$$

Note that the lower bound in (32) is now (13)

$$(10) \quad \frac{v-1}{v} n \log n + \frac{mn}{v} \log p - O(n)$$

(see (10) and (13)), while $D = D(p)$ (as in Definition 8.1) satisfies

$$\log D = (v-1) \log n + m \log p - O(1).$$

Thus $\mathcal{A}$ and (52) give

$$(*) \qquad \log \Phi((G - R) \geq \frac{v-1}{v} n \log n + \frac{mn}{v} \log p - O(n), \qquad \log \Phi(G-R) \geq \log(w(R))$$

while from $\mathcal{R}$ we have, using Lemma 6.1,

$$(**) \qquad \log \Phi(G - R) \leq \frac{n}{2v} [\log((1 + o(1))D) + h(y, G - R)].$$

(In more detail: Lemma 6.1 gives $\log \Phi(G - R) \leq \frac{\overset{1}{\cancel{n}}}{v} \sum_{z \in V \setminus R} h(z, G - R)$; and we have $h(z, G - R) \leq h(y, G - R)$ for at least half the z's in $V \setminus R$, and $h(z, G - R) \leq \log D(z, G - R) < \log((1 + o(1))D)$ for *every* z, using $\mathcal{R}$ and the fact that for any random variable κ, $H(\kappa) \leq \log |\text{range}(\kappa)|$.) (b)

$w'(Z) = \#$ of H-factors in $G - [Y + Z + \{y, z\}]$

Combining (and rearranging) we have (again using $\mathcal{R}$)

$$h(y, G-R) > (v-1)\log n + m\log p - O(1) > \log D_{G-R}(y) - O(1). \tag{53}$$

Let $W = V \setminus (Y \cup \{x,y\})$ and for $Z \in \binom{W}{v-1}$ set

$$\mathsf{w}'(Z) = \Phi(G - ([redacted] \cup Z)).$$ $Y \cup \{x,y\}$

p6

p18

The corresponding weight functions w_y on $\mathcal{H}(y, G-R)$ and w_x on $\mathcal{H}(x, G-S)$ are

$$\mathsf{w}_y(K) = \mathsf{w}'(V(K)\setminus\{y\}) \ \text{ and } \ \mathsf{w}_x(K) = \mathsf{w}'(V(K)\setminus\{x\}).$$

Thus $X(y, G-R)$ (see Section 6) is chosen according to the weights w_y, and similarly for $X(x, G-S)$. Note also that $\mathsf{w}_y(\mathcal{H}(y, G-R)) = \mathsf{w}(R)$ and $\mathsf{w}_x(\mathcal{H}(x, G-S)) = \mathsf{w}(S)$.

$* \sum_{K \ni x,\ K\cap S = \emptyset} \mathsf{w}_x(K)$ [redacted] $= \sum_{K\ni x,\ K \cap S = \emptyset} \Phi[G - K - S] = \Phi(G-S)$

According to Lemma 6.2, (53) implies that there are $a, b \in \mathrm{range}(\mathsf{w}_y)$ $(\subseteq \mathrm{range}(\mathsf{w}'))$? as in (20) for which $J := \mathsf{w}_y^{-1}([a,b])$ satisfies

$J = \{K \ni y,\ K \cap R = \emptyset,\ \mathsf{w}_y(K) \in [a,b]\}$

$$|J| > \Omega(|\mathcal{H}(y, G-R)|) \tag{54}$$

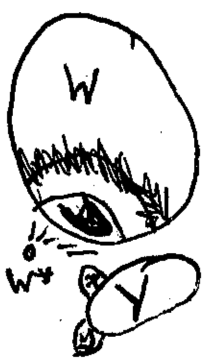
and

$$\mathsf{w}_y(J) > .7\mathsf{w}_y(\mathcal{H}(y, G-R)) = .7\mathsf{w}(R).$$

w_y

Setting $J' = \mathsf{w}_x^{-1}([a,b])$, we thus have

$$\mathsf{w}_y(J) > .7\mathsf{w}(R), \tag{55}$$



while

$$\mathsf{w}_x(J') \le \mathsf{w}(S) < .5\mathsf{w}(R). \tag{56}$$

$* \le \Phi(G - S)$

Once we condition on the value of $G[W]$, $\mathsf{w}_y(J)$ and $\mathsf{w}_x(J')$ are naturally expressed as evaluations of a multilinear polynomial in variables $\{t_u : u \in W\}$, as follows. Given $U \subseteq W$ with $|U| \le v-1$, let G_U^* be the graph obtained from $G[W]$ by adjoining a vertex w^* with neighborhood U. Let $\mathcal{K}_U$ be the set of copies of H in G_U^* containing $\{w^*u : u \in U\}$, and

$K \in \mathcal{K}_U$

$$\alpha_U = \sum\{\mathsf{w}'(V(K)\setminus\{w^*\}) : K \in \mathcal{K}_U, \mathsf{w}'(\overbrace{V(K)\setminus\{w^*\}}^{A}) \in [a,b]\}. \tag{57}$$

The desired polynomial is then

$\alpha_U = \sum_A \#$H-factors in $G - [A + Y + \{x,y\}]$

$$g(t) = \sum_{U \subseteq W} \alpha_U t_U,$$

w^* surrogate for x and y

and $\mathsf{w}_y(J)$ and $\mathsf{w}_x(J')$ are g evaluated at $t' := \mathbf{1}_{\{z\in W : yz \in E(G)\}}$ and $t'' := \mathbf{1}_{\{z \in W : xz \in E(G)\}}$.

Now $\mathcal{R}$ implies $|\mathcal{H}(y, G-R)| = \Theta(n^{v-1}p^m) = \omega(\log n)$. (In more detail: $\mathcal{R}$ implies that $D_G(y) = \Theta(n^{v-1}p^m)$ and, as is easily seen, that the number of copies of H in G containing y and meeting R is $o(n^{v-1}p^m)$.) Thus by (54) we also have

$$|J| = \Theta(n^{v-1}p^m),$$

(*)

and (since $b < O(a)$)

$$\mathsf{w}_y(J) = \Theta(bn^{v-1}p^m). \tag{58}$$

$(b - a)$

(#) Fix $|K| = v$.

25a

$$E(\overbrace{\#\text{ copies of } H \text{ that meet } K \text{ and contain } y}^{Z})$$

$$= O(n^{v-2}p^m) = O\left(\frac{\omega}{n}\right) \qquad \omega = \log n \text{ say}$$

$$\Pr[\# \text{ is not } o(n^{v-1}p^m)] \le n^{-\text{any constant}} \qquad \text{— use Kim Vu (Th. 5.8)}$$

~~$E(Z^k) = O\left(\frac{\omega^k}{n^k}\right)$ for $k = O(1)$~~

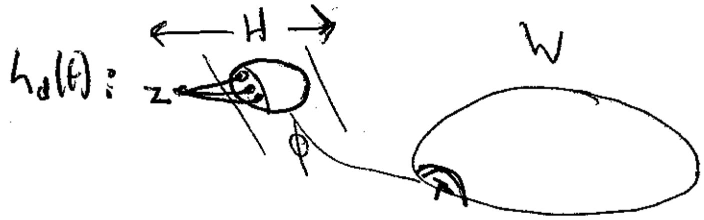
$$E(\# \mid E') = n^{v - v'} p^{m - E'}$$

$$= (n^{v-1}p^m)\, n^{-(v'-1) + E'(v-1)/m}$$

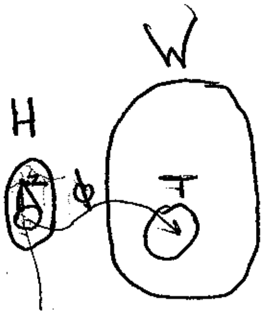
$\uparrow \omega \qquad \uparrow n^{-\epsilon}$

$$W_y(J) = \sum_{\substack{K \ni y, \\ K \text{ is copy of } H \text{ in } G-R, \\ \Theta_K \in [a,b]}} \overbrace{\# H\text{-factors in } G - [K+R]}^{\Theta_K}$$

$$= \sum_{U} \sum_{\substack{i \\ N_G(y,K) = U}} \Theta_K$$



The key idea is now that we can use Corollary 5.7 to deduce that, unless something unlikely has occurred, $\mathsf{E}g$ must be of a similar size.



Fix $T \subseteq W$, say with $|T| = l < v$. For $d = l, \dots, v-1$, and t as in (47) with $V \setminus X$ replaced by W, we consider the polynomial

$t = (t_e : e \in K_W)$

$$h_d(t) = \sum_z \sum_\varphi t_{\varphi(E(H-z))}, \qquad \boxed{G[W] \text{ fixed now}} \tag{59}$$

(or as R.V.)

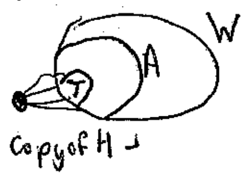
where z ranges over vertices of H of degree d and φ over injections $V(H) \setminus \{z\} \to W$ with $\varphi(N_z) \supseteq T$. Then

$$\alpha_T \leq b \cdot h_l(t)$$

$$\alpha_T = \sum_A \underbrace{\Phi(G - [A + Y + \{x,y\}])}_{\in [a,b]}$$

("$\leq$" because of the restriction $\mathrm{w}'(V(K) \setminus \{w^*\}) \in [a,b]$ in (57)) and

$$\mathsf{E}'_T g \leq b \sum_{d>l} p^{d-l} h_d(t).$$



Let

$$\mathsf{E}^*_d = \max\{\mathsf{E}_L h_d : L \subseteq E(K_W), |L| < m - d\}.$$

We assert that there is a positive constant ε (depending only on H) so that (for each d),

$$p^{d-l} \mathsf{E}^*_d < n^{-\varepsilon} n^{v-1} p^m. \tag{60}$$

The proof of this is essentially identical to that of (50) and we omit it. (Actually (60) is contained in (50): it's enough to prove (60) with E^*_d replaced by $\mathsf{E}^*(z)$ gotten by replacing h_d in the definition of E^*_d by the inner sum in (59); but this inner sum is bounded by the polynomial $h(t)$ in (48) with $r = 1$, $a_1 = z$ and $d_r = d$ ("bounded by" rather than "equal to" because we've replaced $V \setminus \{x_1\}$ by the slightly smaller W); and, finally, note that we actually proved (50) with $\mathsf{E}g$ replaced by $\Omega(p^{d_r + \cdots + d_{v-1}} n^{v-r})$ (see (43).)

Now intending to apply the results of Section 5, we consider $f = \alpha^{-1} g$, where $\alpha = \max_U \alpha_U$. Then

$$f(t') = \alpha^{-1} \mathrm{w}_y(J) = \Theta(\alpha^{-1} b n^{v-1} p^m)$$

(see (58)), and we have

$$f(t') = \omega(\log n) \tag{61}$$

and

$$\max\{\mathsf{E}'_T f : T \subseteq W, T \neq \emptyset\} = n^{-\Omega(1)} f(t'). \tag{62}$$

These are derived from (60) in the same way as (45) and (46) were derived from (50), and we will not repeat the arguments.

In summary, $\overline{\mathcal{ARC}}$ implies that there are Y, x, y and (with notation as above) $a, b \in \text{range}(\mathrm{w}')$ for which we have (61), (62), and, from (55) and (56),

$$f(t'') < .8 f(t'). \tag{63}$$

But for given Y, x, y, a and b, f depends only on $G[W]$. On the other hand, given $G[W]$, t and t' are independent r.v.'s, each with law $\text{Bin}(W,p)$ (where we say $t = (t_w : w \in W)$ has "law $\text{Bin}(W,p)$" if the t_w's are independent, mean p Bernoullis). Thus the following simple consequence of Theorem 5.6 and Corollary 5.7 applies.

Claim. *For any $\varepsilon > 0$ and d the following holds. If f is a multilinear, normal polynomial of degree at most d in n variables, $\zeta(n) = \omega(\log n)$, and t', t'' are independent, each with law $\text{Bin}([n],p)$, then*

$$\Pr(f(t') > \max\{\zeta(n), n^{\varepsilon} \max_{T \neq \emptyset} \mathsf{E}'_T f, (1+\varepsilon) f(t'')\}) = n^{-\omega(1)}.$$

Since there are only polynomially many possibilities for Y, x, y, a and b, this gives Lemma 10.1.

Proof of Claim. Set

$$A = \frac{1}{2} \max\{\zeta(n), n^{\varepsilon} \max_{T \neq \emptyset} \mathsf{E}'_T f\}.$$

If $\mathsf{E}f \leq A$ then Corollary 5.7 gives

$$\Pr(f(t') > A) = n^{-\omega(1)};$$

otherwise, by Theorem 5.6,

$$\begin{aligned} \Pr(f(t') > (1+\varepsilon) f(t'')) &< \Pr(\max\{|f(t') - \mathsf{E}f|, |f(t'') - \mathsf{E}f|\} > (\varepsilon/3)\mathsf{E}f) \\ &= n^{-\omega(1)}. \end{aligned}$$

12. Extensions

As mentioned earlier, we will not repeat the above arguments for general graphs or for hypergraphs; but we do want to stress here that "repeat" is the right word: extending the proof of Theorem 2.4 to produce (the counting versions of) Theorems 2.2, 2.5 and 2.7 involves nothing but some minor formal changes. To elaborate slightly:

While strict balance is used repeatedly in the proof of Theorem 2.4, a little thought shows that all these uses are of the same type: to allow us to say that, for some proper subgraph H' of H, and p in the range under consideration,

$$n^{v(H')-1} p^{e(H')} = n^{\Omega(1)}, \tag{64}$$

as follows from

$$e(H')/(vH') - 1) < m/(v-1) \tag{65}$$

(this is what we get from strict balance) and the fact that $p \geq n^{-1/d(H)}$ (of course it is actually somewhat bigger). See the last few paragraphs of Section 5.

For general graphs we no longer have (65) but can still guarantee (64)—though, of course, at the cost of some precision in the results—by taking $p \geq n^{-1/d^*(H)+\epsilon}$ for some fixed positive ϵ. This leads to Theorem 2.2.

The extensions to hypergraphs are similarly unchallenging, basically amounting to the observation that our arguments really make no use of the assumption that edges have size two.

13. Appendix A: equivalence of Theorems 2.3 and 2.4

Of course we only need to show Theorem 2.4 implies Theorem 2.3.

Set $\vartheta(n) = n^{-1/d(H)}(\log n)^{1/m}$ and $\mu(n,p) = (n^{v-1}p^m)^{n/v}$, and write G for $G(n,p)$ (for whatever p we specify).

Given K, we would like to show that there is a C_K for which $p = p(n) > C_K\vartheta(n)$ implies

$$\Pr(\Phi(G) \leq \mu(n,p)e^{-C_K n}) \leq n^{-K} \;\forall n. \tag{66}$$

Suppose this is not true and for each C, n define $g_C(n)$ by: if $p = p(n) = g_C(n)\vartheta(n)$ then

$$\Pr(\Phi(G(n,p)) \leq \mu(n,p)e^{-Cn}) = n^{-K}.$$

If the sequence $\{g_C(n)\}_n$ is bounded for some C then we have (66) with C_K the maximum of C and the bound on g_C.

So we may assume that $\{g_C(n)\}_n$ is unbounded for every C. We can then choose $n_1 < n_2 < \cdots$ so that $g_C(n_C) > C$ for $C = 1, 2, \ldots$.

Define the sequence $\{g(n)\}$ by $g(n) = g_C(n_C)$, where C is minimum with $n_C \geq n$. Since $g(n) \to \infty$, Theorem 2.4 says that there is a C^* so that if

$$p = p(n) = g(n)\vartheta(n) \tag{67}$$

then

$$\Pr(\Phi(G(n,p)) \leq \mu(n,p)e^{-C^* n}) = n^{-\omega(1)}.$$

Now choose n_0 so that for $n > n_0$ and p as in (67)

$$\Pr(\Phi(G(n,p)) \leq \mu(n,p)e^{-C^* n}) < n^{-K},$$

and then $C > C^*$ with $n_C > n_0$. Then for p still as in (67) and $n = n_C$ we have

$$\Pr(\Phi(G(n,p)) \leq \mu(n,p)e^{-Cn}) < \Pr(\Phi(G(n,p)) \leq \mu(n,p)e^{-C^* n}) < n^{-K}. \tag{68}$$

But this contradicts our definition of g, according to which the l.h.s. of (68) is n^{-K}.

References

[1] N. Alon and J. Spencer, *The Probabilistic Method*, Wiley, New York, 2000.

[2] N. Alon and R. Yuster, Threshold functions for H-factors, *Combin. Probab. Comput.* **2** (1993), 137-144.

[3] B. Bollobás, Random Graphs, pp. 257-274 in *Combinatorics*, London Math. Soc. Lecture Note Ser. 52, Cambridge Univ. Press, Cambridge, 1981.

[4] B. Bollobás, The evolution of sparse graphs, pp. 35-57 in *Graph Theory and Combinatorics*, B. Bollobás, ed., Academic Pr., 1984.

[5] B. Bollobás, *Random Graphs*, Academic Press, London, 1985.

[6] F.R.K. Chung, P. Frankl, R. Graham and J.B. Shearer, Some intersection theorems for ordered sets and graphs, *J. Combinatorial Th. Ser. A.* **48** (1986), 23-37.

[7] C. Cooper, A. Frieze, M. Molloy and B. Reed, Perfect matchings in random r-regular, s-uniform hypergraphs, *Combin. Probab. Comput.* **5** (1996), 1-14.

[8] I. Csiszár and J. Körner, *Information theory. Coding theorems for discrete memoryless systems.* Akadémiai Kiadó, Budapest, 1981.

[9] P. Erdős, On the combinatorial problems which I would most like to see solved, *Combinatorica* **1** (1981), 25-42.

[10] P. Erdős and A. Rényi, On the evolution of random graphs, *Publ. Math. Inst. Hungar. Acad. Sci.* **5** (1960), 17-61.

[11] A. Frieze and S. Janson, Perfect matchings in random s-uniform hypergraphs, *Random Structures & Algorithms* **7** (1995), 41-57.

[12] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*, Wiley, New York, 2000.

[13] J.H. Kim, Perfect matchings in random uniform hypergraphs, *Random Structures & Algorithms* **23** (2003), 111-132.

[14] J.H. Kim and V. Vu, Concentration of multi-variate polynomials and its applications, *Combinatorica* **20** (2000), 417-434.

[15] J. Komlós and E. Szemerédi, Limit distributions for the existence of Hamilton cycles in a random graph, *Discrete Math.* **43** (1983), 55-63.

[16] M. Krivelevich, Triangle factors in random graphs, *Combin. Probab. Comput.* **6** (1997), 337-347.

[17] M. Krivelevich, Perfect fractional matchings in random hypergraphs *Random Structures & Algorithms* **9** (1996), 317-334.

[18] R.J. McEliece, *The Theory of Information and Coding*, Addison-Wesley, London, 1977.

[19] A. Ruciński, Matching and covering the vertices of a random graph by copies of a given graph, *Disc. Math.* **105** (1992), 185-197.

[20] A. Ruciński, Open problem, in *Random Graphs (2)*, Proceedings, Poznań, 1989, A.M. Frieze and T. Łuczak, eds., John Wiley & Sons, New York, 284.

[21] J. Schmidt and E. Shamir, A threshold for perfect matchings in random d-pure hypergraphs, *Disc. Math.* **45** (1983), 287-295.

[22] J. Spencer, Threshold functions for extension statements, *J. Combin. Th. Ser. A* **53** (1990), 286-305.

[23] V. Vu, On the concentration of multi-variate polynomials with small expectation, *Random Structures & Algorithms* **16** (2000), 344-363.

[24] V. Vu, Concentration of non-Lipschitz functions and applications, *Random Structures & Algorithms* **20**(2002), 262-316.

Department of Mathematics, Natural and Computer Sciences, University of Gävle, SWEDEN

E-mail address: ajj@hig.se

Department of Mathematics, Rutgers University, Piscataway, NJ 08854 USA

E-mail address: jkahn@math.rutgers.edu

Department of Mathematics, Rutgers University, Piscataway, NJ 08854 USA

E-mail address: vanvu@math.rutgers.edu