ASYMPTOTICS FOR SHAMIR'S PROBLEM

JEFF KAHN

ABSTRACT. For fixed $r \geq 3$ and n divisible by r, let $\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{H}}^r_{n,M}$ be the random M-edge r-graph on $V = \{1, \dots, n\}$; that is, $\boldsymbol{\mathcal{H}}$ is chosen uniformly from the M-subsets of $\mathcal{K} := \binom{V}{r}$ ($:= \{r\text{-subsets of } V\}$). *Shamir's Problem* (circa 1980) asks, roughly,

for what $M = M(n)$ is $\boldsymbol{\mathcal{H}}$ likely to contain a perfect matching

(that is, n/r disjoint r-sets)?

In 2008 Johansson, Vu and the author showed that this is true for $M > C_r n \log n$. The present paper has two purposes. First, it establishes the asymptotically correct version of the 2008 result:

Theorem 1. *For fixed $\varepsilon > 0$ and $M > (1+\varepsilon)(n/r)\log n$,*

$$\mathbb{P}(\boldsymbol{\mathcal{H}} \text{ contains a perfect matching}) \to 1 \text{ as } n \to \infty.$$

Second, it begins a proof of the definitive "hitting time" statement:

Theorem 2. *If $A_1, \dots$ is a uniform permutation of $\mathcal{K}$, $\boldsymbol{\mathcal{H}}_t = \{A_1, \dots, A_t\}$, and*

$$T = \min\{t : A_1 \cup \cdots \cup A_t = V\},$$

then $\mathbb{P}(\boldsymbol{\mathcal{H}}_T$ contains a perfect matching$) \to 1$ as $n \to \infty$.

It is shown here that Theorem 2 follows from a conditional version of Theorem 1 that will be proved elsewhere. The key ideas in that proof are similar to those for Theorem 1, but the argument is a longer story, and it has seemed best to give the present separate proof of Theorem 1, in which those ideas may appear more clearly.

1. INTRODUCTION

A (simple) *r-graph* (or *r-uniform hypergraph*) is a set $\mathcal{H}$ of r-subsets (*edges*) of a *vertex* set $V = V(\mathcal{H})$; a *matching* of such an $\mathcal{H}$ is a set of disjoint edges; and a *perfect matching* (p.m.) is a matching of size $|V|/r$. Write $\boldsymbol{\mathcal{H}}^r_{n,M}$ for the random M-edge r-graph on $[n] := \{1, \dots, n\}$; that is, $\boldsymbol{\mathcal{H}}^r_{n,M}$ is chosen uniformly from the M-subsets of $\mathcal{K} := \binom{[n]}{r}$. (Some notation is collected at the end of this section.)

We are interested here in *Shamir's Problem*, which asks, roughly, with n ranging over (large) multiples of a fixed r,

for what M is $\boldsymbol{\mathcal{H}}^r_{n,M}$ likely to contain a perfect matching?

In what follows we will always work with a fixed r and omit this from our notation—so $\boldsymbol{\mathcal{H}}_{n,M}$ is $\boldsymbol{\mathcal{H}}^r_{n,M}$—and restrict to n divisible by r.

Shamir's Problem first appeared in print in [13], where Erdős says he heard it from Eli Shamir in 1979, and, following initial results in [27], became one of the most intensively studied questions in probabilistic combinatorics; for example, it and its graph factor analogue (see below) were, according to [20, Section 4.3], "two of the most challenging, unsolved problems in the theory of random structures."

Supported by NSF Grant DMS1501962, BSF Grant 2014290, and a Simons Fellowship.

For a more precise question, recall that $M_0 = M_0(n)$ is *a threshold* for the property of containing a perfect matching if

$$\mathbb{P}(\mathcal{H}_{n,M} \text{ has a perfect matching}) \to \begin{cases} 0 & \text{if } M/M_0 \to 0, \\ 1 & \text{if } M/M_0 \to \infty. \end{cases} \tag{1}$$

This notion was introduced by Erdös and Rényi in [14] and has been a central concern of probabilistic combinatorics since that time (see e.g. [20]). Note (1) depends only on the order of magnitude of M_0, though "the threshold" is a common abuse.

A natural guess—maybe with some hindsight; see below—is that, for any (fixed) r,

$$n \log n \text{ is a threshold for containing a perfect matching.} \tag{2}$$

(When it matters—here it does not—$\log$ is natural logarithm.) We think of this as crudely expressing the idea that *in the random setting* the main obstacle to existence of a perfect matching is isolated vertices (vertices not in any edges)—which typically disappear when $M \approx (n/r)\log n$.

Note that while (2) seems natural (or obvious) today, it was not always so. For example Erdős [13] says "... usually one can guess the answer [for random hypergraph problems] almost immediately. Here we have no idea ...," and [27] gives no guess as to the threshold. It was only in [8] that (2) (in the stronger form (3) below) was first suggested in print, though its likelihood was surely recognized before then.

Following various attempts, the most successful in [16] and [23] (see also e.g. [8, 24]), (2) was proved in [21]:

Theorem 1.1. *For each r there is a C_r such that if $M > C_r n \log n$ then $\mathcal{H}_{n,M}$ has a perfect matching w.h.p.*[1]

(See also [17, Sec. 13.2] for an exposition.)

The challenge since Theorem 1.1 has been to show that isolated vertices are more literally the issue. Ideally this means proving the precise *hitting time* statement: if $A_1 \ldots$ is a uniform permutation of $\mathcal{K}$ then w.h.p. the A_i's include a p.m. as soon as they cover the vertices. (This possibility is suggested in [21], but was by then an obvious guess.)

A somewhat less ambitious goal is to show that Theorem 1.1 holds for any (fixed) $C_r > 1/r$. We may call this *asymptotics of the threshold*: it gives $M_c \sim (n/r)\log n$, where $M_c = M_c(n)$ is *the* threshold, the least M for which

$$\mathbb{P}(\mathcal{H}_{n,M} \text{ has a perfect matching}) \geq 1/2$$

(which *is* a threshold in the Erdős-Rényi sense; see [5] or [20, Theorem 1.24]). It is easy to see that this asymptotic version does follow from the hitting time statement.

The conjecture of [8] is stronger than asymptotics of the threshold but weaker than the hitting time version: if $M = (n/r)(\log n + c_n)$, then

$$\mathbb{P}(\mathcal{H}_{n,M} \text{ has a perfect matching}) \to \begin{cases} 0 & \text{if } c_n \to 0, \\ e^{-e^{-c}} & \text{if } c_n \to c, \\ 1 & \text{if } c_n \to \infty. \end{cases} \tag{3}$$

Equivalently, the probability that $\mathcal{H}_{n,M}$ avoids isolated vertices yet fails to contain a perfect matching tends to zero. For $r = 2$, (3) and the hitting time statement were shown by Erdős and Rényi [15] and Bollobás and Thomason [4] respectively. (So Erdős' comment above might suggest he believed the answer for larger r would be different.)

[1] "with high probability," meaning with probability tending to 1 as $n \to \infty$

In this paper we show that the asymptotics of the threshold are as expected and give the first step in a proof of the hitting time statement that will be completed in [22]; thus:

Theorem 1.2. *For fixed $\varepsilon > 0$ and $M > (1+\varepsilon)(n/r)\log n$, $\boldsymbol{\mathcal{H}}_{n,M}$ has a perfect matching w.h.p.*

Theorem 1.3. *If $A_1, \ldots$ is a uniform permutation of $\mathcal{K}$, $\boldsymbol{\mathcal{H}}_t = \{A_1, \ldots, A_t\}$, and*

$$T = \min\{t : A_1 \cup \cdots \cup A_t = V\}$$

(the hitting time*), then $\boldsymbol{\mathcal{H}}_T$ has a perfect matching w.h.p.*

[We note in passing that Theorem 1.2 is equivalent to its analogue for $\boldsymbol{\mathcal{H}}_{n,p} = \boldsymbol{\mathcal{H}}^r_{n,p}$ (the random r-graph on $[n]$ in which each edge is present with probability p, independent of other choices):

Theorem 1.4. *For fixed $\varepsilon > 0$ and $p > (1+\varepsilon)\binom{n-1}{r-1}^{-1}\log n$, $\boldsymbol{\mathcal{H}}_{n,p}$ has a perfect matching w.h.p.*

See e.g. Propositions 1.12 and 1.13 of [20] for the equivalence.]

The story of these results ran very much contrary to expectations. The author had felt since [21] that a proof of Theorem 1.2 might not be out of the question (maybe a minority opinion), but that Theorem 1.3 was probably hopeless; but in retrospect it is the former that was the bigger step.

Theorem 1.3 is proved by reducing to a statement like Theorem 1.2, but in a conditional space where even routine points from the proof of Theorem 1.2 are not straightforward; so the present organization, with its separate proof of the now subsumed Theorem 1.2, is intended to focus on what seem the most important points. (It should also make the proof of Theorem 1.3 in [22] easier to follow, and will somewhat shorten [22], since some of what we do here *can* be used there directly.)

As in [21] our approach to Theorems 1.2 and 1.3 depends crucially on working with counting versions; with $\Phi(\mathcal{H})$ denoting the number of perfect matchings of $\mathcal{H}$, the corresponding stronger statements are:

Theorem 1.5. *For fixed $\varepsilon > 0$ and $M > (1+\varepsilon)(n/r)\log n$, w.h.p.*

$$\Phi(\boldsymbol{\mathcal{H}}_{n,M}) > \left[e^{-(r-1)}rM/n\right]^{n/r} e^{-o(n)}. \tag{4}$$

Theorem 1.6. *For $\boldsymbol{\mathcal{H}}_t$ and T as in Theorem 1.3, w.h.p.*

$$\Phi(\boldsymbol{\mathcal{H}}_T) > \left[e^{-(r-1)}\log n\right]^{n/r} e^{-o(n)}. \tag{5}$$

The right-hand sides are (of course) roughly the expectations of the left-hand sides; more precisely, they are within subexponential factors of those expectations.

In Section 2 we derive Theorem 1.2 from several statements whose proofs will be the main work of this paper. Outlining that work will be easier once we have the framework of Section 2, so is postponed until then.

Section 10 gives the reduction that is the first step in the proof of Theorem 1.3. (To be precise, we reduce Theorem 1.6 to Theorem 10.1, a conditional variant of Theorem 1.5. The same reduction gets Theorem 1.3 from the weaker version of Theorem 10.1 corresponding to Theorem 1.2, but, again, we don't know how to prove the weaker version without proving the stronger.)

Graph factors

Recall that, for graphs H and G, an *H-factor* of G is a collection of copies of H in G (subgraphs of G isomorphic to H) whose vertex sets partition $V(G)$. The graph-factor counterpart of Shamir's Problem asks (roughly), *when is the random graph $G_{n,M}$ likely to contain an H-factor?* This was originally suggested (for $H = K_3$) by Ruciński [26].

Here the naive guess—that vertices not in copies of H are the main obstruction—is not always correct, though one does expect it to be correct for *strictly balanced* H (see [20]) and *slightly* beyond. For strictly balanced H it is shown in [21] at the level of Erdős-Rényi thresholds (so the analogue of Theorem 1.2; this says, for example, that $n^{4/3}\log^{1/3} n$ is a threshold for existence of a triangle-factor). See Conjecture 1.1 of [21] for what ought to be true in general. Though given in detail only for graphs, the arguments and results of [21] extend essentially unmodified to *r-graph*-factors, where Theorem 1.2 is just the case that H consists of a single edge.

I expect—admittedly, without having thought very seriously—that the present results extend to the general graph (and hypergraph) factor setting, with, as was true in [21], some technical complications but all key ideas already appearing in the arguments for Shamir. Beautiful recent coupling arguments of O. Riordan [25] and A. Heckel [19] show that in some cases—e.g. cliques—the graph factor versions of Theorems 1.1 and 1.2 *follow from* the Shamir versions (e.g. at these levels of accuracy, Ruciński's triangle-factor question is *contained in* Shamir); but there seems little chance of anything analogous for Theorem 1.3.

Usage

Throughout the paper we fix $r \geq 3$; take $V = [n] := \{1, \ldots, n\}$, with $r|n$; and use $\mathcal{K}$ for $\binom{V}{r}$. We use v, w, x, y, z for vertices and $\mathcal{F}, \mathcal{G}, \mathcal{H}$ for r-graphs (a.k.a. subsets of $\mathcal{K}$), or bold versions of these when the r-graphs in question are random. As above, we abbreviate $\boldsymbol{\mathcal{H}}^r_{n,M} = \boldsymbol{\mathcal{H}}_{n,M}$, and the number of perfect matchings (or p.m.s) of $\mathcal{H}$ is denoted $\Phi(\mathcal{H})$.

We use $\mathcal{H}_x = \{A \in \mathcal{H} : x \in A\}$; $\Delta_{\mathcal{H}}$, $\delta_{\mathcal{H}}$ and $D_{\mathcal{H}}$ for maximum, minimum and average degrees in $\mathcal{H}$; and $\mathcal{H} - X = \{A \in \mathcal{H} : A \subseteq V \setminus X\}$. As usual the *codegree* (in $\mathcal{H}$) of x, y is $|\{A \in \mathcal{H} : x, y \in A\}|$. We will often abusively write $Y \cup x$ and $Y \setminus x$ for $Y \cup \{x\}$ and $Y \setminus \{x\}$.

Asymptotic notation is interpreted as $n \to \infty$ (with dependence on n typically suppressed). We use $a \ll b$ and $a = o(b)$ interchangeably and, similarly, $a \lesssim b$ is the same as $a < (1 + o(1))b$. We use both "a.e." and "a.a." to mean "for all but a $o(1)$-fraction." A familiar point that nonetheless seems worth mentioning: given ε, implied quantities in asymptotic expressions not mentioning ε (constants in $O(\cdot)$ and $\Omega(\cdot)$, rates in $o(\cdot)$ and $\omega(\cdot)$) depend on ε, but, for example, the implied constant in $O(\varepsilon)$ does not.

We use $\log$ for natural logarithm and $a \pm b$ for a quantity within b of a. We will always assume n is large enough to support our assertions and, following a common abuse, usually pretend large numbers are integers.

We will sometimes use bold for random objects: consistently for r-graphs, but otherwise only if it seems needed to distinguish a random object from its possible values. We use mathfrak characters ($\mathfrak{A}, \mathfrak{B}, \mathfrak{C}, \mathfrak{D}, \mathfrak{E}, \ldots$) for properties (saying, e.g., "$\mathcal{H}$ has property $\mathfrak{A}$," "$\mathcal{H}$ satisfies $\mathfrak{A}$," "$\mathcal{H} \models \mathfrak{A}$" as convenient) and events (e.g. $\mathfrak{A}_t = \{\boldsymbol{\mathcal{H}}_t \models \mathfrak{A}\}$; see Section 5), and will usually prefer $\mathfrak{A}\mathfrak{R}$ to $\mathfrak{A} \wedge \mathfrak{R}$.

2. Skeleton

Here we prove Theorem 1.5 modulo a few assertions whose justification will be the main content of the paper. As mentioned earlier, the approach is similar to that of [21]; a major, if nearly invisible, difference is the $o(n)$ in (4), which was formerly $O(n)$; see "Orientation" below for a *little* more on this.

Fix ε, let M be as in the statement of Theorem 1.5 (or 1.2), and set $T = \binom{n}{r} - M$. Let $A_1, \ldots, A_{\binom{n}{r}}$ be a uniform ordering of $\mathcal{K}$ $(= \binom{V}{r})$ and set $\mathcal{H}_t = \mathcal{K} \setminus \{A_1, \ldots, A_t\}$; so $\mathcal{H}_0 = \mathcal{K}$ and we may take $\mathcal{H}_{n,M} = \mathcal{H}_T$.

(Note the T and $\mathcal{H}_t$ here disagree with—are nearly the opposite of—those in Theorems 1.3 and 1.6, reflecting the different $\mathcal{H}_0$'s ($\mathcal{K}$ vs. $\emptyset$), but we will not see those theorems again until Section 10, when we are done with the present setup; nor will there ever be any danger of confusing $\mathcal{H}_t$ and $\mathcal{H}_x$ ($= \{A \in \mathcal{H} : x \in A\}$; see Usage).)

Set $\Phi(\boldsymbol{\mathcal{H}}_t) = \Phi_t$ and let ξ_t be the fraction of perfect matchings of $\boldsymbol{\mathcal{H}}_{t-1}$ that contain A_t (so $\xi_t = \Phi(\boldsymbol{\mathcal{H}}_{t-1} - A_t)/\Phi_{t-1}$). Then

$$\Phi_t = \Phi_0(1-\xi_1)\cdots(1-\xi_t);$$

equivalently,

$$\log \Phi_t = \log \Phi_0 + \textstyle\sum_{i=1}^t \log(1-\xi_i). \tag{6}$$

It will be helpful to set

$$\Lambda = (r-1)n/r; \tag{7}$$

this quantity represents one of the crucial differences between the present work and [21] (see "Orientation" following Lemma 4.1).

Notice that (by Stirling's Formula)

$$\log \Phi_0 = \log \frac{n!}{(n/r)!(r!)^{n/r}} = \frac{n}{r}\log \tbinom{n}{r-1} - \Lambda + O(\log n) \tag{8}$$

(recall $\log$ is $\ln$), and that

$$\mathbb{E}\xi_i = \frac{n/r}{\binom{n}{r} - i + 1} =: \gamma_i, \tag{9}$$

since in fact

$$\mathbb{E}[\xi_i | A_1, \ldots, A_{i-1}] = \gamma_i \tag{10}$$

for *any* choice of $A_1, \ldots, A_{i-1}$. (Strictly speaking (10) requires $\Phi(\mathcal{H}_{i-1}) \neq 0$, but this will be true in any case we consider.) Thus

$$\sum_{i=1}^t \mathbb{E}\xi_i = \sum_{i=1}^t \gamma_i = \frac{n}{r}\log\frac{\binom{n}{r}}{\binom{n}{r} - t} + o(1), \tag{11}$$

provided $\binom{n}{r} - t = \omega(n)$.

Let $\mathfrak{A}_t$ be the event

$$\Big\{\log \Phi_t > \log \Phi_0 - \textstyle\sum_{i=1}^t \gamma_i - o(n)\Big\}. \tag{12}$$

Remark. We note, perhaps unnecessarily, that (12) refers to some specific $o(n)$, so that it makes sense to talk about $\mathfrak{A}_t$ for a particular n (as opposed to a sequence). Related points will be common below, and, somewhat departing from common practice, we will elaborate in a couple places where this seems possibly helpful; see following (58) for a first instance.

Combining (8) and the expression for $\sum \gamma_i$ in (11) (with $t = T$, in which case the argument of the log is $\binom{n}{r}/M$), we find that $\mathfrak{A}_T$ says

$$\log \Phi_T > (n/r)\log(rM/n) - \Lambda - o(n), \tag{13}$$

which is the same as (4); so Theorem 1.5 is

$$\mathbb{P}(\overline{\mathfrak{A}}_T) = o(1). \tag{14}$$

(We will in fact show $\mathbb{P}(\cup_{t\leq T}\overline{\mathfrak{A}}_t) = o(1)$; see (19).)

For (14) we use the method of martingales with bounded differences. Here it is natural—though we will need a slight variant—to consider the martingale

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

(it is a martingale by (10)), with associated difference sequence

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

In general proving concentration of such X_t's depends on maintaining some control over the $|Z_i|$'s, to which end we keep track of two sequences of auxiliary events, $\mathfrak{B}_i$ and $\mathfrak{R}_i$ ($i \leq T-1$). These will be defined in Section 5. Roughly, $\mathfrak{B}_i$ says that no edge of $\mathcal{H}_i$ is in too much more than its proper share of perfect matchings, while $\mathfrak{R}_i$ consists of standardish degree restrictions.

For $i \leq T$ it will follow trivially from $\mathfrak{B}_{i-1}$ (see (59)) that

$$\xi_i = O(\gamma_i). \tag{15}$$

This is more than enough for the desired concentration, but can occasionally fail, since $\mathfrak{B}_{i-1}$ may fail. We accordingly slightly modify the above X's and Z's, setting

$$Z_i = \begin{cases} \xi_i - \gamma_i & \text{if } \mathfrak{B}_j \text{ holds for all } j < i, \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

(and $X_t = \sum_{i=1}^t Z_i$). As shown in Section 3, a martingale analysis along the lines of Azuma's Inequality then gives

$$\mathbb{P}(|X_t| > \lambda) < n^{-\omega(1)} \text{ for } \lambda \gg \sqrt{n}. \tag{17}$$

Notice that if we do have $\mathfrak{B}_i$ for all $i < t < T$ (so $X_t = \sum_{i=1}^t(\xi_i - \gamma_i)$) and $X_t < \sqrt{n}\log n$ (say; there is plenty of room here), then we have $\mathfrak{A}_t$; for (15) gives

$$\begin{aligned} \textstyle\sum_{i=1}^t \xi_i^2 &= O(\textstyle\sum_{i=1}^t \gamma_i^2) \\ &= O((n/r)^2 \textstyle\sum\{j^{-2} : j > (n/r)\log n\}) = O(n/\log n); \end{aligned} \tag{18}$$

so (using (6))

$$\log \Phi_t > \log \Phi_0 - \sum_{i=1}^t (\xi_i + \xi_i^2) > \log \Phi_0 - \sum_{i=1}^t \gamma_i - O(n/\log n),$$

where the first inequality uses $\xi_i = o(1)$ (which follows from (15) and (9)), and the $O(n/\log n)$ absorbs the smaller $\sum_{i=1}^t(\xi_i - \gamma_i) = X_t$.

Thus the first failure, if any, of an $\mathfrak{A}_t$ (with $t \leq T$) must occur either because X_t is too large or because $\mathfrak{B}_i$ fails for some $i < t$; formally, we have

$$\mathbb{P}(\cup_{t\leq T}\overline{\mathfrak{A}}_t) < \mathbb{P}(\cup_{t<T}\overline{\mathfrak{R}}_t) + \sum_{t<T}\mathbb{P}(\mathfrak{A}_t\mathfrak{R}_t\overline{\mathfrak{B}}_t) + \sum_{t\leq T}\mathbb{P}((\cap_{i<t}\mathfrak{B}_j) \cap \overline{\mathfrak{A}}_t). \tag{19}$$

Here the last sum is $n^{-\omega(1)}$ by (17) and the discussion following it, and we will show

$$\mathbb{P}(\cup_{t<T}\overline{\mathfrak{R}}_t) = o(1) \tag{20}$$

and, for $i \leq T$,

$$\mathbb{P}(\mathfrak{A}_i\mathfrak{R}_i\overline{\mathfrak{B}}_i) = n^{-\omega(1)}. \tag{21}$$

So the l.h.s. of (19) is $o(1)$, which in particular gives (14) (and Theorem 1.5). □

Remark. Thus most of the exceptional probability comes from the $\mathfrak{R}_i$'s, which include lower bounds on minimum degrees (see (55)) whose failure probability is not all that small. If the process survives the $\mathfrak{R}_i$'s, then the probability that it fails for some other reason is much smaller.

Orientation. What this paper is really about—as was [21]—is bounding the increments ξ_i; that is, establishing (15), which, as already mentioned, follows immediately from $\mathfrak{B}_{i-1}$. The martingale analysis that handles (*via* (17)) the last term in (19) is then pretty standard, and the genericity assertions (20) are also fairly routine.

Thus the heart of the matter is (21), which is proved in Sections 6-9, with the assistance of the entropy machinery developed in Section 4. The most important part of this is Sections 8 and 9, but the Brégman-like bound of Theorem 4.2, which underlies Section 8 and seems of independent interest, is also critical: as mentioned at the beginning of this section, a crucial difference between the present outline and the corresponding discussion in [21] is the $o(n)$—which in [21] was $O(n)$—in the definition of $\mathfrak{A}_t$, and it is Theorem 4.2 that provides the opening to exploiting this.

We will try to comment on particular aspects of the argument when we are in a position to do so more intelligibly.

Outline. After briefly recalling large deviation basics, Section 3 records what we need in the way of martingale concentration, in a form convenient for a second application in Section 9, and gives the calculation for (17). As mentioned above, Section 4 treats entropy, with main point the aforementioned Theorem 4.2. In Section 5 we finally define the events $\mathfrak{B}_i$ and $\mathfrak{R}_i$ as part of a somewhat more general discussion, give the easy derivation of (15) from $\mathfrak{B}_{i-1}$, and, in (58), slightly reformulate (21). The uninteresting proof of (20) is banished to an appendix that the reader is encouraged to skip. And, again, Sections 6-9 prove (58), thus establishing (21) and, according to the above discussion, completing the proof of Theorem 1.5.

3. Concentration

Before turning to the main business of this section we review a couple standard "Chernoff-type" bounds. Recall that a r.v. ξ is *hypergeometric* if, for some s, a and k, it is distributed as $|X \cap A|$, where A is a fixed a-subset of the s-set S and X is uniform from $\binom{S}{k}$.

Theorem 3.1. *If ξ is binomial or hypergeometric with $\mathbb{E}\xi = \mu$, then for $t \geq 0$,*

$$\Pr(\xi \geq \mu + t) \leq \exp\left[-\mu\varphi(t/\mu)\right] \leq \exp\left[-t^2/(2(\mu + t/3))\right], \tag{22}$$

$$\Pr(\xi \leq \mu - t) \leq \exp[-\mu\varphi(-t/\mu)] \leq \exp[-t^2/(2\mu)], \tag{23}$$

where $\varphi(x) = (1 + x)\log(1 + x) - x$ for $x > -1$ and $\varphi(-1) = 1$.

(See e.g. [20, Theorems 2.1 and 2.10].) For larger deviations the following consequence of the finer bound in (22) is helpful.

Theorem 3.2. *For ξ and μ as in Theorem 3.1 and any K,*

$$\Pr(\xi > K\mu) < \exp[-K\mu \log(K/e)].$$

We now turn to martingales and (17). The argument for the latter is about the same as that for the corresponding assertion in [21], but we now present the basic machinery in somewhat greater generality to support a second application in Section 9. There is nothing much new here, but, lacking a convenient reference, we include some details.

Lemma 3.3. *If $Z_1, \ldots, Z_t$ is a martingale difference sequence with respect to the random sequence $Y_1, \ldots, Y_t$ (that is, Z_i is a function of $Y_1, \ldots, Y_i$ and $\mathbb{E}[Z_i|Y_1, \ldots, Y_{i-1}] = 0$), then for $Z = \sum Z_i$ and any $\vartheta > 0$,*

$$\mathbb{E}e^{\vartheta Z} \leq \prod_{i=1}^t \max \mathbb{E}[e^{\vartheta Z_i}|y_1, \ldots, y_{i-1}] \tag{24}$$

and, consequently, for any $\lambda > 0$,

$$\mathbb{P}(Z > \lambda) < e^{-\vartheta\lambda} \prod_{i=1}^t \max \mathbb{E}[e^{\vartheta Z_i}|y_1, \ldots, y_{i-1}] \tag{25}$$

(where y_i ranges over possibilities for Y_i).

Proof. As usual, (25) follows from (24), using $\mathbb{P}(Z > \lambda) = \mathbb{P}(e^{\vartheta Z} > e^{\vartheta\lambda})$ and Markov's Inequality. For (24), with B_i denoting the ith factor on the r.h.s., induction on t gives

$$\begin{aligned} \mathbb{E}e^{\vartheta Z} &= \mathbb{E}\{\mathbb{E}[e^{\vartheta Z}|Y_1, \ldots, Y_{t-1}]\} \\ &= \mathbb{E}\{e^{\vartheta(Z_1+\cdots+Z_{t-1})}\mathbb{E}[e^{\vartheta Z_t}|Y_1, \ldots, Y_{t-1}]\} \\ &\leq B_t \cdot \mathbb{E}[e^{\vartheta(Z_1+\cdots+Z_{t-1})}] \\ &\leq \prod B_i. \end{aligned}$$

□

Both here and in Section 9, bounds on the factors in (24) are given by the next observation.

Proposition 3.4. *For a r.v. $W \in [0, b]$ with $\mathbb{E}W \leq a$, and $\vartheta \in [0, (2b)^{-1}]$,*

$$\max\{\mathbb{E}e^{\vartheta(W-\mathbb{E}W)}, \mathbb{E}e^{-\vartheta(W-\mathbb{E}W)}\} \leq e^{\vartheta^2 ab}. \tag{26}$$

Proof. Since the bound is increasing in a, it is enough to prove it when $\mathbb{E}W = a$. Given this and the bounds on W, convexity implies that each of $\mathbb{E}e^{\vartheta W}$, $\mathbb{E}e^{-\vartheta W}$ is maximized (for any ϑ) when W is b with probability $p := a/b$ and zero otherwise, in which case we have

$$\mathbb{E}e^{\vartheta(W-\mathbb{E}W)} = e^{-\vartheta bp}[1 - p + pe^{\vartheta b}], \ \mathbb{E}e^{-\vartheta(W-\mathbb{E}W)} = e^{\vartheta bp}[1 - p + pe^{-\vartheta b}];$$

and simple calculations show that $e^{-xp}[1 - p + pe^x] \leq e^{x^2 p}$ for $|x| \leq 1/2$ (and any p), implying (26). □

Proof of (17). Let $\varsigma_i = O(\gamma_i)$ be the bound on ξ_i in (15). We will apply Lemma 3.3 with $Y_i = A_i$ and Z_i as in (16) (so $Z = X_t$), using Proposition 3.4 with $b = \varsigma_i$ and $a = \gamma_i$ to bound the factors in (25) (or (24)). (For relevance of the proposition notice that, conditioned on any particular values of $A_1, \ldots, A_{i-1}$, Z_i is either identically zero (as happens if $\mathfrak{B}_j$ has failed for some $j < i$) or $Z_i = \xi_i - \gamma_i$, where $\xi_i \in [0, \varsigma_i]$ has (conditional) expectation γ_i (see (10)).) This combination (i.e. of Lemma 3.3 and Proposition 3.4) gives

$$\mathbb{P}(X_t > \lambda) < \exp[\vartheta^2 \textstyle\sum_{i=1}^t \varsigma_i\gamma_i - \vartheta\lambda]$$

for any $\lambda > 0$, provided, say, $\vartheta \leq 1$ ($\leq (2 \max \varsigma_i)^{-1}$). So with

$$J = \textstyle\sum_{i=1}^t \varsigma_i\gamma_i = O(\sum \gamma_i^2) = O(n/\log n)$$

(see (18)) and $\vartheta = \min\{1, \lambda/(2J)\}$, we have

$$\Pr(X_t > \lambda) < \begin{cases} \exp[-\lambda^2/(4J)] & \text{if } \lambda \leq 2J, \\ \exp[-\lambda/2] & \text{otherwise;} \end{cases}$$

and for $\lambda \gg \sqrt{n}$ (as in (17)) each bound is $n^{-\omega(1)}$ (in the first case since $J = O(n/\log n)$).

The same argument applies to $\mathbb{P}(X_t < -\lambda) = \mathbb{P}(-X_t > \lambda)$ (though this part of (17) isn't needed for the proof of Theorem 1.5). □

4. Entropy

Here we develop what we need in the way of entropy. The main result is Theorem 4.2, an extension (essentially) of Theorem 1.2(a) of [10] (itself more or less a generalization of Brégman's Theorem [6]) that is one main point underlying the present improvement of [21]. The discussion also includes a pair of technical observations, Lemmas 4.3 and 4.4, that support the use of Theorem 4.2 in Section 7.

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

$$H(X) = -\sum_x p(x) \log p(x),$$

where $p(x) = \mathbb{P}(X = x)$. For entropy basics see e.g. [9].

For a hypergraph $\mathcal{H}$ and $v \in V = V(\mathcal{H})$ ($= [n]$ as usual), we use $X(v, \mathcal{H})$ for the edge containing v in a uniformly chosen perfect matching of $\mathcal{H}$, and $h(v, \mathcal{H})$ for $H(X(v, \mathcal{H}))$. (We will not need to worry about $\mathcal{H}$'s without perfect matchings.)

Before turning to our main point we recall one instance of *Shearer's Lemma* [7]; this played a role in [21] corresponding to that of the present Theorem 4.2, and we will find some lesser use for it here.

Lemma 4.1. *For any r-graph* $\mathcal{H}$,

$$\log \Phi(\mathcal{H}) \leq r^{-1} \textstyle\sum_{v \in V} h(v, \mathcal{H}).$$

Orientation. The main purpose of this section is to recover (essentially) a missing $-\Lambda$ ($= -(r-1)n/r$) in the bound of Lemma 4.1. For example when $r = 2$ (so $\Lambda = n/2$), the lemma bounds $\log \Phi(G)$ for a d-regular, n-vertex graph G by $(n/2)\log d$, which an observation of L. Lovász and the author ([10, Eq. (8)] or [1, 11]; it is just the extension of Brégman to not necessarily bipartite G) improves to $\frac{n}{2d}\log(d!) = (n/2)\log d - \Lambda + o(n)$. The missing Λ was irrelevant in [21], since the argument there involved other losses that could not be made smaller than $O(n)$; here the present gain will eventually cancel the $-\Lambda$ in the bound (12) of $\mathfrak{A}_t$ (hidden in the $\log \Phi_0$; see (8)): see the interplay of (76) and (77) in Section 8.

In what follows we will treat a p.m. f as either a set of edges or a function from vertices to $\binom{V}{r-1}$; we use f_v for the edge of f containing v (taking the first view) and $f(v)$ for $f_v \setminus v$ (taking the second).

For Theorem 4.2 we consider a random (not necessarily uniform) p.m. $\mathbf{f}$ of a given r-graph $\mathcal{H}$ (with number of vertices divisible by r). We use v for vertices and Y for $(r-1)$-sets, and always assume $v \notin Y$.

Set $p_v(Y) = \mathbb{P}(\mathbf{f}(v) = Y)$. For a p.m. f, let

$$T(v, f, Y) = \{B \in f : B \neq f_v, B \cap Y \neq \emptyset\}$$

and $\tau(v, f, Y) = |T(v, f, Y)|$. Thus $\tau(v, f, Y) \leq r - 1$, with equality iff the vertices of $v \cup Y$ lie in distinct edges of f (thought of as "generic" behavior of (v, Y) w.r.t. f). With f running over p.m.s of $\mathcal{H}$, set

$$\Gamma_v(Y) = \{f : \tau(v, f, Y) < r - 1\} \tag{27}$$

(note this includes f's with $f(v) = Y$) and $\gamma_v(Y) = \mathbb{P}(\mathbf{f} \in \Gamma_v(Y))$.

Theorem 4.2. *With notation as above,*

$$\begin{aligned} H(\mathbf{f}) \quad < \quad & r^{-1} \textstyle\sum_v H(\mathbf{f}(v)) - \Lambda \\ & + O(\textstyle\sum_v \sum_Y p_v(Y)\gamma_v(Y)^{1/(r-1)}) + O(\log n). \end{aligned} \tag{28}$$

(Of course when $\mathbf{f}$ is uniform, $H(\mathbf{f}(v))$ is another name for $h(v,\mathcal{H})$.) Again, the point here is the "$-\Lambda$"; the ugly terms following it are errors we hope to ignore.

Proof. (The argument here is similar to that for Theorem 1.2(a) in [10].) Note we may assume $\mathcal{H}=\mathcal{K}\,(=\binom{V}{r})$, since we can regard $\mathbf{f}$ as a random matching of $\mathcal{K}$ that doesn't use edges not belonging to $\mathcal{H}$.

We use f_B for the restriction of f (viewed as a function) to $B\subseteq V$. For a permutation σ of V—always thought of as an ordering of V—and $v\in V$, set $B(\sigma,v)=\{w\in V:\sigma(w)<\sigma(v)\}$. Let $\boldsymbol{\sigma}$ be a random (uniform) permutation of V and $\mathbf{X}_v=(\boldsymbol{\sigma},\mathbf{f}_{B(\boldsymbol{\sigma},v)})$. Then (by the "chain rule" for entropy; see [9, Theorem 2.2.1])

$$\begin{aligned} H(\mathbf{f}) &= \frac{1}{n!}\sum_\sigma\sum_v H(\mathbf{f}(v)|\mathbf{f}_{B(\sigma,v)})\\ &= \sum_v\sum_\sigma\sum_g \frac{1}{n!}\mathbb{P}(\mathbf{f}_{B(\sigma,v)}=g)H(\mathbf{f}(v)|\sigma,g)\\ &= \sum_v H(\mathbf{f}(v)|\mathbf{X}_v), \end{aligned} \tag{29}$$

where σ ranges over permutations and, given σ, g ranges over possible values of $\mathbf{f}_{B(\sigma,v)}$ (and the conditioning on (σ,g) has the obvious meaning).

Now let

$$\boldsymbol{Z}_v=\begin{cases} \mathbf{f}_v & \text{if } B(\boldsymbol{\sigma},v)\cap\mathbf{f}(v)\neq\emptyset,\\ (V\setminus\{v\})\setminus\bigcup\{\mathbf{f}_w:w\in B(\boldsymbol{\sigma},v)\} & \text{otherwise.}\end{cases}$$

The condition in the first line just says v is not the first vertex of $\mathbf{f}_v$ in $\boldsymbol{\sigma}$, in which case $\mathbf{f}_v$ is determined by $\mathbf{f}_{B(\boldsymbol{\sigma},v)}$; these cases will be basically ignored in what follows. In the remaining cases $\boldsymbol{Z}_v$ is the set of vertices that can (in principle) belong to $\mathbf{f}(v)$ once we have specified $\mathbf{f}(w)$ for w preceding v in $\boldsymbol{\sigma}$.

Since $\boldsymbol{Z}_v$ is determined by $\mathbf{X}_v$, we have $H(\mathbf{f}(v)|\mathbf{X}_v)\le H(\mathbf{f}(v)|\boldsymbol{Z}_v)$, so, by (29),

$$H(\mathbf{f})\le\sum_v H(\mathbf{f}(v)|\boldsymbol{Z}_v); \tag{30}$$

so we would like to bound $H(\mathbf{f}(v)|\boldsymbol{Z}_v)$.

We now fix v and write $\boldsymbol{Z}$ for $\boldsymbol{Z}_v$. We use Y for values of $\mathbf{f}(v)$ and Z for values of $\boldsymbol{Z}$ *not of the form f*, and set $p_Y=p_v(Y)$ and $\gamma_Y=\gamma_v(Y)$. (See the paragraph preceding Theorem 4.2 for the notation.) We use $\mathbb{P}(Z)$ for $\mathbb{P}(\boldsymbol{Z}=Z)$, $\mathbb{P}(Z|Y)$ for $\mathbb{P}(\boldsymbol{Z}=Z|\mathbf{f}(v)=Y)$ and so on. (It may be worth stressing that $\mathbb{P}$ refers to $\boldsymbol{\sigma}$ *and* $\mathbf{f}$, and that these are independent.)

Since $H(\mathbf{f}(v)|\boldsymbol{Z}=f)=0$, we have

$$\begin{aligned} H(\mathbf{f}(v)|\boldsymbol{Z}) &= \sum_Z\mathbb{P}(Z)\sum_Y\mathbb{P}(Y|Z)\log\frac{1}{\mathbb{P}(Y|Z)}\\ &= \sum_Y\sum_Z\mathbb{P}(Y,Z)\log\frac{\mathbb{P}(Z)}{\mathbb{P}(Y,Z)}\\ &= \sum_Y p_Y\left[\frac{1}{r}\log\frac{1}{p_Y}+\sum_Z\mathbb{P}(Z|Y)\log\frac{\mathbb{P}(Z)}{\mathbb{P}(Z|Y)}\right] \qquad (31)\\ &= r^{-1}H(\mathbf{f}(v))+\sum_Y p_Y\sum_Z\mathbb{P}(Z|Y)\log\frac{\mathbb{P}(Z)}{\mathbb{P}(Z|Y)}\,. \qquad (32) \end{aligned}$$

(For (31) notice that independence of $\mathbf{f}$ and $\boldsymbol{\sigma}$ gives $\sum_Z\mathbb{P}(Z|Y)=1/r$ for any Y for which $p_Y\neq 0$.) We would like to show that the second term in (32) is less than about $-\Lambda$.

Fix Y with $p_Y \neq 0$. Let $\mathfrak{S} = \{B(\boldsymbol{\sigma}, v) \cap \mathbf{f}(v) = \emptyset\}$ (that is, v is the first vertex of f_v under $\boldsymbol{\sigma}$) and for $k \in [n-1]$ set

$$\begin{aligned} q_k &= \sum\{\mathbb{P}(Z|Y) : Z \supseteq Y, |Z| = k\} = \mathbb{P}(\mathfrak{S}, |\boldsymbol{Z}| = k | \mathbf{f}(v) = Y), \\ r_k &= \sum\{\mathbb{P}(Z) : Z \supseteq Y, |Z| = k\} = \mathbb{P}(\mathfrak{S}, |\boldsymbol{Z}| = k, \boldsymbol{Z} \supseteq Y). \end{aligned}$$

(Notice that "$|\boldsymbol{Z}| = k$" and "$\boldsymbol{Z} \supseteq Y$" make sense once we know $\mathfrak{S}$ holds, and that it is not really necessary to specify "$Z \supseteq Y$" in the definition of q_k.) Then the inner sum in (32) is

$$\sum_k q_k \sum \left\{ \frac{\mathbb{P}(Z|Y)}{q_k} \log \frac{\mathbb{P}(Z)}{\mathbb{P}(Z|Y)} : |Z| = k \right\} \leq \sum_k q_k \log \frac{r_k}{q_k} \tag{33}$$

(using Jensen's Inequality), so that (28) will follow from

$$\sum_k q_k \log \frac{r_k}{q_k} \; < \; -(r-1)/r + O(\gamma_Y^{1/(r-1)} + n^{-1} \log n). \tag{34}$$

We next discuss values of the q_k's and r_k's (with justifications to follow). We have

$$q_k = \begin{cases} 1/n & \text{if } k = r-1, 2r-1, \ldots, n-1, \\ 0 & \text{otherwise.} \end{cases} \tag{35}$$

For the r_k's we omit precise specification and settle for upper bounds: with

$$s_k = \begin{cases} \frac{1}{n} \frac{((k-r+1)/r)_{r-1}}{(n/r-1)_{r-1}} & \text{if } k = r-1, 2r-1, \ldots, n-1, \\ 0 & \text{otherwise} \end{cases} \tag{36}$$

(where $(a)_b = a(a-1)\cdots(a-b+1)$), we have

$$r_k \leq \gamma_Y q_k + (1-\gamma_Y) s_k. \tag{37}$$

Justification. In fact (we assert) (35) holds even if we condition on the value of $\mathbf{f}$; that is, (35) is still correct if we replace q_k by

$$q_k(f) = \mathbb{P}(\mathfrak{S}, |\boldsymbol{Z}| = k | \mathbf{f} = f)$$

for any f (but we only use this when $f(v) = Y$). Similarly, we have

$$\forall f \notin \Gamma_v(Y) \; s_k(f) := \mathbb{P}(\mathfrak{S}, |\boldsymbol{Z}| = k, \boldsymbol{Z} \supseteq Y | \mathbf{f} = f) = s_k. \tag{38}$$

To see these, first observe that, given $\{\mathbf{f} = f\}$ and $\mathfrak{S}$, if we order the edges of f according to their first vertices under $\boldsymbol{\sigma}$, then $|\boldsymbol{Z}| - (r-1)$ is r times the number of edges of f that follow f_v, and this number is uniform from $\{0, \ldots, n/r - 1\}$. (So $q_k(f)$ is as in (35).)

Note further that (again, given f) $Y \subseteq \boldsymbol{Z}$ iff each of the $\tau := \tau(v, f, Y)$ edges of $T(v, f, Y)$ follows f_v. But once we know $|\boldsymbol{Z}| = k$, the set of f-edges following f_v is chosen uniformly from the $((k-r+1)/r)$-subsets of the $(n/r-1)$-set $f \setminus \{f_v\}$. This gives (38) and, more generally (though we won't use it),

$$s_k(f) = \frac{1}{n} \frac{((k-r+1)/r)_\tau}{(n/r-1)_\tau}$$

for *any* f, $\tau = \tau(v, f, Y)$ and k as in the first line of (36).

Finally, bounding the second probability in

$$r_k = \textstyle\sum_f \mathbb{P}(f) \mathbb{P}(\mathfrak{S}, |\boldsymbol{Z}| = k, \boldsymbol{Z} \supseteq Y | \mathbf{f} = f)$$

by $q_k(f) = q_k$ for each $f \in \Gamma_v(Y)$ gives (37). □

Now returning to (34) we have, with $\gamma = \gamma_Y$ and t in the sums running from 1 to n/r,

$$\begin{aligned}\sum_k q_k \log \frac{r_k}{q_k} &\le \tfrac{1}{n}\sum \log[\gamma + (1-\gamma)\tfrac{(t-1)_{r-1}}{(n/r-1)_{r-1}}] \\ &< \tfrac{1}{n}\sum \log[\gamma + (1-\gamma)(rt/n)^{r-1}] \\ &= \tfrac{r-1}{n}\sum \log(rt/n) + \tfrac{1}{n}\sum \log(1+\gamma\left(\left(\tfrac{n}{rt}\right)^{r-1} - 1\right)) \\ &< -\tfrac{r-1}{r} + O\left(\tfrac{\log n}{n}\right) + \tfrac{1}{n}\sum \log\left(1+\gamma\left(\tfrac{n}{rt}\right)^{r-1}\right) \qquad (39)\end{aligned}$$

(using $\sum \log(rt/n) = \log[(r/n)^{n/r}(n/r)!]$ and Stirling's formula for the last line).

So for (34) it is enough to bound the sum in (39) by $O(\max\{n\gamma^{1/(r-1)}, 1\})$. For $\gamma < (r/n)^{r-1}$ the sum is less than

$$\sum \log(1 + t^{-(r-1)}) < \sum t^{-(r-1)} = O(1)$$

(recall $r \ge 3$). For larger γ, we set $B = \lfloor (n/r)\gamma^{1/(r-1)} \rfloor$ (noting that now $\gamma(n/r)^{r-1} < (2B)^{r-1}$) and bound the sum in (39) by

$$\sum_{1\le t\le B} \log\left(2\gamma\left(\tfrac{n}{rt}\right)^{r-1}\right) + \sum_{t>B} \gamma\left(\tfrac{n}{rt}\right)^{r-1}.$$

Here the first sum is less than

$$B\log(2\gamma(n/r)^{r-1}) - (r-1)\int_1^B \log x dx$$

$$< B\log(2^r B^{r-1}) - (r-1)[B\log B - B + 1] < B[r\log 2 + r - 1],$$

and for the second we have

$$\begin{aligned}\gamma(n/r)^{r-1}\sum_{t>B} t^{-(r-1)} &< \gamma(n/r)^{r-1}\int_{x>B} x^{-(r-1)}dx \\ &< (2B)^{r-1}\tfrac{1}{r-2}B^{-(r-2)} \le 2^{r-1}B.\end{aligned}$$

So the sum in (39) is $O(B) = O(n\gamma^{1/(r-1)})$ as desired. □

We now turn to the two auxiliary lemmas mentioned at the beginning of this section. The first of these will help in controlling the error terms in (28) when we come to apply Theorem 4.2.

Lemma 4.3. *Suppose* $p_i, \gamma_i \in [0,1]$, $i = 1, \ldots, l$, *satisfy*

$$l = n\Delta, \qquad (40)$$

$$\sum p_i = n, \qquad (41)$$

$$\sum p_i \log(1/p_i) > n\log\Delta - O(n) \qquad (42)$$

and

$$\sum \gamma_i = o(n\Delta). \qquad (43)$$

Then for any nondecreasing $h : [0,1] \to [0,1]$ *with* $h(x) \to 0$ *as* $x \to 0$,

$$\sum p_i h(\gamma_i) = o(n). \qquad (44)$$

Proof. Let $X = \sum \gamma_i$ and specify some ς with $1 \gg \varsigma \gg X/(n\Delta)$. From (41) we have

$$\sum_{\gamma_i \le \varsigma} p_i h(\gamma_i) \le n h(\varsigma) = o(n),$$

so may restrict attention to i's with $\gamma_i > \varsigma$. Let $\sum\{p_i : \gamma_i > \varsigma\} = \alpha n$. Then $\sum\{p_i h(\gamma_i) : \gamma_i > \varsigma\} \le \alpha n$, so it will be enough to show

$$\alpha = o(1). \tag{45}$$

Let $T = |\{i : \gamma_i > \varsigma\}| < X/\varsigma \ll n\Delta$. We have

$$\begin{aligned} n\log\Delta - O(n) &< \sum p_i \log(1/p_i) \\ &< (1-\alpha)n\log[n\Delta/((1-\alpha)n)] + \alpha n\log[T/(\alpha n)] \quad (46)\\ &= n[\log\Delta + H(\alpha) - \alpha\log(n\Delta/T)]. \quad (47)\end{aligned}$$

(For (46) we use the fact that $\sum_{i=1}^{l} x_i = a$ implies $\sum x_i \log(1/x_i) \le a\log(l/a)$.) But then $\alpha\log(n\Delta/T) - H(\alpha) = O(1)$ implies (45). (If $\alpha \ne o(1)$ then the l.h.s. is $\omega(\alpha)$, implying that α *is* $o(1)$.) □

Lemma 4.4. *For a probability distribution $p = (p_1, \ldots, p_l)$ and μ uniform distribution on $[l]$, if $H(p) = \log l - o(1)$, then, with $\|x\| = \sum |x_i|$,*

$$\|p - \mu\| = o(1); \tag{48}$$

equivalently, for some $\varsigma = o(1)$ and $\mathcal{B} = \{i : p_i \ne (1 \pm \varsigma)/l\}$,

$$|\mathcal{B}| = o(l) \tag{49}$$

and

$$\sum_{i \in \mathcal{B}} p_i = o(1). \tag{50}$$

Proof. For the equivalence, note that (49) and (50) imply

$$\|p - \mu\| \le \sum_{i\in\mathcal{B}} |p_i - 1/l| + \varsigma \le \sum_{i\in\mathcal{B}} (p_i + 1/l) + \varsigma = o(1),$$

while

$$\|p - \mu\| \ge \begin{cases} \varsigma|\mathcal{B}|/l, \\ \sum_{i\in\mathcal{B}}(p_i - 1/l) \end{cases}$$

with (say) $\varsigma = \|p - \mu\|^{1/2}$ shows that (48) implies (49) and then (50).

That the hypothesis of the lemma implies (48) is an instance of the next observation, whose elementary proof we omit.

Proposition 4.5. *If I is an interval of $\Re$ and $f : I \to \Re$ is twice differentiable with $f(0) = 0$ and $f'' > 0$, then for any r.v. X with $\mathbb{E}X = 0$,*

$$\mathbb{E}f(X) = \Omega(\min\{\mathbb{E}^2|X|, 1\})$$

(where the implied constant depends on f).

Here, with $\|p - \mu\| = \delta$, we may take $f(x) = (1 + x)\log(1 + x)$ and let X be $\alpha_i := lp_i - 1$, with i chosen uniformly from $[l]$. Then, noting that $\mathbb{E}X = \sum p_i - 1 = 0$ and $\mathbb{E}|X| = \delta$, and applying Proposition 4.5, we have

$$H(p) = \frac{1}{l}\sum(1 + \alpha_i)\log\frac{l}{1 + \alpha_i} = \log l - \mathbb{E}f(X) = \log l - \Omega(\delta^2),$$

which with $H(p) = \log l - o(1)$ implies $\delta = o(1)$. □

5. Properties $\mathfrak{A}$, $\mathfrak{R}$ and $\mathfrak{B}$

Properties here and in later sections are defined for a general r-graph $\mathcal{H}$, and then, for example, the event $\mathfrak{A}_t$ in Section 2 is $\{\boldsymbol{\mathcal{H}}_t \models \mathfrak{A}\}$. In this section we use n and m as defaults for the numbers of vertices and edges of $\mathcal{H}$, so

$$nD_{\mathcal{H}} = mr \tag{51}$$

(recall D is average degree). We will always use

$$t = \tbinom{n}{r} - m, \tag{52}$$

and will *tend* to use A for edges and Z or U for general r-sets (members of $\mathcal{K}$). We assume throughout that we have fixed some positive ε (it will be essentially the one in Theorem 1.5), upon which the implied constants in "$O(\cdot)$" and "$\Omega(\cdot)$" depend.

We say $\mathcal{H}$ has the property $\mathfrak{A}$ (or $\mathcal{H}$ satisfies $\mathfrak{A}$, or $\mathcal{H} \models \mathfrak{A}$) if

$$\log \Phi(\mathcal{H}) > \log \Phi_0 - \textstyle\sum_{i=1}^t \gamma_i - o(n), \tag{53}$$

(see (9) for γ_i), and the property $\mathfrak{R}$ if

$$\text{a.a. degrees in } \mathcal{H} \text{ are asymptotic to } D_{\mathcal{H}}, \tag{54}$$

$$\Delta_{\mathcal{H}} = O(D_{\mathcal{H}}),\ \delta_{\mathcal{H}} = \Omega(D_{\mathcal{H}}), \tag{55}$$

and

$$\text{all codegrees in } \mathcal{H} \text{ are } o(D_{\mathcal{H}}). \tag{56}$$

As noted above, $\mathfrak{A}_t$ and $\mathfrak{R}_t$ of Section 2 are then $\{\boldsymbol{\mathcal{H}}_t \models \mathfrak{A}\}$ and $\{\boldsymbol{\mathcal{H}}_t \models \mathfrak{R}\}$. Note that $\mathfrak{R}$ is "robust," in that

$$\text{if } \mathcal{H} \text{ satisfies } \mathfrak{R} \text{ then so does } \mathcal{H} - Z \text{ for every } Z \in \mathcal{K}. \tag{57}$$

(We omit the easy justification, just noting that (55) implies $D_{\mathcal{H}-Z} \sim D_{\mathcal{H}}$ and that each of (54)-(56) for $\mathcal{H} - Z$ depends on having (56) for $\mathcal{H}$.)

For $\mathfrak{B}$ a little notation will be helpful. For a finite set S and $\mathsf{w} : S \to \Re^+$ $(:= [0, \infty))$, set

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

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

and

$$\text{maxr } \mathsf{w}(S) = \overline{\mathsf{w}}(S)^{-1} \max \mathsf{w}(S).$$

For $\mathcal{H} \subseteq \mathcal{K}$ define $\mathsf{w}_{\mathcal{H}} : \mathcal{K} \to \Re^+$ by

$$\mathsf{w}_{\mathcal{H}}(Z) = \Phi(\mathcal{H} - Z),$$

and say $\mathcal{H}$ has the property $\mathfrak{B}$ if

$$\text{maxr } \mathsf{w}_{\mathcal{H}}(\mathcal{H}) = O(1)$$

(so the number of p.m.s containing any particular $A \in \mathcal{H}$ is not too large compared to the average). Then $\mathfrak{B}_t$ in Section 2 is $\{\boldsymbol{\mathcal{H}}_t \models \mathfrak{B}\}$, and (21) is

$$\text{for } m > (1+\varepsilon)(n/r)\log n,\ \mathbb{P}(\boldsymbol{\mathcal{H}}_{n,m} \models \mathfrak{A}\mathfrak{R}\overline{\mathfrak{B}}) = n^{-\omega(1)}. \tag{58}$$

(More formally: there is a fixed C, depending on the particular $o(\cdot)$'s and implied constants in $\mathfrak{A}$ and $\mathfrak{R}$, such that $\mathbb{P}(\{\boldsymbol{\mathcal{H}} \models \mathfrak{A}\mathfrak{R}\} \wedge \{\text{maxr } \mathsf{w}_{\boldsymbol{\mathcal{H}}}(\boldsymbol{\mathcal{H}}) > C\}) = n^{-\omega(1)}$.)

As mentioned at the end of Section 2, (58) is shown in Sections 6-9, and with (20) (likelihood of the $\mathfrak{R}_i$'s, proved in Section 11) will complete the proof of Theorem 1.5.

We conclude this section with the promised

$$\mathfrak{B}_{t-1} \text{ implies (15).} \tag{59}$$

(Recall (15) says $\xi_t = O(\gamma_t)$, where γ_t is now $n/(r(m+1))$; see (9) and (52).)

Proof. Given $\boldsymbol{\mathcal{H}}_{t-1} = \mathcal{H}$, we have $\xi_t \leq \max_{A\in\mathcal{H}} \mathsf{w}_{\mathcal{H}}(A)/\Phi(\mathcal{H})$, while γ_t is the average of these ratios, since

$$\textstyle\sum_{A\in\mathcal{H}} \mathsf{w}_{\mathcal{H}}(A) = \Phi(\mathcal{H})n/r$$

(and $|\mathcal{H}| = m+1$). This gives (59). □

6. More properties

We will get at $\mathfrak{B}$ (and (58)) *via* several auxiliary properties. We introduce the first three of these here (there will be a couple more in Section 9), together with assertions concerning them that together imply (58). Proofs of the assertions are mostly postponed to later sections.

Given $\mathcal{H}$, we now use D for $D_{\mathcal{H}}$. The first three auxiliary properties (for $\mathcal{H}$) are:

$\mathfrak{C}$: if $Z \in \mathcal{K}$ satisfies

$$\mathsf{w}_{\mathcal{H}}(Z) > \Phi(\mathcal{H})e^{-o(n)}, \tag{60}$$

then for any $x \in Z$,

$$\mathsf{w}_{\mathcal{H}}((Z \setminus x) \cup y) \gtrsim \mathsf{w}_{\mathcal{H}}(Z)d(x)/D \text{ for a.e. } y \in V \setminus Z; \tag{61}$$

$\mathfrak{E}$: $\mathsf{w}_{\mathcal{H}}(A) \sim \Phi(\mathcal{H})/D$ for a.e. $A \in \mathcal{H}$;

$\mathfrak{F}$: $\mathsf{w}_{\mathcal{H}}(Z) \sim \Phi(\mathcal{H})/D$ for a.e. $Z \in \mathcal{K}$.

(More formally, e.g. for $\mathfrak{E}$: there is $\varsigma = \varsigma(n) = o(1)$ such that $|\{A \in \mathcal{H} : \mathsf{w}_{\mathcal{H}}(A) \neq (1\pm\varsigma)\Phi(\mathcal{H})/D\}| < \varsigma|\mathcal{H}|$.) For perspective on $\mathfrak{E}$ and $\mathfrak{F}$—and for use below—note that (using (51))

$$(\overline{\mathsf{w}}_{\mathcal{H}}(\mathcal{H}) =)\ |\mathcal{H}|^{-1}\sum_{A\in\mathcal{H}} \mathsf{w}_{\mathcal{H}}(A) = |\mathcal{H}|^{-1}\Phi(\mathcal{H})n/r = \Phi(\mathcal{H})/D. \tag{62}$$

We now use $\boldsymbol{\mathcal{H}}$ for $\boldsymbol{\mathcal{H}}_{n,m}$ and $\mathcal{H}$ for a general m-edge r-graph on n, with n and m as in (58), and sometimes write

$$\mathfrak{X} \overset{*}{\Longrightarrow} \mathfrak{Z}$$

for $\mathbb{P}(\mathfrak{X}\overline{\mathfrak{Z}}) = n^{-\omega(1)}$; so e.g. the conclusion of (58) becomes

$$\{\boldsymbol{\mathcal{H}} \models \mathfrak{A}\mathfrak{R}\} \overset{*}{\Longrightarrow} \{\boldsymbol{\mathcal{H}} \models \mathfrak{B}\}. \tag{63}$$

The aforementioned assertions are as follows.

Lemma 6.1. *If $\mathcal{H}$ satisfies $\mathfrak{A}\mathfrak{R}$ then it satisfies $\mathfrak{E}$.*

Lemma 6.2. $\{\boldsymbol{\mathcal{H}} \models \mathfrak{A}\mathfrak{R}\} \overset{*}{\Longrightarrow} \{\boldsymbol{\mathcal{H}} \models \mathfrak{F}\}$.

Lemma 6.3. *For $x \in Z \in \mathcal{K}$,*

$$\{\boldsymbol{\mathcal{H}} \models \mathfrak{R}\} \wedge \{\boldsymbol{\mathcal{H}} - Z \models \mathfrak{F}\} \overset{*}{\Longrightarrow} \{(\boldsymbol{\mathcal{H}}, Z, x) \models (61)\}.$$

Lemma 6.4. *If $\mathcal{H}$ satisfies $\mathfrak{RFC}$ then it satisfies $\mathfrak{B}$.*

The nonprobabilistic Lemma 6.1, which is based mainly on the material of Section 4, allows us to replace $\mathfrak{AR}$ by $\mathfrak{ARE}$ in Lemma 6.2. That lemma then embodies the idea that $\mathfrak{E}\overline{\mathfrak{F}}$ is unlikely because the distribution of the $\mathsf{w}_{\boldsymbol{\mathcal{H}}}(A)$'s ($A \in \boldsymbol{\mathcal{H}}$) should reflect that of the $\mathsf{w}_{\boldsymbol{\mathcal{H}}}(Z)$'s ($Z \in \mathcal{K}$). We regard this natural point, and more particularly (83) below, as the heart of our argument; certainly it was the part whose proof took longest to find.

Lemmas 6.1-6.4 are stated in the order in which they are used in proving (58), but shown below in ascending order of difficulty and interest. Thus we prove Lemma 6.3 in Section 7, Lemma 6.1 in Section 8, and Lemma 6.2 in Section 9, with the easy Lemma 6.4 proved here following the derivation of (58).

Proof of (58). We first observe that $\{\mathcal{H} \models \mathfrak{R}\}$ implies $\{\mathcal{H} - Z \models \mathfrak{R}\}$ for any Z ($\in \mathcal{K}$); $\{\mathcal{H} \models \mathfrak{A}\}$ implies $\{\mathcal{H} - Z \models \mathfrak{A}\}$ for any Z as in (60); and Lemma 6.2 also holds with $\boldsymbol{\mathcal{H}} - Z$ in place of $\boldsymbol{\mathcal{H}}$ (for any Z). The first of these was already noted in (57) and the second is trivial, so we just need the (routine) justification of the third:

With $\boldsymbol{h} = |\boldsymbol{\mathcal{H}} - Z|$ and $m' = (1 - \varepsilon/2)m$, Theorem 3.1 gives

$$\mathbb{P}(\boldsymbol{h} < m') = \exp[-\Omega(\varepsilon^2 m)] = n^{-\omega(1)}. \tag{64}$$

So with $\mathfrak{X} = \mathfrak{AR}\overline{\mathfrak{F}}$, we have

$$\mathbb{P}(\boldsymbol{\mathcal{H}} - Z \models \mathfrak{X}) \le \mathbb{P}(\boldsymbol{h} < m') + \sum_{h \ge m'} \mathbb{P}(\boldsymbol{h} = h)\mathbb{P}(\boldsymbol{\mathcal{H}} - Z \models \mathfrak{X} | \boldsymbol{h} = h),$$

which is $n^{-\omega(1)}$ by (64) and application of Lemma 6.2 to the summands.

We thus have, in addition to Lemma 6.2,

$$\text{for } Z \text{ as in (60)}, \{\boldsymbol{\mathcal{H}} \models \mathfrak{AR}\} \stackrel{*}{\Longrightarrow} \{\boldsymbol{\mathcal{H}} - Z \models \mathfrak{F}\},$$

which with Lemma 6.3 gives

$$\{\boldsymbol{\mathcal{H}} \models \mathfrak{AR}\} \stackrel{*}{\Longrightarrow} \{\boldsymbol{\mathcal{H}} \models \mathfrak{C}\} \tag{65}$$

(since $\{\boldsymbol{\mathcal{H}} \models \mathfrak{C}\} = \{(\boldsymbol{\mathcal{H}}, Z, x) \models$ (61) for all Z as in (60) and $x \in Z\}$). Finally, using Lemma 6.4 with Lemma 6.2 and (65) gives (58) (in the form (63)). □

Proof of Lemma 6.4. We need one more property (again, for a given $\mathcal{H}$):

$\mathfrak{D}$: if $Z_0 \in \mathcal{K}$ satisfies (60) then

$$\mathsf{w}_{\mathcal{H}}(Z) \gtrsim \mathsf{w}_{\mathcal{H}}(Z_0) D^{-r} \textstyle\prod_{x \in Z_0} d_{\mathcal{H}}(x) \;\text{ for a.e. } Z \in \mathcal{K}. \tag{66}$$

The next two assertions give Lemma 6.4.

$$\text{If } \mathcal{H} \text{ satisfies } \mathfrak{RC} \text{ then it satisfies } \mathfrak{D}. \tag{67}$$

$$\text{If } \mathcal{H} \text{ satisfies } \mathfrak{RDF} \text{ then it satisfies } \mathfrak{B}. \tag{68}$$

For (67) notice that if $\mathcal{H}$ satisfies $\mathfrak{RC}$ and $Z_0 = \{x_1, \ldots, x_r\}$ satisfies (60), then induction on $i \in [r]$ shows that for a.e. choice of distinct $y_1, \ldots, y_r \in V \setminus Z_0$ we have, with $Z_i = (Z_{i-1} \setminus x_i) \cup y_i$,

$$\forall i \; \mathsf{w}_{\mathcal{H}}(Z_i) \gtrsim \mathsf{w}_{\mathcal{H}}(Z_{i-1}) d_{\mathcal{H}}(x_i)/D \gtrsim \mathsf{w}_{\mathcal{H}}(Z_0) D^{-i} \textstyle\prod_{j \le i} d_{\mathcal{H}}(x_j). \tag{69}$$

(The only thing to observe here is that (69) for Z_{i-1} (with (60) for Z_0) implies (60) for Z_i, since $\delta_{\mathcal{H}} = \Omega(D)$ (see (55)) implies that the r.h.s. of (69) is $\Omega(\mathsf{w}_{\mathcal{H}}(Z_0))$.) This gives $\mathfrak{D}$, since it implies that a.e. $Z \in \mathcal{K}$ is Z_r for some $y_1, \ldots, y_r$ supporting (69).

For (68) choose $Z_0 \in \mathcal{K}$ with $\mathsf{w}_{\mathcal{H}}(Z_0)$ maximum and note Z_0 satisfies (60) (since $\mathsf{w}_{\mathcal{H}}(Z_0)$ is at least the l.h.s. of (62)). Thus $\mathfrak{D}$ (and $\delta_{\mathcal{H}} = \Omega(D)$) give $\mathsf{w}_{\mathcal{H}}(Z) = \Omega(\mathsf{w}_{\mathcal{H}}(Z_0))$ for a.e. $Z \in \mathcal{K}$, which with $\mathfrak{F}$ implies $\Phi(\mathcal{H})/D = \Omega(\mathsf{w}_{\mathcal{H}}(Z_0))$. But this gives $\mathfrak{B}$, since $\overline{\mathsf{w}}_{\mathcal{H}}(\mathcal{H}) = \Phi(\mathcal{H})/D$ (again see (62)) and $\max_{\mathcal{H}}(\mathcal{H}) \leq \mathsf{w}_{\mathcal{H}}(Z_0)$. □

7. Proof of Lemma 6.3

We now use D for $D_{\boldsymbol{\mathcal{H}}}$ (with $\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{H}}_{n,m}$) and set $\boldsymbol{\mathcal{G}} = \boldsymbol{\mathcal{H}} - Z$, $Y = Z \setminus x$ and $W = V \setminus Z$. Notice to begin that, for any $y \in W$,

$$\mathsf{w}_{\boldsymbol{\mathcal{H}}}(Y \cup y) = \sum \{\mathsf{w}_{\boldsymbol{\mathcal{G}}}(S \cup y) : S \in \tbinom{W \setminus y}{r-1},\ S \cup x \in \boldsymbol{\mathcal{H}}\}. \tag{70}$$

Let $\boldsymbol{\mathcal{H}}' = \{A \in \boldsymbol{\mathcal{H}} : A \cap Z = \{x\}\}$ and $\boldsymbol{\mathcal{H}}'' = \boldsymbol{\mathcal{H}} \setminus \boldsymbol{\mathcal{H}}'$. We think of choosing first $\boldsymbol{\mathcal{H}}''$ (which determines $\boldsymbol{\mathcal{G}}$) and then $\boldsymbol{\mathcal{H}}'$. If $\boldsymbol{\mathcal{H}} \models \mathfrak{R}$ then (using (55) for (71) and (55)-(56) for (72))

$$D_{\boldsymbol{\mathcal{G}}} \sim D, \tag{71}$$

$$d'(x) := |\boldsymbol{\mathcal{H}}'| \ (= |\{S \in \tbinom{W}{r-1} : S \cup x \in \boldsymbol{\mathcal{H}}\}|) \sim d_{\boldsymbol{\mathcal{H}}}(x) = \Omega(D), \tag{72}$$

and (in view of (71)) $\mathfrak{F}$ for $\boldsymbol{\mathcal{G}}$ is

$$\mathsf{w}_{\boldsymbol{\mathcal{G}}}(U) \sim \Phi' := \Phi(\boldsymbol{\mathcal{G}})/D \ \text{ for a.e. } U \in \tbinom{W}{r}. \tag{73}$$

Thus Lemma 6.3 will follow from

$$\mathbb{P}((\boldsymbol{\mathcal{H}}, Z, x) \models (61) | \{\boldsymbol{\mathcal{G}} \models (73)\} \wedge \{d'(x) \models (72)\}) = 1 - n^{-\omega(1)}. \tag{74}$$

So we assume $\boldsymbol{\mathcal{H}}''$ has been chosen so that the conditioning event holds (note this *is* decided by $\boldsymbol{\mathcal{H}}''$), and proceed to choosing $\boldsymbol{\mathcal{H}}'$.

From (73) we have

$$\text{for a.e. } y \in W,\ \mathsf{w}_{\boldsymbol{\mathcal{G}}}(S \cup y) \sim \Phi' \ \text{ for a.e. } S \in \tbinom{W \setminus y}{r-1}; \tag{75}$$

so for (74) it is enough to show that if y is as in (75) then the inequality in (61) holds with probability $1 - n^{-\omega(1)}$. But for such a y, Theorem 3.2 (using (72) and $D = \Omega(\log n)$) says that with probability $1 - n^{-\omega(1)}$, $\mathsf{w}_{\boldsymbol{\mathcal{G}}}(S \cup y) \sim \Phi'$ for all but $o(d'(x))$ of the S's in (70); and whenever this is true we have (as desired)

$$\mathsf{w}_{\boldsymbol{\mathcal{H}}}(Y \cup y) \gtrsim \Phi' d'(x) \ (\sim \mathsf{w}_{\boldsymbol{\mathcal{H}}}(Z) d_{\boldsymbol{\mathcal{H}}}(x)/D).$$

8. Proof of Lemma 6.1

Here $\mathcal{H}$ is a general m-edge (n-vertex) r-graph satisfying $\mathfrak{AR}$. We again use D for $D_{\mathcal{H}}$.

Setting $p = \left(\binom{n}{r} - t\right) / \binom{n}{r}$ $(= m/\binom{n}{r})$, and using (8) and (11), we may rewrite the lower bound in (53) as

$$\frac{r-1}{r} n \log n - \Lambda - \frac{n}{r} \log[(r-1)!] + \frac{n}{r} \log p - o(n),$$

while

$$\log D = (r-1)\log n - \log[(r-1)!] + \log p + o(1).$$

Thus $\mathfrak{A}$ for $\mathcal{H}$ says

$$\log \Phi(\mathcal{H}) \ > \ \frac{n}{r} \log D - \Lambda - o(n) \ > \ \frac{1}{r} \sum \log d(v) - \Lambda - o(n), \tag{76}$$

the second inequality following from $\sum \log d(v) \leq n \log(\sum d(v)/n)$. (Here and in the rest of this argument, v runs over vertices and $d(v)$ is $d_{\mathcal{H}}(v)$.)

On the other hand, we claim that

$$\log \Phi(\mathcal{H}) < \frac{1}{r} \sum h(v, \mathcal{H}) - \Lambda + o(n). \tag{77}$$

(Recall from the third paragraph of Section 4 that $h(v, \mathcal{H})$ is the entropy of the edge containing v in a uniform p.m. of $\mathcal{H}$.)

Proof of (77). This will follow from Theorem 4.2, applied with $\mathbf{f}$ a uniform p.m. of $\mathcal{H}$ (so $H(\mathbf{f}(v)) = h(v, \mathcal{H})$), once we show

$$\textstyle\sum_v \sum_Y p_v(Y) \gamma_v(Y)^{1/(r-1)} = o(n). \tag{78}$$

(Recall from the passage preceding Theorem 4.2 that $p_v(Y) = \mathbb{P}(\mathbf{f}(v) = Y)$ and $\gamma_v(Y)$ is the probability that fewer than r edges of $\mathbf{f}$ meet $Y \cup v$. Of course here, for a given v, the only relevant Y's are those with $Y \cup v \in \mathcal{H}$, and we restrict to these in the following discussion.)

For (78) we apply Lemma 4.3 with $h(x) = x^{1/(r-1)}$, i running over pairs (v, Y), and, for $i = (v, Y)$, $p_i = p_v(Y)$ and $\gamma_i = \gamma_v(Y)$; thus $l = \sum d(v)$ $(= nD)$ and $\Delta = D$. Then (78) becomes (44), so we need (40)-(43). The first two of these are immediate and the third follows from (76) *via* Lemma 4.1:

$$\begin{aligned} \sum p_i \log(1/p_i) &= \sum \sum p_v(Y) \log(1/p_v(Y)) \\ &= \sum h(v, \mathcal{H}) \;\geq\; r \log \Phi(\mathcal{H}) \;>\; n \log \Delta - O(n) \end{aligned}$$

(with the first inequality given by Lemma 4.1 and the second by the first part of (76)).

For (43) we use (56): with κ $(= o(D))$ the largest codegree in $\mathcal{H}$ (and $\Gamma_v(Y)$ as in (27)), we have

$$\textstyle\sum \sum \gamma_v(Y) = \sum_f \mathbb{P}(\mathbf{f} = f) |\{(v, Y) : f \in \Gamma_v(Y)\}| \leq \frac{n}{r} \binom{r}{2} \kappa r = o(nD)$$

(where the third expression bounds each of the cardinalities in the preceding sum, since $f \in \Gamma_v(Y)$ iff the edge $Y \cup v$ meets some member of f more than once).

(For our random $\mathcal{H}$—as opposed to one just assumed to satisfy $\mathfrak{A}$ and $\mathfrak{R}$—this last bit is particularly crude since most codegrees will be *much* smaller than κ.)

□

Now combining (76) and (77) we have

$$\sum h(v, \mathcal{H}) > \sum \log d(v) - o(n),$$

implying (note $h(v, \mathcal{H}) \leq \log d(v)$ is trivial)

$$h(v, \mathcal{H}) > \log d(v) - o(1) \;\text{ for a.e. } v. \tag{79}$$

But Lemma 4.4 says that for any v as in (79) there is a set of $(1 - o(1)) d(v)$ edges A at v with $\mathsf{w}_{\mathcal{H}}(A) = (1 \pm o(1)) \Phi(\mathcal{H})/d(v)$ (note $p_v(Y) = \mathsf{w}_{\mathcal{H}}(Y \cup v)/\Phi(\mathcal{H})$), and combining this with (54) gives $\mathfrak{E}$.

9. Proof of Lemma 6.2

We again use $\mathcal{H}$ for a general m-edge r-graph, $\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{H}}_{n,m}$ and $D = mr/n$ $(= D_{\mathcal{H}} = D_{\boldsymbol{\mathcal{H}}})$.

By Lemma 6.1, Lemma 6.2 is the same as

$$\{\boldsymbol{\mathcal{H}} \models \mathfrak{A}\mathfrak{R}\mathfrak{E}\} \overset{*}{\Longrightarrow} \{\boldsymbol{\mathcal{H}} \models \mathfrak{F}\}. \tag{80}$$

Note that, in view of this, we may assume

$$m = |\mathcal{K}| - \Omega(|\mathcal{K}|), \tag{81}$$

since otherwise $\mathfrak{E}$ and $\mathfrak{F}$ are equivalent and (80) is vacuous. (This rather silly point will be needed for (86).)

It will be convenient to further reformulate as follows. For any $\mathcal{H}$ set

$$\alpha(\mathcal{H}) = \inf\{\alpha : |\{U \in \mathcal{K} : \mathsf{w}_{\mathcal{H}}(U) \neq (1 \pm \alpha)\Phi(\mathcal{H})/D_{\mathcal{H}}\}| < \alpha|\mathcal{K}|\}.$$

Then $\{\mathcal{H} \models \mathfrak{F}\} = \{\alpha(\mathcal{H}) = o(1)\}$ and (80) is equivalent to[2]

$$\text{for any fixed } \theta > 0,\ \mathbb{P}(\{\boldsymbol{\mathcal{H}} \models \mathfrak{A}\mathfrak{R}\mathfrak{E}\} \wedge \{\alpha(\boldsymbol{\mathcal{H}}) > 2\theta\}) = n^{-\omega(1)}. \tag{82}$$

(The 2θ will be convenient below.) So for the rest of this section we fix $\theta > 0$ and aim for (82).

Set

$$\Phi' = \Phi(\boldsymbol{\mathcal{H}})/D.$$

Notice that $\{\boldsymbol{\mathcal{H}} \models \mathfrak{E}\} \wedge \{\alpha(\boldsymbol{\mathcal{H}}) > 2\theta\}$ implies

$\mathfrak{Q}$: $\mathsf{w}_{\boldsymbol{\mathcal{H}}}(A) \sim \Phi'$ *for a.e.* $A \in \boldsymbol{\mathcal{H}}$, *but* $\mathsf{w}_{\boldsymbol{\mathcal{H}}}(U) \neq (1 \pm 2\theta)\Phi'$ *for at least a* (2θ)*-fraction of the* U*'s in* $\mathcal{K} \setminus \boldsymbol{\mathcal{H}}$.

So for (82) it is enough to show

$$\mathbb{P}(\boldsymbol{\mathcal{H}} \models \mathfrak{A}\mathfrak{R}\mathfrak{Q}) < n^{-\omega(1)}. \tag{83}$$

For the proof of this we work with an auxiliary random set $\boldsymbol{\mathcal{T}}$ chosen uniformly from $\binom{\boldsymbol{\mathcal{H}}}{\tau}$, where τ, which will be specified later (see the paragraph containing (90)-(94)), will at least satisfy

$$\log n \ll \tau \ll \log^2 n. \tag{84}$$

We take $\boldsymbol{\mathcal{F}} = \boldsymbol{\mathcal{H}} \setminus \boldsymbol{\mathcal{T}}$ and

$$\zeta = e^{-\tau/D},$$

and will be interested in a property of the pair $(\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{T}})$ (or $(\boldsymbol{\mathcal{F}}, \boldsymbol{\mathcal{T}})$), *viz.*

$\mathfrak{V}$: $\mathsf{w}_{\boldsymbol{\mathcal{F}}}(A) \sim \zeta\Phi'$ *for a.e.* $A \in \boldsymbol{\mathcal{T}}$, *but* $\mathsf{w}_{\boldsymbol{\mathcal{F}}}(U) \neq (1 \pm \theta)\zeta\Phi'$ *for at least a* θ*-fraction of the* U*'s in* $\mathcal{K} \setminus \boldsymbol{\mathcal{H}}$.

(Note $\zeta\mathsf{w}_{\boldsymbol{\mathcal{H}}}(U)$ is a natural asymptotic value for $\mathsf{w}_{\boldsymbol{\mathcal{F}}}(U)$ since each p.m. of $\boldsymbol{\mathcal{H}} - U$ survives in $\boldsymbol{\mathcal{F}}$ with probability about $(1 - \tau/m)^{n/r-1} \sim \zeta$; *cf.* (109).)

Here we exploit the familiar leverage derived from the interplay of two natural ways of generating the pair $(\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{T}})$:

[2] With $\mathfrak{G} = \{\boldsymbol{\mathcal{H}} \models \mathfrak{A}\mathfrak{R}\mathfrak{E}\}$ and $\mathfrak{H}(\nu) = \{\alpha(\boldsymbol{\mathcal{H}}) > \nu\}$, (80) says

$$\text{there is } \varsigma = o(1) \text{ such that } \mathbb{P}(\mathfrak{G} \wedge \mathfrak{H}(\varsigma)) = n^{-\omega(1)},$$

while (82) implies

$$\forall k,\ \mathbb{P}(\mathfrak{G} \wedge \mathfrak{H}(1/k)) < n^{-k} \text{ for } n \geq n_k;$$

and we get the former from the latter by taking $\varsigma(n) = (\max\{k : n_k \leq n\})^{-1}$.

(A) choose $\boldsymbol{\mathcal{H}}$ and then $\boldsymbol{\mathcal{T}}$ (as above);
(B) choose $\boldsymbol{\mathcal{F}}$ and then $\boldsymbol{\mathcal{T}}$ (determining $\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{F}} \cup \boldsymbol{\mathcal{T}}$).

Now writing simply $\mathfrak{A}$ for $\{\boldsymbol{\mathcal{H}} \models \mathfrak{A}\}$ and similarly for $\mathfrak{R}$ and $\mathfrak{Q}$, and $\mathfrak{V}$ for $\{(\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{T}}) \models \mathfrak{V}\}$, we will show

$$\mathbb{P}(\mathfrak{V}|\mathfrak{A}\mathfrak{R}\mathfrak{Q}) > 1 - o(1) \tag{85}$$

and

$$\mathbb{P}(\mathfrak{V}) = n^{-\omega(1)}. \tag{86}$$

These give (83), since

$$\mathbb{P}(\mathfrak{A}\mathfrak{R}\mathfrak{Q}) = \mathbb{P}(\mathfrak{A}\mathfrak{R}\mathfrak{Q}\mathfrak{V})/\mathbb{P}(\mathfrak{V}|\mathfrak{A}\mathfrak{R}\mathfrak{Q}) \leq \mathbb{P}(\mathfrak{V})/\mathbb{P}(\mathfrak{V}|\mathfrak{A}\mathfrak{R}\mathfrak{Q}).$$

(So (85) is more than is needed here.) We first dispose of the easier (86).

Proof of (86). Here we use viewpoint (B). The (natural) idea is: $\boldsymbol{\mathcal{F}}$ determines the weights $\mathsf{w}_{\boldsymbol{\mathcal{F}}}(U)$ (for all $U \in \mathcal{K}$, though here we are only interested in $U \in \mathcal{K} \setminus \boldsymbol{\mathcal{F}}$), and $\mathfrak{V}$ then requires that $\boldsymbol{\mathcal{T}}$ be (pathologically) drawn almost entirely from U's with weights close to $\zeta\Phi'$, though this group excludes an $\Omega(1)$-fraction of $\mathcal{K} \setminus \boldsymbol{\mathcal{F}}$.

A small complication is that $\boldsymbol{\mathcal{F}}$ doesn't determine Φ'. Among several ways of dealing with this, the following seems nicest. Given $\boldsymbol{\mathcal{F}}$, let $U_1, \ldots$ be an ordering of $\mathcal{K} \setminus \boldsymbol{\mathcal{F}}$ with $\mathsf{w}_{\boldsymbol{\mathcal{F}}}(U_1) \leq \mathsf{w}_{\boldsymbol{\mathcal{F}}}(U_2) \leq \cdots$, and let $\mathcal{Y}$ and $\mathcal{Z}$ be (resp.) the first and last $\theta|\mathcal{K} \setminus \boldsymbol{\mathcal{F}}|/3$ of the U_i's. Then, *whatever* Φ' turns out to be, the second part of $\mathfrak{V}$ requires that at least one of $\mathcal{Y}$, $\mathcal{Z}$ be contained in

$$\mathcal{W} := \{U : \mathsf{w}_{\boldsymbol{\mathcal{F}}}(U) \neq (1 \pm \theta)\zeta\Phi'\}$$

(or (81), with $\tau \ll m$, implies $|\mathcal{W} \setminus \boldsymbol{\mathcal{H}}| < |\mathcal{Y}| + |\mathbb{Z}| < 2\theta(|\mathcal{K} \setminus \boldsymbol{\mathcal{H}}| + \tau)/3 < \theta|\mathcal{K} \setminus \boldsymbol{\mathcal{H}}|$). But if this is true then the first part of $\mathfrak{V}$ requires that (say)

$$\min\{|\boldsymbol{\mathcal{T}} \cap \mathcal{Y}|, |\boldsymbol{\mathcal{T}} \cap \mathcal{Z}|\} < \theta\tau/4; \tag{87}$$

and, since

$$\mathbb{E}|\boldsymbol{\mathcal{T}} \cap \mathcal{Y}| = \mathbb{E}|\boldsymbol{\mathcal{T}} \cap \mathcal{Z}| = \theta\tau/3$$

(and θ is fixed), Theorem 3.1 bounds the probability of (87) by $e^{-\Omega(\tau)}$, which is $n^{-\omega(1)}$ by (84). □

Proof of (85). We now need to pay some attention to parameters. We first observe that if $\mathcal{H} \models \mathfrak{A}\mathfrak{R}$, then there is $\gamma = o(1)$ (depending on the $o(n)$ in $\mathfrak{A}$ and, in $\mathfrak{R}$, the (explicit or implicit) $o(\cdot)$'s in (54) and (56), and the implied constants in (55)), such that for each $U \in \mathcal{K}$ with (say)

$$(\mathsf{w}_{\mathcal{H}}(U) =)\ \Phi(\mathcal{H} - U) > \Phi(\mathcal{H})n^{-r}, \tag{88}$$

$\mathcal{H}^* := \mathcal{H} - U$ and $\Phi^* := \Phi(\mathcal{H}^*)$ satisfy

$$\sum\{\mathsf{w}_{\mathcal{H}^*}(A) : A \in \mathcal{H}^*, \mathsf{w}_{\mathcal{H}^*}(A) \neq (1 \pm \gamma)\Phi^*/D\} < \gamma n\Phi^*. \tag{89}$$

To see this, notice that each relevant $\mathcal{H}^*$ satisfies $\mathfrak{A}\mathfrak{R}$ (see (57) for $\mathfrak{R}$), so also $\mathfrak{C}$ by Lemma 6.1. But then $\mathcal{H}^*$ contains $(1 - o(1))|\mathcal{H}^*| \sim nD/r$ edges of $\mathsf{w}_{\mathcal{H}^*}$-weight $(1 \pm o(1))\Phi^*/D_{\mathcal{H}^*} \sim \Phi^*/D$, with (both) asymptotics following easily from $\mathcal{H} \models \mathfrak{R}$ (see (55)); so such edges account for all but a $o(1)$-fraction of the total weight $\Phi^*(n - r)/r \sim \Phi^*n/r$. This gives (89) for a suitable $\gamma = o(1)$.

We now choose $\tau = \nu \log n$—noting that then

$$\zeta\ (= e^{-\tau/D})\ > e^{-\nu} \tag{90}$$

(since $D > \log n$)—together with M and η, satisfying

$$\log n \gg \nu \gg 1 \tag{91}$$

(which is (84));

$$e^{-\nu} \gg \gamma; \tag{92}$$

$$\tau \gg M \begin{cases} \gg \gamma\tau, \\ > 1+\gamma; \end{cases} \tag{93}$$

and

$$e^{-\nu} \gg \eta \gg \sqrt{\tau M}/\log n. \tag{94}$$

Note this is possible: we may choose $\nu \to \infty$ as slowly as we like (which in particular gives (91) and (92)); we then want to choose M as in (93) satisfying (to leave room for η) $e^{-\nu} \gg \sqrt{\tau M}/\log n$; and this is possible if $e^{-\nu} \gg \max\{\nu\sqrt{\gamma}, \sqrt{\nu/\log n}\}$, which is true for a slow enough ν.

For the proof of (85) we use viewpoint (A) (choose $\boldsymbol{\mathcal{H}}$, then $\boldsymbol{\mathcal{T}}$). We assume we have chosen $\boldsymbol{\mathcal{H}} = \mathcal{H}$ satisfying $\mathfrak{A}\mathfrak{R}\mathfrak{Q}$; so $\mathbb{P}$ now refers just to the choice of $\boldsymbol{\mathcal{T}}$, and (85) will follow from

$$\mathbb{P}((\mathcal{H}, \boldsymbol{\mathcal{T}}) \models \mathfrak{V}) = 1 - o(1). \tag{95}$$

It will be enough to show that for $U \in \mathcal{K}$ as in (88) (i.e. $\mathsf{w}_{\mathcal{H}}(U) > \Phi(\mathcal{H})n^{-r}$),

$$\mathbb{P}(\mathsf{w}_{\boldsymbol{\mathcal{F}}}(U) \sim \zeta\mathsf{w}_{\mathcal{H}}(U)) = 1 - o(1) \quad \text{if } U \in \mathcal{K} \setminus \mathcal{H}, \tag{96}$$

$$\mathbb{P}(\mathsf{w}_{\boldsymbol{\mathcal{F}}}(U) \sim \zeta\mathsf{w}_{\mathcal{H}}(U) | U \in \boldsymbol{\mathcal{T}}) = 1 - o(1) \qquad \text{if } U \in \mathcal{H}. \tag{97}$$

Before proving this we show that it does give (95). If $\mathcal{H}$ satisfies $\mathfrak{Q}$ then for a suitable $\varsigma = o(1)$,

$$|\{A \in \mathcal{H} : \mathsf{w}_{\mathcal{H}}(A) \neq (1 \pm \varsigma)\Phi'\}| \ll |\mathcal{H}|. \tag{98}$$

Thus, with $\mathcal{H}^0$ the set in (98), we have $\mathbb{E}|\boldsymbol{\mathcal{T}} \cap \mathcal{H}^0| = \tau|\mathcal{H}^0|/|\mathcal{H}| \ll \tau$, so

$$|\boldsymbol{\mathcal{T}} \cap \mathcal{H}^0| \ll \tau \ \text{ w.h.p.}$$

(by Theorem 3.1 or just Markov's Inequality). But for the first part of $\mathfrak{V}$ to fail we must have either $|\boldsymbol{\mathcal{T}} \cap \mathcal{H}^0| = \Omega(\tau)$, which we have just said occurs with probability $o(1)$, or

$$|\{A \in \boldsymbol{\mathcal{T}} \setminus \mathcal{H}^0 : \mathsf{w}_{\boldsymbol{\mathcal{F}}}(A) \not\sim \zeta\mathsf{w}_{\mathcal{H}}(A)\}| = \Omega(\tau),$$

which has probability $o(1)$ by (97) (and Markov).

Similarly, failure of the second part of $\mathfrak{V}$ implies

$$\mathsf{w}_{\boldsymbol{\mathcal{F}}}(U) = (1 \pm \theta)\zeta\Phi' \not\sim \zeta\mathsf{w}_{\mathcal{H}}(U) \tag{99}$$

for at least $\theta|\mathcal{K} \setminus \mathcal{H}|$ of those U's in the second part of $\mathfrak{Q}$ that satisfy

$$\mathsf{w}_{\mathcal{H}}(U) > (1-\theta)\zeta\Phi' > n^{-o(1)}\Phi(\mathcal{H})/D \tag{100}$$

(since those failing (100) *cannot* satisfy (99); for the second bound in (100) see (90) and (91)). But since the bound in (100) is larger than the one in (88), (96) implies that the probability that (99) holds for such a set of U's is $o(1)$.

Finally, we prove (96); the proof of (97) is almost literally the same and is omitted. (Note the probability in (97) is just $\mathbb{P}(\mathsf{w}_{\mathcal{H}\setminus\boldsymbol{\mathcal{T}}_0}(U) \sim \zeta\mathsf{w}_{\mathcal{H}}(U))$, with $\boldsymbol{\mathcal{T}}_0$ uniform from $\binom{\mathcal{H}\setminus\{U\}}{\tau-1}$.)

Proof of (96). We now fix U as in (88) (and recall $\mathcal{H}^* = \mathcal{H} - U$ and $\Phi^* = \Phi(\mathcal{H}^*)$).

Say $A \in \mathcal{H}$ is *heavy* if $A \in \mathcal{H}^*$ and $\mathsf{w}_{\mathcal{H}^*}(A) > M\Phi^*/D$, and note that by (89) (and $M > 1+\gamma$ from (93)),

(101) the number of heavy edges in $\mathcal{H}$ is less than $\gamma nD/M = \gamma mr/M$,

implying

(102) $$\mathbb{P}(\mathcal{T} \text{ contains a heavy edge}) < \gamma\tau r/M = o(1)$$

(see (93)). So it is enough to show (96) conditioned on

(103) $$\{\mathcal{T} \text{ contains no heavy edges}\}.$$

We will instead show a slight variant, replacing $\mathcal{T}$ by $\mathcal{T}' = \{A_1, \ldots, A_\tau\}$, with the A_i's chosen uniformly and *independently* from the non-heavy edges of $\mathcal{H}$; thus:

(104) $$\mathbb{P}(\mathsf{w}_{\mathcal{H}\setminus\mathcal{T}'}(U) \sim \zeta \mathsf{w}_{\mathcal{H}}(U)) = 1 - o(1).$$

Of course this suffices: we may couple $\mathcal{T}$ (conditioned on (103)) and $\mathcal{T}'$ so they agree whenever the edges of $\mathcal{T}'$ are distinct, which occurs w.h.p. (more precisely, with probability at least $1 - \tau^2/m$), and the probability in (96) is then at least the probability in (104) minus $\mathbb{P}(\mathcal{T}' \neq \mathcal{T})$.

For the proof of (104), let

$$X = X(A_1, \ldots, A_\tau) = \Phi(\mathcal{H}^* \setminus \{A_1, \ldots, A_\tau\}) = \mathsf{w}_{\mathcal{H}\setminus\mathcal{T}'}(U).$$

Since $\eta \ll \zeta$ (see (90) and (94)), (104) will follow from (recall $\mathsf{w}_{\mathcal{H}}(U) = \Phi^*$)

(105) $$\mathbb{E}X \sim \zeta\Phi^*$$

and

(106) $$\mathbb{P}(|X - \mathbb{E}X| > \eta\Phi^*) = o(1).$$

Proof of (105). Let M_i run through the p.m.s of $\mathcal{H}^*$ and let x_i be the number of heavy edges in M_i. Then with m' the number of non-heavy edges in $\mathcal{H}$, we have

$$\mathbb{E}X = \textstyle\sum_i (1 - (n/r - 1 - x_i)/m')^\tau$$

and, by (89),

(107) $$\sum x_i = \sum \{\mathsf{w}_{\mathcal{H}^*}(A) : A \in \mathcal{H}^*, A \text{ heavy}\} < \gamma n\Phi^*.$$

These imply, with $\varrho = (n/r - 1)/m'$,

$$\begin{aligned} (1-\varrho)^\tau \Phi^* \quad &\leq \quad \mathbb{E}X \ < \ \textstyle\sum_i e^{-(\varrho - x_i/m')\tau} \\ &< \quad \left[e^{-\varrho\tau} + \gamma rn/(n-r)\right]\Phi^*. \end{aligned} \tag{108}$$

Here the last inequality follows from (107) and convexity of the exponential function, which imply that the sum in (108) is at most what it would be with $\frac{\gamma n\Phi^*}{n/r-1} = \frac{\gamma rn\Phi^*}{n-r}$ of the x_i's equal to $n/r - 1$ and the rest (the number of which we just bound by Φ^*) equal to zero.

In view of (108), (105) will follow from

(109) $$(1-\varrho)^\tau \sim e^{-\varrho\tau} \sim e^{-\tau/D} \ (= \zeta)$$

(and $\gamma \ll \zeta$, which is given by (90) and (92)). For the two parts of (109) we need (resp.) $\varrho^2 \ll 1/\tau$ and $|\varrho - 1/D| \ll 1/\tau$. The first of these follows from (101) (which gives $m' \sim m$, though here $m' = \Omega(m)$ would suffice) and (84). For the second, now using (101) more precisely (and recalling $D = mr/n$), we have

$$\left|\frac{n/r-1}{m'} - \frac{n/r}{m}\right| \le \frac{1}{m'} + \frac{n}{r}\frac{m-m'}{mm'} < \frac{1}{m'} + \frac{n}{r}\frac{\gamma r}{Mm'} \ll \frac{1}{\tau},$$

with the last inequality a (weak) consequence of (93). □

Proof of (106). We consider the (Doob) martingale

$$X_i = X_i(A_1, \dots, A_i) = \mathbb{E}[X|A_1, \dots, A_i] \ (i = 0, \dots, \tau), \tag{110}$$

with difference sequence $Z_i = X_i - X_{i-1}$ ($i \in [\tau]$) and $Z = \sum Z_i$ ($= X - \mathbb{E}X$). For the next little bit we use $\mathbb{E}_S$ for expectation with respect to $(A_i : i \in S)$.

Given $A_1, \dots, A_{i-1}$ we may express

$$Z_i = \mathbb{E}W - W, \tag{111}$$

where $\mathbb{E}$ refers to A chosen uniformly from the non-heavy edges of $\mathcal{H}$ and

$$\begin{aligned} W(A) = {} & \mathbb{E}_{[i+1,\tau]}\Phi(\mathcal{H}^* \setminus \{A_1, \dots, A_{i-1}, A_{i+1}, \dots, A_\tau\}) \\ & - \mathbb{E}_{[i+1,\tau]}\Phi(\mathcal{H}^* \setminus \{A_1, \dots, A_{i-1}, A, A_{i+1}, \dots, A_\tau\}). \end{aligned} \tag{112}$$

For (111) just notice that

$$X_i(A_1, \dots, A_{i-1}, A) = \mathbb{E}_{[i+1,\tau]}\Phi(\mathcal{H}^* \setminus \{A_1, \dots, A_{i-1}, A, A_{i+1}, \dots, A_\tau\}),$$

while

$$X_{i-1}(A_1, \dots, A_{i-1}) = \mathbb{E}_{[i,\tau]}\Phi(\mathcal{H}^* \setminus \{A_1, \dots, A_{i-1}, A_i, A_{i+1}, \dots, A_\tau\}).$$

(The first term on the r.h.s. of (112), which is chosen to give (113), doesn't depend on A so doesn't affect (111).)

We also have

$$0 \le W(A) \le \mathsf{w}_{\mathcal{H}^*}(A), \tag{113}$$

since these bounds hold even if we remove the $\mathbb{E}$'s in (112). Thus W satisfies the conditions in Proposition 3.4 with $b = M\Phi^*/D$ and $a = \Phi^*/D$ (the latter since $|\mathcal{H}|^{-1}\sum_{A\in\mathcal{H}} \mathsf{w}_{\mathcal{H}^*}(A) = |\mathcal{H}|^{-1}\Phi^*(n/r-1) < \Phi^*/D$—note $\mathsf{w}_{\mathcal{H}^*}(A) := 0$ if $A \notin \mathcal{H}^*$—and averaging instead only over non-heavy edges can only decrease this). So for any

$$\vartheta \in [0, (2b)^{-1}], \tag{114}$$

we may apply Lemma 3.3 to each of Z, $-Z$, using Proposition 3.4 (with (111)) to bound the factors in (24), yielding

$$\max\{\mathbb{E}e^{\vartheta Z}, \mathbb{E}e^{-\vartheta Z}\} \le e^{\tau\vartheta^2 ab} = \exp[\tau\vartheta^2 M(\Phi^*/D)^2]$$

and, for any $\lambda > 0$,

$$\max\{\mathbb{P}(Z > \lambda), \mathbb{P}(Z < -\lambda)\} < \exp[\tau\vartheta^2 M(\Phi^*/D)^2 - \vartheta\lambda]. \tag{115}$$

For (106) we use (115) with $\lambda = \eta\Phi^*$ and

$$\vartheta = \min\left\{\tfrac{\eta\Phi^*}{2\tau M(\Phi^*/D)^2}, \tfrac{D}{2M\Phi^*}\right\} = \tfrac{D}{2M\Phi^*}\min\left\{\tfrac{\eta D}{\tau}, 1\right\} \tag{116}$$

(the first value in "min" minimizes the r.h.s. of (115) and the second enforces (114)), and should show that the exponent in (115) is then $-\omega(1)$.

Suppose first that $\eta D \leq \tau$, so ϑ takes the first value(s) in (116). Then the negative of the exponent in (115) is (using (94) and $D \geq \log n$)

$$\frac{(\eta\Phi^*)^2}{4\tau M(\Phi^*/D)^2} = \frac{\eta^2 D^2}{4\tau M} = \omega(1).$$

If instead $\eta D > \tau$, then $\vartheta = D/(2M\Phi^*)$ and the exponent in (115) is

$$\frac{D^2}{(2M\Phi^*)^2}\tau M\left(\frac{\Phi^*}{D}\right)^2 - \frac{D}{2M\Phi^*}\eta\Phi^* = \frac{\tau}{4M} - \frac{\eta D}{2M} < -\frac{\tau}{4M} = -\omega(1),$$

where we used the assumed $\eta D > \tau$ and, from (93), $\tau \gg M$. □

□

□

10. Reduction

In this section we derive Theorem 1.6 from the following statement, which will be proved in [22].

Theorem 10.1. *Fix a small positive ε and suppose $\delta_x \sim \varepsilon \log n$ for each $x \in W := [n]$. Let $M \sim (n/r)\log n$ and let $\boldsymbol{\mathcal{H}}^*$ be distributed as $\boldsymbol{\mathcal{H}}_{n,M}$ conditioned on*

$$\{d_{\boldsymbol{\mathcal{H}}}(x) \geq \delta_x \ \forall x \in W\}. \tag{117}$$

Then w.h.p.

$$\Phi(\boldsymbol{\mathcal{H}}^*) > \left[e^{-(r-1)}\log n\right]^{n/r} e^{-o(n)}.$$

In other words: for $\varsigma \ll 1$ there is $\varrho \ll 1$ such that if $M = (1 \pm \varsigma)(n/r)\log n$ and $\delta_x = (1 \pm \varsigma)\varepsilon \log n$ for each x, then

$$\Pr\left(\Phi(\boldsymbol{\mathcal{H}}^*) \leq \left[e^{-(r-1)}\log n\right]^{n/r} e^{-\varrho n}\right) < \varrho.$$

(The n here will not be exactly the one in Theorem 1.6, and will be renamed n' when we come to use it.)

For the rest of this section $\boldsymbol{\mathcal{H}}_T$ is as in Theorem 1.6. Our discussion through Lemma 10.2 is adapted from [12].

We employ the following standard device for handling the process $\{\boldsymbol{\mathcal{H}}_t\}$ of Theorems 1.3 and 1.6. Let ξ_A, $A \in \mathcal{K}$, be independent random variables, each uniform from $[0,1]$, and set $\boldsymbol{\mathcal{G}}_\lambda = \{A \in \mathcal{K} \ : \ \xi_A \leq \lambda\}$. Members of $\boldsymbol{\mathcal{G}}_\lambda$ are λ*-edges* and we use d_λ for degree in $\boldsymbol{\mathcal{G}}_\lambda$. Of course with probability one the ξ_A's are distinct. If they *are* distinct—which we assume henceforth—they define the discrete process $\{\boldsymbol{\mathcal{H}}_t\}$ in the natural way (add edges A in the order in which the ξ_A's appear in $[0,1]$).

Fix a small positive ε. Let $\delta_0 = \lfloor \varepsilon \log n \rfloor$, let g be a suitably slow $\omega(1)$, and set:

$$\Lambda = \min\{\lambda : \boldsymbol{\mathcal{G}}_\lambda \text{ has no isolated vertices}\}$$

(so $\boldsymbol{\mathcal{H}}_T = \boldsymbol{\mathcal{G}}_\Lambda$);

$$\sigma = \frac{\log n - g(n)}{\binom{n-1}{r-1}} \text{ and } \beta = \frac{\log n + g(n)}{\binom{n-1}{r-1}};$$

$$W_\sigma = \{v \in [n] \ : \ d_\sigma(v) < \delta_0\};$$

and

$$Y = W_\sigma \cup \bigcup\{A : A \in \boldsymbol{\mathcal{G}}_\beta, A \cap W_\sigma \neq \emptyset\}.$$

Parts (b) and (c) of the next lemma are (a) and (c) of Lemma 5.1 in [12].

Lemma 10.2. *With the above setup, w.h.p.*

(a) $|W_\sigma| < n^{2\alpha}$*, with* $\alpha \sim \varepsilon \log(e/\varepsilon)$*;*

(b) $\Lambda \in (\sigma, \beta)$*;*

(c) *in* $\boldsymbol{\mathcal{G}}_\beta$*, no edge meets* W_σ *more than once and no* $u \notin W_\sigma$ *lies in more than one edge meeting* $Y \setminus \{u\}$*.*

Remarks. Once $\Lambda < \beta$ as in (b), the initial W_σ in the definition of Y is superfluous. For Theorem 1.3, $|W_\sigma| = o(n)$ in (a) would suffice, but for Theorem 1.6 we need a little more (precisely, $|W_\sigma| = o(n/\log\log n)$), to make the bound on Φ in Theorem 10.1 (in which, again, $|W|$ will not be exactly the present n) an instance of (5); see (128).

Proof of (a). Since (for any v) $d_\sigma(v)$ is binomial with mean $\mu := \binom{n-1}{r-1}\sigma \sim \log n$, Theorem 3.1 gives

$$\mathbb{P}(v \in W_\sigma) < \exp[-\mu\varphi(-(1-\varepsilon-o(1)))] < n^{-1+\alpha}, \tag{118}$$

with α as in (a); and (a) then follows *via* Markov's Inequality. □

It will now be convenient to fix some linear ordering "$\prec$" of $\mathcal{K}$. Choose (and condition on)

$$W_\sigma,$$

$$\{A \in \boldsymbol{\mathcal{G}}_\sigma : A \cap W_\sigma \neq \emptyset\}, \tag{119}$$

and the *ordering* of $\{\xi_A : A \cap W_\sigma \neq \emptyset, \xi_A > \sigma\}$.

Notice that W_σ and the set in (119) are enough to tell us whether $\Lambda > \sigma$, which by Lemma 10.2 holds w.h.p. If it does hold—which we now assume—then the above choices determine $\{A \in \boldsymbol{\mathcal{G}}_\Lambda : A \cap W_\sigma \neq \emptyset\}$, so in particular, for each $x \in W_\sigma$, the first (under $\prec$) Λ-edge, say A_x, containing x. (They do not determine Λ, but we don't need this and avoid conditioning on a zero-probability event.)

By Lemma 10.2, w.h.p.

$$|W_\sigma| < n^{2\alpha} \text{ and the } A_x\text{'s are distinct and disjoint} \tag{120}$$

(if $\Lambda < \beta$, as in (b) of the lemma, then the A_x's are all in $\boldsymbol{\mathcal{G}}_\beta$, so the second part of (120) is contained in (c)); so we assume these properties and set $U = \cup_{x \in W_\sigma} A_x \setminus W_\sigma$.

Next, choose (and condition on)

$$\{A \in \boldsymbol{\mathcal{G}}_\sigma : A \cap U \neq \emptyset = A \cap W_\sigma\} \tag{121}$$

(from (119) we already know the members of $\boldsymbol{\mathcal{G}}_\sigma$ that do meet W_σ). Set

$$W = V \setminus (W_\sigma \cup U),\ n' = |W| > n - rn^{2\alpha} \tag{122}$$

(using the first part of (120)), and

$$\boldsymbol{\mathcal{H}}^* = \boldsymbol{\mathcal{G}}_\sigma[W]$$

(meaning, as for graphs, the set of edges of $\mathcal{G}_\sigma$ contained in W). Again by Lemma 10.2, w.h.p.

$$\text{no vertex of } W \text{ lies in more than one } \sigma\text{-edge meeting } W_\sigma \cup U, \tag{123}$$

and we add this assumption to those above.

Since $\Phi(\mathcal{H}_T) = \Phi(\mathcal{G}_\Lambda) \geq \Phi(\mathcal{H}^*)$, Theorem 1.6 will follow from

$$\text{w.h.p. } \Phi(\mathcal{H}^*) > \left[e^{-(r-1)} \log n\right]^{n/r} e^{-o(n)}. \tag{124}$$

We will get this from Theorem 10.1.

For $x \in W$ let

$$\delta_x = \delta_0 - |\{A \in \mathcal{G}_\sigma : x \in A, A \cap (W_\sigma \cup U) \neq \emptyset\}| \in \{\delta_0, \delta_0 - 1\} \tag{125}$$

(with the membership assertion given by (123)), and notice that

$$\mathcal{H}^* \textit{ is distributed as } \mathcal{H}_{W,\sigma} \textit{ conditioned on } \mathfrak{L} := \{d(x) \geq \delta_x \ \forall x \in W\} \tag{126}$$

(where $\mathcal{H}_{W,\sigma}$—and $\mathcal{H}_{W,m}$ below—have the obvious meanings and d is degree in $\mathcal{H}_{W,\sigma}$).

To complete the reduction to Theorem 10.1 we then just want to replace $\mathcal{H}_{W,\sigma}$ by a suitable combination of $\mathcal{H}_{W,m}$'s. (Note (125) says the δ_x's are as in the theorem.)

By (118) and Harris' Inequality [18] we have

$$\mathbb{P}(\mathcal{H}_{W,\sigma} \models \mathfrak{L}) > (1 - n^{-1+\alpha})^{n'} \ (\sim \exp[-n^\alpha]), \tag{127}$$

with $\alpha \sim \varepsilon \log(e/\varepsilon)$ (as in (118), the substitution of n' for n and δ_x for δ_0 having no significant effect). On the other hand, with

$$\mu := \mathbb{E}|\mathcal{H}_{W,\sigma}| = \tbinom{n'}{r}\sigma \ (\sim (n/r)\log n),$$

$\gamma = n^{-1/3}$ (say), and $I = ((1-\gamma)\mu, (1+\gamma)\mu)$, Theorem 3.1 gives

$$\mathbb{P}(|\mathcal{H}_{W,\sigma}| \notin I) < \exp[-\Omega(n^{1/3})],$$

which with (127) implies $\mathbb{P}(|\mathcal{H}_{W,\sigma}| \notin I | \mathfrak{L}) < \exp[-\Omega(n^{1/3})]$.

According to Theorem 10.1 there is

$$\Phi^* \ > \ [e^{-(r-1)} \log(n')]^{n'/r} e^{-o(n')} \ = \ [e^{-(r-1)} \log n]^{n/r} e^{-o(n)} \tag{128}$$

(the equality follows easily from (122)) such that

$$\max_{m \in I} \mathbb{P}(\Phi(\mathcal{H}_{W,m}) < \Phi^* | \mathfrak{L}) = o(1).$$

So, finally,

$$\begin{aligned}
\mathbb{P}(\Phi(\mathcal{H}_{W,\sigma}) < \Phi^* | \mathfrak{L}) &= \textstyle\sum_m \mathbb{P}(|\mathcal{H}_{W,\sigma}| = m | \mathfrak{L}) \mathbb{P}(\Phi(\mathcal{H}_{W,m}) < \Phi^* | \mathfrak{L}) \\
&< \max_{m \in I} \mathbb{P}(\Phi(\mathcal{H}_{W,m}) < \Phi^* | \mathfrak{L}) + o(1) \\
&= o(1),
\end{aligned}$$

which gives (124). ☐

Acknowledgment. Thanks to Eyal Lubetzky for stimulating conversations.

References

[1] N. Alon and S. Friedland, The maximum number of perfect matchings in graphs with given degree sequence, *Elec. J. Combin.* **15** (2008). 9

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

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

[4] B. Bollobás and A. Thomason, Random graphs of small order, pp. 47-97 in *Random graphs '83*, North-Holland, Amsterdam, 1985. 2

[5] B. Bollobás and A. Thomason, Threshold functions, *Combinatorica* **7** (1987), 35-38. 2

[6] L. Brégman, Some properties of nonnegative matrices and their permanents, *Math. Doklady* **14** (1973), 945-949. 9

[7] 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. 9

[8] 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. 2

[9] T.M. Cover and J.A. Thomas, *Information Theory*, Wiley, New York, 1991. 9, 10

[10] W. Cuckler and J. Kahn, Entropy bounds for perfect matchings and Hamiltonian cycles, *Combinatorica* **29** (2009), 327-335. 9, 10

[11] J. Cutler and A.J. Radcliffe, An entropy proof of the Kahn-Lovász Theorem, *Elec. J. Comb.* **18** (2011), Issue 1. 9

[12] P. Devlin and J. Kahn, Perfect fractional matchings in k-out hypergraphs, *Elec. J. Comb.* **24** (2017), Issue 3. 24, 25

[13] P. Erdős, On the combinatorial problems which I would most like to see solved, *Combinatorica* **1** (1981), 25-42. 1, 2

[14] P. Erdős and A. Rényi, On the evolution of random graphs, *Publ. Math. Inst. Hungar. Acad. Sci.* **5** (1960), 17-61. 2

[15] P. Erdős and A. Rényi, On the existence of a factor of degree one of a connected random graph, *Acta Math. Acad. Sci. Hungar.* **17** (1966), 359-368. 2

[16] A. Frieze and S. Janson, Perfect matchings in random s-uniform hypergraphs, *Random Structures & Algorithms* **7** (1995), 41-57. 2

[17] A. Frieze and M. Karonski, *Introduction to Random Graphs,* Cambridge Univ. Pr., Cambridge, 2016. 2

[18] T.E. Harris, A lower bound on the critical probability in a certain percolation process, *Proc. Cam. Phil. Soc.* **56** (1960), 13-20. 26

[19] A. Heckel, Random triangles in random graphs, arXiv:1802.08472 [math.CO] 4

[20] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*, Wiley, New York, 2000. 1, 2, 3, 4, 7

[21] A. Johansson, J. Kahn and V. Vu, Factors in random graphs *Random Structures & Algorithms* **33** (2008), 1-28. 2, 3, 4, 5, 7, 9

[22] J. Kahn, Hitting times for Shamir's Problem, in preparation. 3, 24

[23] J.H. Kim, Perfect matchings in random uniform hypergraphs, *Random Structures & Algorithms* **23** (2003), 111-132. 2

[24] M. Krivelevich, Perfect fractional matchings in random hypergraphs *Random Structures & Algorithms* **9** (1996), 317-334. 2

[25] O. Riordan, Random cliques in random graphs, arXiv:1802.01948 [math.CO] 4

[26] A. Ruciński, Open problem, in *Random Graphs (2)*, Proceedings, Poznań, 1989, A.M. Frieze and T. Łuczak, eds., John Wiley & Sons, New York, 284. 4

[27] J. Schmidt and E. Shamir, A threshold for perfect matchings in random d-pure hypergraphs, *Disc. Math.* **45** (1983), 287-295. 1, 2

11. Appendix: generics

Here we prove (20). We regard this item as a necessary (actually, for anyone who's gotten this far surely *un*necessary) evil and aim to be brief.

With $D_t = D_{\mathcal{H}_t}$, (20) says

$$\text{w.h.p. } \mathcal{H}_t \text{ satisfies (54)-(56) with } D = D_t \text{ for all } t \leq T. \tag{129}$$

For (54), (56) and the upper bound in (55), a naive union bound will suffice here, as failure probabilities for individual t's are very small. A little more care is needed for the lower bound in (55), since for t near T we can only say

$$\mathbb{P}(\mathcal{H}_t \text{ violates (55)}) < n^{-\alpha} \tag{130}$$

with α some small (positive) constant depending on ε. But even this is enough: with $M_t = \binom{n}{r} - t \, (= |\mathcal{H}_t|)$ and

$$I = \{t : M_t = 2^i M \text{ for some } i \in \{0, 1, \ldots\}\}$$

(recall $M = M_T$), (55) holds for all $t \le T$ if it holds for all $t \in I$, since then for $t' = \min\{l \in I : l \ge t\}$ we have $D_t \le 2D_{t'}$ and $\delta_{\mathcal{H}_t} \ge \delta_{\mathcal{H}_{t'}} = \Omega(D_{t'}) = \Omega(D_t)$; and (130) gives $\sum_{t\in I} \mathbb{P}(\mathcal{H}_t \text{ violates } (55)) = O(n^{-\alpha}\log n)$.

We proceed to failure probabilities, now writing d_t for degree in $\mathcal{H}_t$ and beginning with (54). For $W \subseteq V$, we have

$$\xi(W) := \textstyle\sum_{v\in W} d_t(v) = \sum_{j=1}^r j\xi_j(W),$$

where $\xi_j(W) := |\{A \in \mathcal{H}_t : |A \cap W| = j\}|$ is hypergeometric with mean

$$\frac{\binom{|W|}{j}\binom{n-|W|}{r-j}}{\binom{n}{r}} M_t.$$

If $\theta = \theta(n) = (\log n)^{-1/3}$ (say) and $|W| = \theta n$, then $\mathbb{E}\xi_j(W) \sim \theta^j\binom{r}{j} M_t$ and

$$\mu := \mathbb{E}\xi(W) \sim \mathbb{E}\xi_1(W) \sim \theta r M_t = \theta n D_t > \theta n \log n;$$

so for $\lambda = \theta\mu$, Theorem 3.1 gives $\mathbb{P}(|\xi_j(W) - \mathbb{E}\xi_j(W)| > \lambda) < \exp[-\Omega(\theta^2\mu)]$ for $j \in [r]$ (with the true value much smaller if $j \ne 1$), implying

$$\mathbb{P}(|\xi(W) - \mu| > r\lambda) < \exp[-\Omega(\theta^2\mu)] = \exp[-\Omega(\theta^3 n \log n)]. \tag{131}$$

But if (54) fails (for $\mathcal{H}_t$) then there *must* be some $W \in \binom{V}{\theta n}$ with $|\xi(W) - \mu| > r\lambda$, and (131) bounds the probability that this happens by

$$\binom{n}{\theta n} e^{-\Omega(\theta^3 n\log n)} < \exp[\theta n \log(e/\theta) - \Omega(\theta^3 n \log n)] = e^{-\Omega(n)}\ (= n^{-\omega(1)}).$$

This gives (54).

For (55) we apply Theorems 3.1 and 3.2 to the $d_t(v)$'s, each of which is hypergeometric with mean $D_t > (1+\varepsilon)\log n$. For the upper bound, Theorem 3.2 gives (say)

$$\mathbb{P}(d_t(v) > 3rD_t) < \exp[-3rD_t\log(3r/e)] < n^{-3r}$$

so the probability that some $d_t(v)$ exceeds $3rD_t$ is less than n^{-2r-1}. For the lower bound, with $\gamma = \varepsilon/(2\log(1/\varepsilon))$, a simple calculation using the first bound in (23) (*cf.* (118); the weaker second bound will not do here) gives (say)

$$\mathbb{P}(d_t(v) < \gamma D_t) < n^{-(1+\varepsilon/3)}, \tag{132}$$

implying (130) (and, as discussed above, the lower bound in (55)).

Finally, for (56): Each codegree $d_t(v,w)$ is hypergeometric with mean $(r-1)D_t/(n-1)$; so for ς with (say) $1 \gg \varsigma \gg \max\{D_t^{-1}, n^{-1/2}\}$, Theorem 3.2 gives

$$\mathbb{P}(d_t(v,w) > \varsigma D_t) < \exp[-\varsigma D_t \log(e\varsigma n/r)] = n^{-\omega(1)}.$$

Email address: jkahn@math.rutgers.edu

DEPARTMENT OF MATHEMATICS, RUTGERS UNIVERSITY, HILL CENTER FOR THE MATHEMATICAL SCIENCES, 110 FRELINGHUYSEN RD., PISCATAWAY, NJ 08854-8019, USA