The Johansson-Kahn-Vu solution of the Shamir problem

Deepak Bal and Alan Frieze

February 5, 2013

1 Introduction

A k-uniform hypergraph $H = (V, E)$ on vertex set $V = [n]$ is a collection $E = \{e_1, e_2, \ldots, e_m\}$ where $e_i \in \binom{V}{k}, 1 \leq i \leq m$ i.e. each e_i is a k-element subset of V. A factor or perfect matching of H is a set of disjoint edges $e_i, i \in I$ that partition V. The existence of a perfect matching or 1-factor requires that n is a multiple of k. When $k = 2$ this reduces to the ordinary notion of a perfect matching in a graph.

Next let $H_{n,m;k}$ denote the uniform random hypergraph where E is a random m-subset of $\binom{V}{k}$ and let $H_{n,p;k}$ denote the random hypergraph where each element of $\binom{V}{k}$ is included independently with probability p.

Theorem 1.1. *[Johannson, Kahn, Vu]*
Fix $k \geq 2$. Then there exists a constant $K > 0$ such that if $m \geq Kn \log n$ then

$$\lim_{n \to \infty} \mathbb{P}(H_{kn,m;k} \text{ has a factor}) = 1.$$

In the following, K will be taken to be sufficiently large that all inequalities involving it are valid.

2 Proof of Theorem 1.1

Assume from now on that k divides n and let $e_1, e_2, \ldots, e_N, N = \binom{n}{k}$ be a random ordering of the edges of $H_{k,n}$, the complete k-uniform hypergraph on vertex set $V = [n]$. Let $H_i = H_{k,n} - \{e_1, \ldots, e_i\}$ and $E_i = E(H_i)$ and $m_i = N - i = |E_i|$.

H_i is distributed as $H_{n,m_i;k}$ and the idea is to show that w.h.p. H_i has many factors as long as $m_i \geq Kn \log n$.

For a k-uniform hypergraph $H = (V, E)$, where $k \mid |V|$ we let $\mathcal{F}(H)$ denote the set of factors of H and

$$\Phi(H) = |\mathcal{F}(H)|.$$

Let $\mathcal{F}_t = \mathcal{F}(H_t)$. Then

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

or

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

where

$$\log |\mathcal{F}_0| = \log \frac{n!}{(n/k)!(k!)^{n/k}} = \frac{k-1}{k} n \log n - O(n). \tag{2.1}$$

We also have

$$\mathbb{E}\xi_i = \gamma_i = \frac{n/k}{\binom{n}{k} - i + 1} \leq \frac{1}{kK \log n}. \tag{2.2}$$

for $i \leq T = N - Kn \log n$.

Equation (2.2) becomes, with

$$p_t = \frac{\binom{n}{k} - t}{\binom{n}{k}},$$

$$\sum_{i=1}^{t} \mathbb{E}\xi_i = \sum_{i=1}^{t} \gamma_i = \frac{n}{k}\left(\log \frac{\binom{n}{k}}{\binom{n}{k} - t} + O\left(\frac{1}{\binom{n}{k} - t}\right)\right) = \frac{n}{k}\left(\log \frac{1}{p_t} + O\left(\frac{1}{\binom{n}{k} - t}\right)\right) \tag{2.3}$$

using the fact that $\sum_{i=1}^{N} \frac{1}{i} = \log N + (euler's\ constant) + O(1/N)$.

For $t = T$ this will give

$$p_T = \frac{Kn \log n}{N}$$

and so

$$\sum_{i=1}^{T} \gamma_i = \frac{k-1}{k} n \log n - \frac{n}{k} \log \log n + O(n).$$

Our basic goal is then to prove that if

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

then

$$\mathbb{P}(\bar{\mathcal{A}}_t) \leq n^{-K/10} \text{ for } t \leq T. \tag{2.4}$$

We need the following notation: Suppose $\mathbf{w} : A \to [0, \infty)$ where A is a finite set. Then

$$\bar{\mathbf{w}}(A) = |A|^{-1} \sum_{a \in A} \mathbf{w}(a), \quad \max \mathbf{w}(A) = \max_{a \in A} \mathbf{w}(a), \quad \text{maxr } \mathbf{w}(A) = \frac{\max \mathbf{w}(A)}{\bar{\mathbf{w}}(A)}$$

and

$$\text{med}\mathbf{w}(A) \text{ is the median value of } \mathbf{w}(a), a \in A.$$

We let $V_r = \binom{V}{r}$. For $Z \in V_k$ we let $\mathbf{w}_i(Z) = |\Phi(H_i - Z)|$. Now define property

$$\mathcal{B}_i = \text{maxr } \mathbf{w}_i(E_i) \leq L = K^{1/2}.$$

We also define

$$\mathcal{R}_i = \text{For each } x \in V, \left| D(x, H_i) - \binom{n-1}{k-1} p_i \right| \leq \frac{1}{K^{1/2}} \binom{n-1}{k-1} p_i$$

where

$$D(x,H_i) = |\{e \in E_i : x \in e\}| \text{ is the number of edges of } H_i \text{ that contain } x.$$

We consider the first time $t \le T$, if any, where $\mathcal{A}_t$ fails. Then,

$$\bar{\mathcal{A}}_t \cap \bigcap_{i<t} \mathcal{A}_i \subseteq \left[\bigcup_{i<t} \bar{\mathcal{R}}_i\right] \cup \left[\bigcup_{i<t} \mathcal{A}_i \mathcal{R}_i \bar{\mathcal{B}}_i\right] \cup \left[\bar{\mathcal{A}}_t \cap \bigcap_{i<t} (\mathcal{B}_i \mathcal{R}_i)\right]$$

We can therefore write

$$\mathbb{P}\left(\bar{\mathcal{A}}_t \cap \bigcap_{i<t} \mathcal{A}_i\right) < \sum_{i<t} \mathbb{P}(\bar{\mathcal{R}}_i) + \sum_{i<t} \mathbb{P}(\mathcal{A}_i \mathcal{R}_i \bar{\mathcal{B}}_i) + \mathbb{P}\left(\bar{\mathcal{A}}_t \cap \bigcap_{i<t} (\mathcal{B}_i \mathcal{R}_i)\right). \tag{2.5}$$

The hypergraph H_i is distributed as $H(n, m_i; k)$, the random k-uniform hypergraph on vertex set $[n]$ with $m_i = N - i$ edges. It is easier to work with $H(n, p_i; k)$ where each possible edge occurs independently with probability p_i. Now $\mathbb{P}(H(n, p_i; k)$ has exactly m_i edges is $\Omega(m_i^{-1/2})$ and so we can use $H(n, p_i; k)$ as our model if we multiply the probability of unlikely events by $O(m_i^{1/2})$ – $\mathbb{P}(A \mid B) \le \mathbb{P}(A)/\mathbb{P}(B)$. It then follows that the Chernoff bounds imply that

$$\mathbb{P}(\exists i \le T : \neg \mathcal{R}_i) = O(n^{-K/3}). \tag{2.6}$$

This deals with the first sum in (2.5).

We show next that

$$\mathcal{B}_{i-1} \Rightarrow \xi_i \le \frac{1}{K^{1/2} \log n}.$$

This enables us to use a standard concentration argument to show that $\mathcal{A}_T$ holds.

We first compute

$$\begin{aligned} \mathbf{w}_{i-1}(E_{i-1}) &= \sum_{e \in E_{i-1}} \sum_{F \in \mathcal{F}_{i-1}} 1_{e \in F} \\ &= \sum_{F \in \mathcal{F}_{i-1}} \frac{n}{k}. \end{aligned}$$

Hence, for any $e \in E_{i-1}$,

$$\begin{aligned} \Phi(H_{i-1}) &= \frac{k}{n} \mathbf{w}_{i-1}(E_{i-1}) \\ &\ge \frac{k}{Ln} |E_{i-1}| \max \mathbf{w}_{i-1}(E_{i-1}) \\ &\ge \frac{kN}{Ln} p_{i-1} \mathbf{w}_{i-1}(e). \end{aligned}$$

Hence,

$$\xi_i \le \max_{e \in E_{i-1}} \frac{\mathbf{w}_{i-1}(e)}{\Phi(H_{i-1})} \le \frac{Ln}{kNp_{i-1}} \le \frac{L}{kK \log n} \le \frac{1}{K^{1/2} \log n} \tag{2.7}$$

Now define

$$Z_i = \begin{cases} \xi_i - \gamma_i & \mathcal{B}_j, \mathcal{R}_j \text{ hold for } j < i \\ 0 & \textit{otherwise} \end{cases}$$

and

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

We show momentarily that

$$\mathbb{P}(X_t \geq n) \leq e^{-\Omega(n)}. \tag{2.8}$$

So if we do have $\mathcal{B}_i, \mathcal{R}_i$ for $i < t \leq T$ (so that $X_t = \sum_{i=1}^{t}(\xi_i - \gamma_i)$) and $X_t \leq n$ then

$$\sum_{i=1}^{t} \xi_i < \sum_{i=1}^{t} \gamma_i + n \leq \frac{k-1}{k} n \log n$$

and hence

$$\sum_{i=1}^{t} \xi_i^2 \leq \frac{1}{K^{1/2} \log n} \sum_{i=1}^{t} \xi_i = O(n).$$

So,

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

This deals with the third term in (2.5). (If $\bigcap_{i<t}(\mathcal{B}_t \mathcal{R}_t)$ holds then $\mathcal{A}_t$ holds with sufficient probability).

Let us now verify (2.8). Note that $|Z_i| \leq \frac{1}{K^{1/2} \log n}$ and that for any $h > 0$

$$\mathbb{P}(X_t \geq n) = \mathbb{P}(e^{h(Z_1+\cdots+Z_t)} \geq e^{hn}) \leq E(e^{h(Z_1+\cdots+Z_t)}) e^{-hn} \tag{2.9}$$

Since $Z_i = \xi_i - \gamma_i$ (or zero) and $\mathbb{E}(\xi_i \mid e_1, \ldots, e_{i-1}) = \gamma_i$ and $0 \leq \xi_i \leq \varepsilon = \log^{-1} n$ we have, with h a sufficiently small positive constant,

$$\mathbb{E}(e^{hZ_i} \mid e_1, \ldots, e_{i-1}) \leq e^{-h\gamma_i} \left(1 - \frac{\gamma_i}{\varepsilon} + \frac{\gamma_i}{\varepsilon} e^{h\varepsilon}\right) \leq e^{h^2 \varepsilon \gamma_i}.$$

So,

$$\mathbb{E}(e^{h(Z_1+\cdots+Z_t)}) \leq e^{h^2 \varepsilon \sum_{i=1}^{t} \gamma_i}$$

and going back to (2.9) we get

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

Now $\sum_{i=1}^{t} \gamma_i = O(n \log n)$ and so putting h equal to a small enough positive constant makes the RHS of the above less than $e^{-hn/2}$ and (2.8) follows.

It only remains to deal with the second term in (2.5) and show

$$\mathbb{P}(\mathcal{A}_i \mathcal{R}_i \bar{\mathcal{B}}_i) < n^{-K/4} \tag{2.10}$$

For $|Y| \leq k$ we let

$$V_{k,Y} = \{Z \in V_k : Z \supset Y\}$$

and

$$\mathcal{C}_i = \left\{ \max \mathbf{w}_i(V_{k,Y}) \leq \max \left\{ n^{-(k+1)} \Phi(H_i), 2\mathrm{med}\mathbf{w}_i(V_{k,Y}) \right\} \text{ for all } Y \in V_{k-1} \right\}$$

This event "replaces" the average of $\mathbf{w}_i$ by the median of $\mathbf{w}_i$. A subtle, but vital idea.

We will prove

$$\mathbb{P}(\mathcal{R}_i\mathcal{C}_i\bar{\mathcal{B}}_i) < \varepsilon = n^{-\delta_1 K} \text{ where } \delta_1 = \frac{K\delta}{20k!2^k}. \tag{2.11}$$

$$\mathbb{P}(\mathcal{A}_i\mathcal{R}_i\bar{\mathcal{C}}_i) < n^{-K/3} \tag{2.12}$$

And then use

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

Proof of (2.11)

We make the following assumption: $\mathbb{P}(\mathcal{R}_i\mathcal{C}_i) \geq \varepsilon$, for if not, (2.11) will be trivially satisfied.

Suppose that $|V| = n$ and $\mathbf{w} : V_k \to \Re^+$. For $X \subseteq V$ with $|X| \leq k$ we let $\psi(X) = \max \mathbf{w}(V_{k,X})$.

The following lemma is proved in the appendix:

Lemma 2.1. *Suppose that for each $Y \in V_{k-1}$ and $\psi(Y) \geq B$ we have*

$$\left|\left\{Z \in V_{k,Y} : \mathbf{w}(Z) \geq \frac{1}{2}\psi(Y)\right\}\right| \geq \frac{n-k}{2}.$$

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

$$\left|\left\{Z \in V_{k,X} : \mathbf{w}(Z) \geq \frac{1}{2^j}\psi(X)\right\}\right| \geq \left(\frac{n-k}{2}\right)^j \frac{1}{(j-1)!}. \tag{2.13}$$

Applying the lemma with $B = (2n)^{-(k-1)}\Phi(H_i)$ we see that if $\mathcal{C}_i$ holds then $\psi(Y) > B$ implies that $2\mathrm{med}\mathbf{w}_i(V_{k,Y}) \geq \max \mathbf{w}_i(V_{k,Y})$ and so

$$\left|\left\{Z \in V_{k,Y} : \mathbf{w}_i(Z) \geq \frac{1}{2}\psi(Y)\right\}\right| \geq \frac{n-k}{2}.$$

Putting $j = k$ so that $X = \emptyset$ and $\psi(\emptyset) = \max \mathbf{w}_i(V_k)$,

$$\left|\left\{K \in V_k = V_{k,\emptyset} : \mathbf{w}_i(K) \geq \frac{\max \mathbf{w}_i(V_k)}{2^k}\right\}\right| \geq \delta \frac{n^k}{(k-1)!} \tag{2.14}$$

where $\delta = 1/2^k$.

Let

$$E_i^* = \{e \in E_i : \mathbf{w}_i(e) \geq \delta \max \mathbf{w}_i(E_i)/2\}.$$

We show that (2.14) implies

$$\mathbb{P}\left(|E_i^*| \leq \frac{\delta^2}{2(k!)^2}n^k p_i \middle| \mathcal{R}_i\mathcal{C}_i\right) \leq n^{-\delta_1 K}. \tag{2.15}$$

Now (2.14) implies that there are $\frac{\delta}{k!}n$ vertices x_1 such that for $\frac{\delta}{k!}n^{k-1}$ choices for $x_2, \ldots, x_k$ we have

$$\mathbf{w}_i(x_1, \ldots, x_k) > \delta \max \mathbf{w}_i(E_i). \tag{2.16}$$

Fix such an x_1 and use $H(n,p_i;k)$, but condition on $\mathcal{R}_i\mathcal{C}_i$ holding. For any $x_2,\ldots,x_k$ write $e = \{x_1,\ldots,x_k\}$. Then write, for some enumeration of the edges of $E_i \setminus \{e\}$,

$$\max \mathbf{w}_i(E_i \setminus \{e\}) = \max\{\mathbf{w}_i(e_2),\ldots,\mathbf{w}_i(e_m)\} = \max\left\{\mathbf{w}_i'(e_2)+\mathbf{w}_i''(e_2),\ldots,\mathbf{w}_i'(e_m)+\mathbf{w}_i''(e_m)\right\}$$

where $\mathbf{w}_i'(e_i) \le \mathbf{w}_i(e)$ counts factors that include e_i and e and $\mathbf{w}_i''(e_i)$ counts factors that include e_i but do not include e. (This statement does not assume that $e \in E_i$). Now without conditioning on the occurence of $\mathcal{R}_i\mathcal{C}_i$ we have that $\{e \in E_i\}$ is independent of $\max \mathbf{w}_i''(E_i \setminus \{e\})$ and so, using (2.16) and $\mathbf{w}_i \ge \mathbf{w}_i''$ and the simple inequality $\mathbb{P}(A \mid B) \le \mathbb{P}(A)/\mathbb{P}(B)$ we see that with probability at least

$$1-\varepsilon^{-1}\mathbb{P}(Bin(\delta n^{k-1}/k!,p_T) \le \delta n/2k!) \ge 1-\varepsilon^{-1}n^{-1-2\delta_1 K} = 1-n^{-1-\delta_1 K}$$

there are $\frac{\delta}{2k!}n^{k-1}p_i$ sequences $x_2,\ldots,x_k$ such that e is an edge and $\mathbf{w}_i(e) \ge \delta \max \mathbf{w}_i''(E_i \setminus \{e\})$. Let $x_2,\ldots,x_k$ be such a choice. Now $\max \mathbf{w}_i(E_i) \le \mathbf{w}_i(e) + \max \mathbf{w}_i''(E_i \setminus \{e\})$ and so $\mathbf{w}_i(e) \ge \delta \max \mathbf{w}_i(E_i)/2$ for such $x_2,\ldots,x_k$. ($\mathbf{w}_i(e)/2 \ge \delta \max \mathbf{w}_i''(E_i \setminus \{e\})/2$ and $\mathbf{w}_i(e)/2 > \delta \mathbf{w}_i(e)/2$).

There are at most n choices for x_1 and so with probability $1-n^{-\delta_1 K}$ we have that for each choice of x_1 there are $\frac{\delta}{2k!}n^{k-1}p_i$ choices for $x_2,\ldots,x_k$ such that $\{x_1,\ldots,x_k\}$ is an edge and $\mathbf{w}_i(x_1,\ldots,x_k) > \delta \max \mathbf{w}_i(E_i)/2$. This verifies (2.15) and we have

$$\frac{\sum_{e\in E_i}\mathbf{w}_i(e)}{\max \mathbf{w}_i(E_i)} \ge \frac{\sum_{e\in E_i^*}\mathbf{w}_i(e)}{\max \mathbf{w}_i(E_i)} \ge \frac{\delta|E_i^*|}{2} \ge \frac{\delta^3}{4(k!)^2}n^k p_i \ge \frac{\delta^3}{10k!}|E_i|$$

which implies property $\mathcal{B}_i$ if $K^{1/2} \ge 10k!\delta^{-3}$.

Proof of (2.12)

We need the following two lemmas that are proved in the appendix:
Given $y \in V$ we let $X(y,H)$ denote the edge e containing y in a uniformly random factor of H. We let

$$h(y,H) = -\sum_{e \ni y}\mathbb{P}(X(y,H)=e)\ \log \mathbb{P}(X(y,H)=e)$$

denote the entropy of $X(y,H)$.

Lemma 2.2.

$$\log \Phi(H) \le \frac{1}{k}\sum_{y\in V} h(y,H).$$

For the next lemma let S be a finite set and $\mathbf{w}: S \to \Re^+$ and let X be the random variable with

$$\mathbb{P}(X=x) = \frac{\mathbf{w}(x)}{\mathbf{w}(S)}.$$

Let $h(X) = -\sum_{x\in S}\mathbb{P}(x)\log \mathbb{P}(x)$ be the entropy of X.

Lemma 2.3. *If $h(X) \ge \log|S| - O(1)$ then there exist $a,b \in range(\mathbf{w})$ with $a \le b \le O(a)$ such that for $J = \mathbf{w}^{-1}[a,b]$ we have*

$$|J| = \Omega(|S|) \text{ and } \mathbf{w}(J) > .7\mathbf{w}(S).$$

Assume that we have $\mathcal{A}_i$ and $\mathcal{R}_i$ and that $\mathcal{C}_i$ fails at Y. Let $e = Y \cup \{x\} \in V_k$ satisfy $\mathbf{w}_i(e) = \max \mathbf{w}_i(V_{k,Y})$. Note that

$$\mathbf{w}_i(e) > n^{-(k+1)}\Phi(H_i) = e^{\Omega(n \log\log n)}. \tag{2.17}$$

Choose $y \in V \setminus Y$ with $\mathbf{w}_i(Y \cup \{y\}) \leq \text{med}\mathbf{w}_i(V_{k,Y})$ and with $h(y, H_i - e)$ maximum subject to this restriction and set $f = Y \cup \{y\}$. Note that $y \neq x$ by its definition.

We argue that $h(y, H_i - e) \geq \log D(y, H_i - e)$ and apply Lemma 2.3 to get J, a, b. Then we take $W = V \setminus (Y + \{x, y\})$ and for $Z \in \binom{W}{k-1}$ we let $f(Z) = \Phi(H_i - (Y + Z + \{x, y\})$. Then we let $W_0 = \{Z \in W : f(Z) \in [a, b]\}$. Then we consider

$$\alpha = \sum_{Z \in W_0} f(Z) 1_{Z+y \in E} \text{ and } \beta = \sum_{Z \in W_0} f(Z) 1_{Z+x \in E}.$$

Because $Z, Z' \in W_0$ implies that $f(Z)/f(Z') = O(1)$ we see that w.h.p. $\alpha \sim \beta$. But the definitions imply that

$$\alpha > .7\mathbf{w}_i(e) > .5\mathbf{w}_i(e) \geq \mathbf{w}_i(f) \geq \beta.$$

Continuing the more detailed argument, we have

$$\mathbf{w}_i(e) > 2\text{med}\mathbf{w}_i(V_{k,Y}) \geq 2\mathbf{w}_i(f).$$

Since we have $\mathcal{A}_i$, we have

$$\log |\Phi(H_i)| > \log |\Phi(H_0)| - \sum_{t=1}^{i} \gamma_t - O(n) = \frac{k-1}{k} n \log n + \frac{n}{k} \log p_i - O(n).$$

This and (2.17) implies that

$$\log \Phi(H_i - e) = \log \mathbf{w}_i(e) \geq \frac{k-1}{k} n \log n + \frac{n}{k} \log p_i - O(n) \tag{2.18}$$

But Lemma 2.2 implies that

$$\log \Phi(H_i - e) \leq \frac{1}{k} \sum_{z \in V \setminus e} h(z, H_i - e)$$

and by our choice of y we have $h(z, H_i - e) \leq h(y, H_i - e)$ for at least half the z's in $V \setminus e$ and that for all $z \in V \setminus e$ we have

$$h(z, H_i - e) \leq \log D(z, H_i - e) \leq \log \left((1 + o(1)) \binom{n}{k-1} p_i \right).$$

This implies that

$$\log \Phi(H_i - e) \leq \frac{n}{2k} \left(\log \left((1 + o(1)) \binom{n}{k-1} p_i \right) + h(y, H_i - e) \right). \tag{2.19}$$

Combining (2.18) and (2.19) we get

$$h(y, H_i - e) > (k-1) \log n + \log p_i - O(1) = \log D(y, H_i - e) - O(1). \tag{2.20}$$

Let $W = V \setminus (Y + \{x, y\})$ and for $Z \in W_{k-1}$ let

$$\mathbf{w}'_i(Z) = \Phi(H_i - (Y \cup Z \cup \{x, y\}).$$

Then define

$$\mathbf{w}_y \text{ on } W_y = \{K \subseteq V \setminus e : \ |K| = k, y \in K\}$$

and

$$\mathbf{w}_x \text{ on } W_x = \{K \subseteq V \setminus f : \ |K| = k, x \in K\}$$

by

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

Then $X(y, H_i - e)$ is chosen according to the weights $\mathbf{w}_y$ and $X(x, H_i - f)$ is chosen according to the weights $\mathbf{w}_x$. Note also that $\mathbf{w}_y(W_y) = \mathbf{w}_i(e)$ and $\mathbf{w}_x(W_x) = \mathbf{w}_i(f)$.

Let $a, b \in range(\mathbf{w}_y) \subseteq range(\mathbf{w}'_i)$ be as defined in Lemma 2.3, using (2.20). Then with $M = |W_{k-1}| = \Omega(n^{k-1})$,

$$\mathbf{w}_y(J = \mathbf{w}_y^{-1}[a, b]) =$$

$$\sum_{\substack{Z \subseteq W \\ |Z|=k-1 \\ \Phi(H_i-(Y+Z+x+y)) \in [a,b]}} \Phi(H_i - (Y + Z + x + y))1_{Z+y \in E} = \sum_{j=1}^{M} a_j \zeta_1^{(A)} > .7\mathbf{w}_y(W_y) = .7\mathbf{w}_i(e)$$

$$\sum_{\substack{Z \subseteq W \\ |Z|=k-1 \\ \Phi(H_i-(Y+Z+x+y)) \in [a,b]}} \Phi(H_i - (Y + Z + x + y))1_{Z+x \in E} = \sum_{j=1}^{M} a_j \zeta_1^{(B)} \leq \mathbf{w}_x(W_x) = \mathbf{w}_i(f) \leq .5\mathbf{w}_i(e)$$

The probability of this is $n^{-\omega(1)}$. This is because $M = \Omega(n^{k-1})$ and the a_j's are all within $O(1)$ of each other and using $H(n, p_i; k)$, the $\zeta_i^{(A)}, \zeta_i^{(B)}$ are both collections of independent 0-1 variables with the same mean p_i. This completes the proof of (2.12).

References

[1] A. Johansson, J. Kahn and V. Vu, Factors in Random Graphs, *Random Structures and Algorithms* 33 (2008) 1-28.

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

Appendix

A Proof of Lemma 2.1

Write N_i for the r.h.s. of (2.13). 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 $\mathbf{w}(Z)$ maximum (i.e. $\mathbf{w}(Z) = \psi(X)$). Let

$y \in Z \ X$ and $Y = X \cup \{y\}$. Then $|Y| = v(i1)$ and $\psi(Y) = \psi(X) \geq 2^{i1}B(\geq 2^{i2}B)$; so by our induction hypothesis there are at least N_{i1} sets $Z' \in \mathcal{H}_0(Y)$ with $\mathbf{w}(Z') \geq 2^{(i1)}\psi(Y)(= 2^{(i1)}\psi(X))$. For each such $Z', Z' \setminus \{y\}$ is a $(v1)$-subset of V with $\psi(Z' \setminus \{y\}) \geq \mathbf{w}(Z') \geq B$. So (again, for each such Z') there are at least $(nv)/2$ sets $Z'' \in \mathcal{H}_0(Z' \setminus \{y\})$ with

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

The number of these pairs (Z', Z'') is thus at least $N_{i1}(nv)/2$. 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.

B Proof of Lemma 2.2

This follows from what is referred to as “Shearer's Lemma,“ or more precisely what Shearer's proof actually gives. The lemma may be stated as follows. Suppose that $Y = (Y_i : i \in I)$ 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 S. Then the entropy $h(Y) \leq t^{-1}\sum_{S\in\mathcal{S}} h(Y_S)$, where Y_S is the random vector $(Y_i : i \in S)$. To get Lemma 2.2 from this, let Y be the indicator of the random factor (so that I is the set of edges of the complete k-uniform hypergraph $H_{k,n}$) and $\mathcal{S} = (S_v : v \in V)$, where S_v is the set of edges of $H_{k,n}$ containing v.

C Proof of Lemma 2.3

Let

$$H(X) = \log|S| - K \tag{C.1}$$

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

$$H(X) \leq \log 3 + \frac{\mathbf{w}(L)}{\mathbf{w}(S)}\log|L| + \frac{\mathbf{w}(J)}{\mathbf{w}(S)}\log|J| + \frac{\mathbf{w}(U)}{\mathbf{w}(S)}\log|U|. \tag{C.2}$$

Then we have a few observations. First, $|U| < |S|/C$ implies that the r.h.s. of (C.2) is less than

$$\log 3 + \log|S|\frac{\mathbf{w}(U)}{\mathbf{w}(S)}\log C$$

which with (C.1) implies

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

Of course this also implies $|U| < |S|/4$. Second, combining (C.3) with the trivial $\mathbf{w}(L) < w(S)/C$, we have (say) $\mathbf{w}(J) > .7\mathbf{w}(S)$. But then (third) since the r.h.s. of (C.2) is at most

$$\log 3 + \log|S| + \frac{\mathbf{w}(J)}{\mathbf{w}(S)}\log\frac{|J|}{|S|} < \log 3 + \log|S| + .7\log\frac{|J|}{|S|},$$

we have

$$|J| \geq \exp\left\{(.7)^1(K + \log 3)\right\}|S|(= \Omega(|S|)).$$