Perfect Matchings in Random s-Uniform Hypergraphs

Alan Frieze* Svante Janson†

November 6, 2001

1 Introduction

Let $E = \{X_1, X_2, \ldots, X_m\}$ where the $X_i \subseteq V$ for $1 \leq i \leq m$ are distinct. The hypergraph $G = (V, E)$ is said to be *s-uniform* if $|X_i| = s$ for $1 \leq i \leq m$. Thus, for example, a $2-$uniform hypergraph is a graph. A set of edges $M = \{X_i : i \in I\}$ is a *perfect matching* if

(i) $i \neq j \in I$ implies $X_i \cap X_j = \emptyset$, and

(ii) $\bigcup_{i \in I} X_i = V$.

In this paper we consider the question of whether a random s-uniform hypergraph contains a perfect matching. More precisely, where $[k] = \{1, 2, \ldots, k\}$, we consider the random hypergraph $G = G(n, m, s) = (V = [n], E = \{X_1, X_2, \ldots, X_m\})$ where the X_i are chosen randomly (without replacement)

*Department of Mathematics, Carnegie Mellon University, Pittsburgh PA15213, U.S.A., Supported by NSF grant CCR-9225008

†Department of Mathematics, Uppsala University, PO Box 480, S-751 06 Uppsala, Sweden, Supported in part by the Göran Gustafsson Foundation for Research in Natural Sciences and Medicine.

Much of this work was done at the IMA Workshop on Emerging Applications of Probability at The University of Minnesota

from the set $\binom{V}{s}$ of all s-subsets of V. Equivalently we choose a hypergraph uniformly at random from the set $\mathcal{G}(n,m,s)$ of all $\binom{\binom{n}{s}}{m}$ s-uniform hypergraphs with vertex set $[n]$ and m distinct edges.

To avoid trivialities we assume from now on that $n \to \infty$ in steps of size s so that $s \mid n$ always.

One of the most interesting and difficult problems in Probabilistic Combinatorics is to determine the threshold growth for $m = m(n)$ so that when s is fixed

$$\lim_{n\to\infty} \mathbf{Pr}(G(n,m,s) \text{ has a perfect matching}) = 1. \tag{1}$$

The case $s = 2$ was solved completely by Erdős and Rényi [6]. This is the question of when does the random graph $G_{n,m}$ have a perfect matching **whp**.[1] The first serious work on this problem for $s \geq 3$ was done by Schmidt and Shamir [10]. Using the second moment method directly, they show that if $m/n^{3/2} \to \infty$ then $G(n,m,s)$ has a perfect matching **whp**. Since then, determining the threshold for (1) has become known as Shamir's problem which seems to ignore Schmidt's contribution.

Cooper, Frieze, Molloy and Reed [5] considered the problem of perfect matchings in random r-regular, s-uniform hypergraphs. Surprisingly, this seems to be an easier problem and is solvable by the Analysis of Variance method of Robinson and Wormald [8, 9]. In that paper they make the following conjecture which we repeat here:

CONJECTURE. Assume s is a positive integer constant and let

$$m = n(\log n + c_n)/s.$$

Then

$$\lim_{n\to\infty} \mathbf{Pr}(G \text{ has a perfect matching }) = \begin{cases} 0 & c_n \to -\infty, \\ e^{-e^{-c}} & c_n \to c, \\ 1 & c_n \to \infty. \end{cases}$$

The right-hand side of the above expression is simply the limiting probability that $\bigcup_{i=1}^m X_i = V$.

[1]An event $\mathcal{E}_n$ is said to occur **whp** (with high probability) if $\mathbf{Pr}(\mathcal{E}_n) = 1 - o(1)$ as $n \longrightarrow \infty$.

There has been no other progress on this problem as far as we can tell. Our main result is an improvement of the upper bound of Schmidt and Shamir to

Theorem 1 *Let $s \geq 3$ be fixed. If $m/n^{4/3} \to \infty$, then the random hypergraph $G(n, m, s)$ contains a perfect matching* **whp**.

(The result and the proof below hold also for $s = 2$, but then the result by Erdős and Rényi [6] is much stronger.) Our proof is also by the second moment method, but we obtain some reduction in variance by sampling randomly from hypergraphs with a fixed degree sequence.

Schmidt and Shamir also considered the s-partite version of the problem. Here $V = [n]$ is partitioned into s equally large sets $V_1, V_2, \ldots, V_s$ and the edges X_i, $i = 1, 2, \ldots m$ are chosen randomly from $\{X : |X| = s \text{ and } |X \cap V_t| = 1, t = 1, 2, \ldots, s\}$ i.e. each edge contains a single member of each V_i. Let $G^*(n, m, s)$ denote the random hypergraph produced. We will sketch a proof of

Theorem 2 *Let $s \geq 3$ be fixed. If $m/n^{4/3} \to \infty$, then the random hypergraph $G^*(n, m, s)$ contains a perfect matching* **whp**.

2 Proof of Theorem 1

We will consider (multi)hypergraphs with a fixed degree sequence. As an interim model we let $\tilde{G}(n, m, s)$ be defined as follows: let $\mathbf{x} = x_1, x_2, \ldots, x_{sm}$ be chosen randomly from $[n]^{sm}$ i.e. $\mathbf{x}$ is a random sequence of integers from $[n]$ (sampled with replacement). Let $G(\mathbf{x})$ be the (multi)hypergraph with vertex set $[n]$ and edges $\tilde{E}_i = \{x_{(i-1)s+1}, \ldots, x_{is}\}$ for $1 \leq i \leq m$. Note that $G(\mathbf{x})$ may contain repeated edges and *deficient* edges with fewer than s vertices. Let $\hat{G} = \hat{G}(\mathbf{x})$ be obtained from $G(\mathbf{x})$ by deleting edge repetitions and deficient edges.

Observe that if $\hat{G}$ has $\hat{m} \leq m$ edges then $\hat{G}$ is distributed as $G(n, \hat{m}, s)$, since each member of $\mathcal{G}(n, \hat{m}, s)$ arises from the same number of $\mathbf{x} \in [n]^{sm}$ in this way. We define a perfect matching of $G(\mathbf{x})$ to consist of n disjoint edges of size s and so we see that any perfect matching of $G(\mathbf{x})$ is also a perfect matching of $\hat{G}$. Since $\hat{m} \leq m$ and having a perfect matching is a monotone property we have

$$\mathbf{Pr}(G(n,m,s) \text{ has a perfect matching}) \geq \mathbf{Pr}(G(\mathbf{x}) \text{ has a perfect matching})$$

and we can concentrate on showing that

$$G(\mathbf{x}) \text{ has a perfect matching } \mathbf{whp} \text{ if } m/n^{4/3} \to \infty. \tag{2}$$

The degree $deg_{\mathbf{x}}(i)$ of $i \in [n]$ in the hypergraph $G(\mathbf{x})$ is simply $|\{j : x_j = i\}|$. For a fixed degree sequence $\mathbf{d} = d_1, d_2, \ldots, d_n$ we let

$$X(\mathbf{d}) = \{\mathbf{x} \in [n]^{sm} : deg_{\mathbf{x}} = \mathbf{d}\},$$

where $m = s^{-1} \sum_i d_i$ (we assume that this is an integer).

Schmidt and Shamir worked directly with $\mathcal{G}(n,m,s)$. We will work with $G(\mathbf{x})$ where $\mathbf{x}$ is chosen randomly from $X(\mathbf{d})$. This will eliminate some of the variability in the number of perfect matchings and enable us to improve their result.

A degree sequence is said to be ϵ-*smooth* if it satisfies properties **P1–P4** below. The notation used is

$$\begin{aligned}
d_{\min} &= \min\{d_i : 1 \leq i \leq n\}, \\
d &= \frac{1}{n}\sum_{i=1}^{n} d_i, \\
R &= \sum_{i=1}^{n} d_i^{-1}, \\
Q &= \sum_{i=1}^{n} d_i^{-2}.
\end{aligned}$$

The properties are

P1 $\quad d_{\min} > 0.$

P2 $\quad d \leq \epsilon n.$

P3 $\left|R - \frac{n}{d}\right| \leq \epsilon \left(\frac{n}{d}\right)^{1/2}.$

P4 $Q \leq \epsilon R.$

Note that Jensen's inequality (or the arithmetic-harmonic inequality) implies $R \geq n/d$. Moreover, Cauchy-Schwarz' inequality yields $R^2 \leq nQ$, and thus $n/d \leq R \leq nQ/R$ or $d \geq R/Q$. Hence **P4** implies

P5 $d \geq 1/\epsilon.$

Our proof of (2) is divided into two lemmas. Since the probability of $G(n, m, s)$ having a perfect matching increases with m, there is no loss in assuming that $m = o(n)$.

Lemma 1 *If* $\mathbf{x}$ *is chosen uniformly at random from* $[n]^{sm}$ *where* $m = \omega n^{4/3}$, $\omega(n) \to \infty$ *and* $\omega(n) = o(n^{2/3})$, *then* **whp** $deg_{\mathbf{x}}$ *is* $\epsilon(n)$*-smooth for some* $\epsilon(n) \to 0$.

Lemma 2 *Let (for each* n*)* $\mathbf{d}$ *be an* $\epsilon(n)$*-smooth degree sequence where* $\epsilon(n) \to 0$. *Let* $\mathbf{x}$ *be chosen uniformly at random from* $X(\mathbf{d})$. *Then* **whp**

$G(\mathbf{x})$ *has a perfect matching.*

2.1 Configuration Model

To analyse $G(\mathbf{x})$ where $\mathbf{x}$ is chosen uniformly from $X(\mathbf{d})$ we generalise the models of Bender and Canfield [1] and Bollobás [2] for the case of graphs with a fixed degree sequence ($s = 2$).

Let $W_i = \{i\} \times [d_i]$ for $i = 1, 2, \ldots, n$ and $W = \bigcup_{i=1}^n W_i$. A *configuration* is a partition of W into m disjoint subsets of size s. Let $\Omega = \Omega(\mathbf{d})$ be the set of all such configurations, and let F be chosen randomly from Ω.

For $x = (v, i) \in W$ we let $\sigma(x) = v$. If $F \in \Omega$ and $S \in F$ we let $\sigma(S) = \{\sigma(x) : x \in S\}$. We define the multihypergraph $\mu(F) = (V, E_\mu = \{\sigma(S) : S \in F\})$. It is not difficult to see that $\mu(F)$ and $G(\mathbf{x}), \mathbf{x} \in X(\mathbf{d})$ have the same distribution. (We can generate $\mathbf{x} \in X(\mathbf{d})$ by taking d_i "copies" of i for $1 \leq i \leq n$ and then placing these sm copies in random order. The relationship with $\mu(F)$ should be clear.) This idea was used in Bollobás and Frieze [3] and Bollobás, Fenner and Frieze [4] to analyse random graphs with minimum degree bounded from below.

Before continuing with the proofs of Lemmas 1 and 2 we need to define a perfect matching in $F \in \Omega(\mathbf{d})$.

A perfect matching M of F is a set $\{S_i : i \in I\} \subseteq F$ such that

(i) $|\sigma(S_i)| = s$, $i \in I$,

(ii) $i, j \in I, i \neq j$ implies $\sigma(S_i) \cap \sigma(S_j) = \emptyset$, and

(iii) $\bigcup_{i \in I} \sigma(S_i) = V$.

Thus F contains a perfect matching if and only if $\mu(F)$ contains one. If M is a perfect matching of F we let $\sigma(M) = \{\sigma(E_i) : i \in I\}$ be the partition of $[n]$ into s-sets that it induces.

We can now begin the proof proper.

2.2 Proof of Lemma 1

We first note that $d = sm/n = s\omega n^{1/3}$, which shows **P2**.

Next, $d_1, d_2, \ldots d_n$ are random variables, each having the binomial distribution $\mathrm{Bi}(sm, n^{-1}) = \mathrm{Bi}(nd, n^{-1})$. By standard estimates, **whp**

$$d_{\min} \geq s\omega n^{1/3}/2,$$

which will account for **P1** and **P4**.

For **P3** we have already observed that Jensen's inequality implies $R \geq n/d$. For an upper bound, we note that if $d_{\min} \geq d/2$, then

$$\frac{1}{d_i} = \frac{1}{d} + \frac{d - d_i}{dd_i} = \frac{1}{d} + \frac{d - d_i}{d^2} + \frac{(d - d_i)^2}{d^2 d_i} \leq \frac{1}{d} + \frac{d - d_i}{d^2} + 2\frac{(d - d_i)^2}{d^3}$$

and thus

$$R \leq \frac{n}{d} + \frac{2}{d^3}\sum_{i=1}^{n}(d_i - d)^2.$$

Consequently, if $\delta = \omega^{-1}$ then,

$$\begin{aligned}
&\mathbf{Pr}\left(|R - \tfrac{n}{d}| \geq \delta\left(\tfrac{n}{d}\right)^{1/2}\right) \leq \\
&\qquad \leq \mathbf{Pr}\left(\frac{2}{d^3}\sum_{i=1}^{n}(d_i - d)^2 \geq \delta\left(\frac{n}{d}\right)^{1/2}\right) + \mathbf{Pr}\left(d_{\min} < \frac{d}{2}\right) \\
&\qquad \leq \delta^{-1}\left(\frac{n}{d}\right)^{-1/2}\frac{2}{d^3}\mathbf{E}\left(\sum_{i=1}^{n}(d_i - d)^2\right) + o(1) \\
&\qquad = 2\delta^{-1}n^{-1/2}d^{-5/2}nnd\frac{1}{n}\left(1 - \frac{1}{n}\right) + o(1) \\
&\qquad < 2\delta^{-1}n^{1/2}d^{-3/2} + o(1) \\
&\qquad = 2s^{-3/2}\omega^{-1/2} + o(1) \\
&\qquad = o(1).
\end{aligned}$$

This verifies **P3** and the Lemma follows.

2.3 Proof of Lemma 2

We will work with the configuration model and show that if F is chosen randomly from $\Omega(\mathbf{d})$ then **whp**

$$\mu(F) \text{ has a perfect matching.} \tag{3}$$

Let $Y(F)$ denote the number of perfect matchings of F. Since

$$\mathbf{Pr}(Y(F) > 0) \geq \frac{\mathbf{E}(Y)^2}{\mathbf{E}(Y^2)}$$

we can prove (3) by showing that

$$\frac{\mathbf{E}(Y^2)}{\mathbf{E}(Y)^2} \to 1 \qquad \text{as } n \to \infty.$$

For positive integers a, b with $a = sb$ let

$$\psi(a) = \frac{a!}{b!(s!)^b}$$

denote the number of ways of partitioning $[a]$ into b s-subsets.

Then

$$\mathbf{E}(Y) \quad = \quad d_{[n]}\psi(n)\frac{\psi(sm-n)}{\psi(sm)},$$

where

$$d_S \quad = \quad \prod_{i \in S} d_i \qquad\qquad \text{for } S \subseteq [n].$$

Explanation: There are $\psi(n)$ choices for the sets $\sigma(S_i), i \in I$ of the perfect matching. There are then $d_{[n]}$ choices for the sets $S_i, i \in I$ and the remaining factor is the probability that F contains a particular matching.

Furthermore, where $\nu = n/s$,

$$\mathbf{E}(Y^2) = \sum_{k=0}^{\nu} \sum_{|S|=ks} \psi(sk)\psi(n-sk)^2 d_{[n]} \left(\prod_{x \notin S}(d_x - 1)\right) \frac{\psi(sm-2n+sk)}{\psi(sm)}.$$

Explanation: We first remind the reader that $G(\mathbf{x})$ can have multiple edges. We have to sum over the probabilities that two matchings M_1, M_2 exist in F. In the above sum k denotes the size of $M_1 \cap M_2$ and $S = \{x : x \in E \in \sigma(M_1 \cap M_2)\} \subseteq [n]$ is the set of σ images of elements of common edges. Having fixed k, S we determine $\sigma(M_1 \cap M_2), \sigma(M_1 \setminus M_2)$ and $\sigma(M_2 \setminus M_1)$. This gives $\psi(sk)\psi(n-sk)^2$ choices. We then have to actually choose M_1, M_2. If $x \in S$ there are d_x choices of w such that $\sigma(w) = x$ and otherwise we must choose two distinct elements to map onto x which gives us a factor $d_x(d_x - 1)$ for $x \in \bar{S} = [n] \setminus S$. The final factor is the probability that F contains the two matchings.

Hence,

$$\frac{\mathbf{E}(Y^2)}{\mathbf{E}(Y)^2} = \sum_{k=0}^{\nu} \frac{(sk)!}{k!}\frac{(n-sk)!^2}{(\nu-k)!^2}\frac{(s(m-2\nu+k))!}{(m-2\nu+k)!}\frac{(sm)!}{m!}\frac{\nu!^2}{n!^2}\frac{(m-\nu)!^2}{(sm-n)!^2}D_{n-sk} \tag{4}$$

where

$$D_j = d_{[n]}^{-1} \sum_{|A|=j} \prod_{x \in A} (d_x - 1), \qquad 0 \le j \le n.$$

Denote the combination of factorials in (4) by a_k, so that

$$\frac{\mathbf{E}(Y^2)}{\mathbf{E}(Y)^2} = \sum_{k=0}^{\nu} a_k D_{n-sk}. \tag{5}$$

By Stirling's formula,

$$\begin{aligned} a_k &= s^{1/2} k^{k(s-1)} (\nu - k)^{2(s-1)(\nu-k)} (m - 2\nu + k)^{(s-1)(m-2\nu+k)} \\ &\quad m^{(s-1)m} \nu^{-2(s-1)\nu} (m - \nu)^{-2(s-1)(m-\nu)} (1 + \varepsilon_k), \end{aligned} \tag{6}$$

where $\varepsilon_k = O\left(\frac{1}{k+1} + \frac{1}{\nu-k+1}\right)$.

Let $x = k/\nu$, then (6) can be written, using $m/\nu = sm/n = d$,

$$\begin{aligned} a_k &= \sqrt{s}(1 + \varepsilon_k) \left[x^x (1-x)^{2(1-x)} (d-2+x)^{d-2+x} d^d (d-1)^{-2(d-1)} \right]^{(s-1)\nu} \\ &= \sqrt{s}(1 + \varepsilon_k) e^{\nu(s-1)f(x)} \end{aligned} \tag{7}$$

with

$$\begin{aligned} f(x) &= x \log x + 2(1-x) \log(1-x) + (d-2+x) \log(d-2+x) \\ &\quad + d \log d - 2(d-1) \log(d-1). \end{aligned}$$

We have

$$\begin{aligned} f'(x) &= \log x - 2 \log(1-x) + \log(d-2+x), \\ f''(x) &= \frac{1}{x} + \frac{2}{1-x} + \frac{1}{d-2+x} > 0, \end{aligned} \tag{8}$$

and find after simple calculations

$$f(1/d) = f'(1/d) = 0.$$

(This is thus the minimum of f.)

We turn to the term D_{n-sk} in (5).

Lemma 3 *If* **P1–P4** *hold, then*

$$D_{n-j} \le c_j \exp(j - R - j\log(j/R)), \tag{9}$$

with $c_j = c_j(n) = O((j+1)^{-1/2} + R^{-1/2})$ *(uniformly in* j *and* n*) and* $c_j \sim (2\pi n/d)^{-1/2}$ *for any sequence* $j = j(n)$ *with* $j \sim n/d$.

Proof: Define the generating function

$$D(z) = \sum_0^n D_{n-j} z^j = \prod_1^n d_k^{-1} \prod_1^n (z + (d_k - 1)) = \prod_1^n \left(1 + \frac{z-1}{d_k}\right)$$

By Cauchy's formula,

$$D_{n-j} = \frac{1}{2\pi i} \oint_\gamma D(z) z^{-j-1} dz$$

for any simple closed curve γ around the origin. We let γ be the circle $|z| = r$, where r will be chosen later, and obtain by taking absolute values

$$D_{n-j} \le \frac{1}{2\pi} \int_{-\pi}^{\pi} |D(re^{it})| r^{-j} dt. \tag{10}$$

We next observe that if $0 \le r \le 1$ and $0 \le a \le 1$, then

$$\begin{aligned}
|1 + a(re^{it} - 1)|^2 &= (1 - a(1 - r\cos t))^2 + a^2 r^2 \sin^2 t \\
&= (1 - a(1-r) - ar(1 - \cos t))^2 + a^2 r^2 \sin^2 t \\
&= (1 - a(1-r))^2 - 2(1 - a(1-r))ar(1 - \cos t) + \\
&\qquad a^2 r^2 (1 - 2\cos t + \cos^2 t + \sin^2 t) \\
&\le (1 - a(1-r))^2 - 2(1 - a(1-r))ar(1 - \cos t) + \\
&\qquad a^2 r (1 - a(1-r)) 2 (1 - \cos t) \\
&\le (1 - a(1-r) - ar(1 - \cos t) + a^2 r (1 - \cos t))^2
\end{aligned}$$

and thus if $0 \le r \le 1$,

$$|1 + a(re^{it} - 1)| \le 1 - a(1-r) - (a - a^2) r (1 - \cos t). \tag{11}$$

(The right hand side is always ≥ 0, since it is so when $r = 0$ and $r = 1$).

If $r > 1$ we use (11) and the triangle inequality to obtain

$$\begin{aligned}|1 + a(re^{it} - 1)| &\leq |1 + a(e^{it} - 1)| + |ae^{it}(r - 1)| \\ &\leq 1 - (a - a^2)(1 - \cos t) + a(r - 1), \qquad r > 1\end{aligned}$$

Hence we obtain, for every r and $d_j \geq 1$,

$$\begin{aligned}|1 + d_j^{-1}(re^{it} - 1)| &\leq 1 + d_j^{-1}(r - 1) - (d_j^{-1} - d_j^{-2})(1 \wedge r)(1 - \cos t) \\ &\leq \exp\left[d_j^{-1}(r - 1) - (d_j^{-1} - d_j^{-2})(1 \wedge r)(1 - \cos t)\right]\end{aligned}$$

and

$$\begin{aligned}|D(re^{it})| &\leq \exp\left[\sum_j d_j^{-1}(r - 1) - \sum_j (d_j^{-1} - d_j^{-2})(1 \wedge r)(1 - \cos t)\right] \\ &= \exp[R(r - 1) - (R - Q)(1 \wedge r)(1 - \cos t)]. \qquad (12)\end{aligned}$$

We substitute the estimate (12) in (10) and pause for another lemma.

Lemma 4 *Let*

$$\varphi(b) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-b(1-\cos t)} dt, \qquad b \geq 0.$$

Then $\varphi(b) = O(1 \wedge b^{-1/2})$, *and* $\varphi(b) = \frac{1}{\sqrt{2\pi}} b^{-1/2}(1 + o(1))$ *as* $b \to \infty$.

Proof: Clearly $\varphi(b) \leq 1$. On the other hand,

$$\sqrt{b}\varphi(b) = \frac{1}{2\pi} \int_{-\pi\sqrt{b}}^{\pi\sqrt{b}} e^{-b(1-\cos(u/\sqrt{b}))} du \to \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-u^2/2} du = \frac{1}{\sqrt{2\pi}}$$

as $b \to \infty$ by dominated convergence

$$\left(\exp(-b(1 - \cos(u/\sqrt{b}))) \leq \exp\left(-\frac{2}{\pi^2}u^2\right) \text{ when } |u| \leq \pi\sqrt{b}\right).$$

□

We may now complete the proof of Lemma 3. With φ as in Lemma 4, (10) and (12) yield

$$D_{n-j} \le r^{-j}\varphi((R-Q)(1\wedge r))\exp(Rr-R).$$

We choose $r = j/R$ and obtain

$$D_{n-j} \le \varphi\left((1-Q/R)(R\wedge j)\right)\exp\left(j-R-j\log(j/R)\right),$$

which is (9) with $c_j = \varphi\left((1-Q/R)(R\wedge j)\right)$. The estimates for c_j follow from Lemma 4 because by assumptions **P2**–**P4**, $n/d\to\infty$, $R\sim n/d$ and $Q/R\to 0$. □

Continuing with the proof of Lemma 2, we define

$$g(x) = sx - R/\nu - sx\log(nx/R)$$

and obtain by Lemma 3 $D_{n-sk} \le c_{sk}\exp(\nu g(k/\nu))$ and thus by (5) and (7)

$$\frac{\mathbf{E}Y^2}{(\mathbf{E}Y)^2} \le \sum_{k=0}^{\nu}\sqrt{s}(1+\varepsilon_k)c_{sk}\exp\left(\nu(s-1)f(k/\nu)+\nu g(k/\nu)\right).$$

Here we may absorb the factor $1+\varepsilon_k$ into c_{sk} without changing the estimates given in Lemma 3.

We also define

$$h(x) = (s-1)f(x) + g(x)$$

and obtain

$$\frac{\mathbf{E}Y^2}{(\mathbf{E}Y)^2} \le \sum_{k=0}^{\nu}\sqrt{s}c_{sk}\exp(\nu h(k/\nu)). \tag{13}$$

Differentiation yields, cf. (8),

$$\begin{aligned} g'(x) &= -s\log(nx/R) \\ g''(x) &= -\frac{s}{x} \\ h''(x) &= (s-1)f''(x)+g''(x) = -\frac{1}{x}+\frac{2(s-1)}{1-x}+\frac{s-1}{d-2+x} \end{aligned} \tag{14}$$

and (for $d>3$)

$$h'''(x) = \frac{1}{x^2}+\frac{2(s-1)}{(1-x)^2}-\frac{s-1}{(d-2+x)^2} > 0. \tag{15}$$

Moreover, $g(x) = sx\left(\log \frac{R}{nx} - \frac{R}{nx} + 1\right) \le 0$, and so, using **P3** again,

$$h(1/d) = g(1/d) = O\left(\frac{s}{d}\left(\frac{dR}{n} - 1\right)^2\right) = o\left(\frac{1}{\nu}\right), \tag{16}$$

$$h'(1/d) = g'(1/d) = s\log(Rd/n) = o\left(\left(\frac{d}{\nu}\right)^{1/2}\right), \tag{17}$$

$$h''(1/d) = -d + O(1). \tag{18}$$

If all d_i are equal to d, $R = n/d$ and h has a (local) maximum 0 at $1/d$; in general h has a small positive maximum close to $1/d$.

As we will see shortly, the main contribution to the sum in (13) indeed comes from terms with k/ν close to $1/d$. Note, however, that for larger x, $h(x)$ has a minimum and then increases again, and we have to check that $h(x)$ is small for large x. We argue as follows, assuming that n, and thus d, is large enough (note that **P5** shows that $d \to \infty$).

By (14) and (15), h'' increases from $-\infty$ to $+\infty$ on $(0,1)$; hence h'' is negative, and h concave, on $(0, x_0]$ while h is convex on $[x_0, 1)$ for some $x_0 \in (0,1)$. By (14), $h''(1/2s) < 0$ and thus $x_0 > 1/2s$. Moreover, if $x \in (0, 2/d)$, then (14) yields

$$h''(x) < -d/2 + O(1) < -d/3 \tag{19}$$

and thus by a Taylor expansion and the inequality $ab \le 3a^2 + \frac{1}{12}b^2$ (a case of the arithmetic-geometric inequality), using (16) and (18),

$$\begin{aligned} h(x) &\le h\left(\frac{1}{d}\right) + \left(x - \frac{1}{d}\right)h'\left(\frac{1}{d}\right) - \frac{1}{2}\left(x - \frac{1}{d}\right)^2\frac{d}{3} \\ &\le 0 + \frac{3}{d}\left|h'\left(\frac{1}{d}\right)\right|^2 + \frac{1}{12}d\left(x - \frac{1}{d}\right)^2 - \frac{1}{6}d\left(x - \frac{1}{d}\right)^2 \\ &= -\frac{1}{12}d\left(x - \frac{1}{d}\right)^2 + o\left(\frac{1}{\nu}\right), \qquad 0 \le x \le \frac{2}{d}. \end{aligned} \tag{20}$$

In particular, $h(2/d) \le -\frac{1}{12d} + \frac{1}{\nu}$.

By (17) and (19), also

$$h'\left(\frac{2}{d}\right) \le h'\left(\frac{1}{d}\right) - \frac{d}{3}\cdot\frac{1}{d} = -\frac{1}{3} + o\left(\left(\frac{d}{\nu}\right)^{1/2}\right) < -\frac{1}{4},$$

and since h' is decreasing on $[0, x_0]$, we obtain

$$h(x) \le h\left(\frac{2}{d}\right) - \frac{1}{4}\left(x - \frac{2}{d}\right) \le -\frac{1}{12d} - \frac{1}{4}\left(x - \frac{2}{d}\right) + \frac{1}{\nu}, \qquad \frac{2}{d} \le x \le x_0.$$

It follows that the sum of the terms in (13) with $2/d \le k/\nu \le x_0$ is at most, for some $C < \infty$

$$C \sum_{j=0}^{\infty} \exp\left(-\frac{\nu}{12d} - \frac{1}{4}j + 1\right) = O\left(e^{-\nu/12d}\right) = o(1).$$

We also obtain

$$h(x_0) \le -\frac{x_0}{4} + O\left(\frac{1}{d}\right) \le -\frac{1}{9s},$$

while

$$f(1) = d\log d - (d-1)\log(d-1) = \log d + (d-1)\log\left(1 + \frac{1}{d-1}\right) < \log d + 1$$

and, using **P3** again,

$$\begin{aligned} g(1) &= s - R/\nu + s\log(R/n) = O(1) - s\log d + s\log(Rd/n) \\ &= -s\log d + O(1) \end{aligned}$$

whence

$$h(1) = (s-1)f(1) + g(1) < -\log d + O(1).$$

Since h is convex on $[x_0, 1]$ we thus find $h(x) \le \max(h(x_0), h(1)) \le -\frac{1}{9s}$ on $[x_0, 1]$, and the contribution to (13) from terms with $k/\nu > x_0$ is exponentially small.

Let Σ_1 denote the sum of the remaining terms in (13), i.e. the sum of the terms with $k < 2\nu/d$. Defining $k(y) = \lfloor \nu/d + y\sqrt{\nu/d} \rfloor = \nu(1/d + y/\sqrt{\nu d} + O(1/\nu))$, we rewrite the sum as an integral,

$$\Sigma_1 = \int_{-\sqrt{\nu/d}}^{\sqrt{\nu/d}+\theta\sqrt{d/\nu}} \sqrt{s}c_{sk(y)}e^{\nu h(k(y)/\nu)}\sqrt{\frac{\nu}{d}}\,dy, \tag{21}$$

where $0 \le \theta \le 1$. For any fixed real y, **P2** yields $k(y) \sim \nu/d$; hence $c_{sk(y)} \sim (2\pi n/d)^{-1/2}$ by Lemma 3. A Taylor expansion yields

$$\begin{aligned} h(k(y)/\nu) &= h(1/d) + \left(y/\sqrt{\nu d} + O\left(\frac{1}{\nu}\right)\right) h'(1/d) \\ &- \frac{1}{2}\left(y/\sqrt{\nu d} + O\left(\frac{1}{\nu}\right)\right)^2 h''\left(\frac{1}{d}\right) + O\left(\left(\frac{1}{\sqrt{\nu d}}\right)^3 d^2\right) \\ &= -\frac{y^2}{2\nu} + o\left(\frac{1}{\nu}\right). \end{aligned}$$

Hence the integrand in (21) is asymptotically equal to

$$s^{1/2}(2\pi n/d)^{-1/2}(\nu/d)^{1/2}e^{-y^2/2} = (2\pi)^{-1/2}e^{-y^2/2}.$$

We can use dominated convergence in (21), because (20) yields

$$h(k(y)/\nu) \le -\frac{d}{12}\left(\frac{y}{\sqrt{\nu d}}\right)^2 + O\left(\frac{1}{\nu}\right) = -\frac{y^2}{12\nu} + O\left(\frac{1}{\nu}\right)$$

while **P2** and **P3** yield $R^{-1} = O(d/\nu)$ and thus, rather crudely

$$c_j = O\left(\left(\frac{d}{\nu}\right)^{1/2} + \left(1 - \frac{dj}{n}\right)_+\right),$$

which implies $\sqrt{\frac{\nu}{d}}c_{sk(y)} = O\left(1 + |y|\right)$, so the integrand is $O((1+|y|)e^{-y^2/12})$. We thus get

$$\Sigma_1 \to \int_{-\infty}^{\infty} (2\pi)^{-1/2}e^{-y^2/2}dy = 1,$$

which completes the proof of Lemma 2. □

3 Sketch proof of Theorem 2

We obtain the same expressions for $\mathbf{E}(Y)$ and $\mathbf{E}(Y)^2$ as before, except that $n! = (s\nu)!$ is replaced by $\nu!^s$, and similarly $(sm)!$ by $m!^s$ and $(n - sm)!$

by $(\nu - m)!^s$; moreover, $D_{s(\nu-k)}$ is replaced by $\prod_{\ell=1}^s D^{(\ell)}_{\nu-k}$, where $D^{(\ell)}_i = \prod_{i=1}^{\nu} d_{\ell i} \sum_{|A|=j} \prod_{i\in A}(d_{\ell i} - 1)$, $(d_{\ell i})_i$ being the degrees in the ℓ'th partition.

This gives

$$\frac{\mathbf{E}(Y^2)}{\mathbf{E}(Y)^2} = \sum_{k=0}^{\nu} a_k \prod_{\ell} D^{(\ell)}_{\nu-k}$$

where Stirling's formula yields

$$a_k = \left(2\pi\nu x(1-x)^2 d(d-2+x)(d-1)^{-2}\right)^{(s-1)/2} (1+\varepsilon_k)e^{\nu(s-1)f(x)}$$

with the same $f(x)$ as before.

Let us define $R^{(\ell)} = \sum_{i=1}^{\nu} d_{\ell i}^{-1}$ and $R = \sum_{\ell=1}^{s} R^{(\ell)}$. Since $\log x$ is concave, $\sum_{\ell=1}^{s} \log R^{(\ell)} \le s\log(R/s)$, and Lemma 3 yields

$$\prod_{\ell=1}^{s} D^{(\ell)}_{\nu-j} \le \prod_{\ell=1}^{s} c^{(\ell)}_j \exp\left(sj - R - sj\log(sj/R)\right)$$

or

$$\prod_{\ell=1}^{s} D^{(\ell)}_{\nu-k} \le \prod_{\ell=1}^{s} c^{(\ell)}_k \exp(\nu g(k/\nu)).$$

The proof is completed as before with minor differences; the important fact is that for $x \sim 1/d$,

$$a_k \sim (2\pi\nu/d)^{(s-1)/2} e^{\nu(s-1)f(x)}$$

and

$$\prod_{\ell=1}^{s} c^{(\ell)}_k \sim \left(\frac{2\pi\nu}{d}\right)^{-s/2},$$

while we get a final factor $(2\pi\nu/d)^{1/2}$ from the integration over y as before.□

4 Complements

4.1 The number of perfect matchings

In the case $m/n^{4/3} \to \infty$ but $m = O(n^{3/2})$ (exactly the case *not* covered by Schmidt and Shamir [10]), our method also yields the asymptotic distribution of the number Y of perfect matchings, which turns out to be log-normal. For simplicity we consider only the multihypergraph $\tilde{G}(n,m,s)$; we believe that the result carries over to $G(n,m,s)$ and $G^*(n,m,s)$ (with slightly different constants) but we have not investigated this in detail. We conjecture also that the result below remains true when $m/n^{3/2} \to \infty$, which in that case would yield an asymptotic normal distribution of Y, but that cannot be proved by the present method.

Theorem 3 *Let $s \geq 2$ be fixed. If $m/n^{4/3} \to \infty$ but $m = O(n^{3/2})$, then, for the model $\tilde{G}(n,m,s)$,*

$$\frac{m}{n^{3/2}}(\ln Y - \alpha_n) \xrightarrow{d} \mathrm{N}\left(0, \frac{1}{2s^2}\right),$$

where

$$\alpha_n = \frac{n}{s}\ln\frac{sm}{n} - \left(1 - \frac{1}{s}\right)n - \frac{n^2}{2s^2m} - \frac{(3s+2)n^3}{12s^3m^2} + \tfrac{1}{2}\ln s.$$

This result is new also for the case $s = 2$ (graphs), although Janson [7] has obtained similar results for $G(n,m) = G(n,m,2)$ when $m/n^{3/2}$ is bounded *below*. In the overlapping case $m \asymp n^{3/2}$, the result of [7] can be written as above with α_n replaced by $\alpha_n + 3/4$; in other words, the distributions of Y for $G(n,m,2)$ and $\tilde{G}(n,m,2)$ differ asymptotically by a shift by a factor $e^{3/4}$. (It can be shown that the deficient edges (loops) in $\tilde{G}(n,m,2)$ contribute a factor $e^{-1/2}$ and the multiple edges a further factor $e^{-1/4}$.)

Proof The proof of Lemma 2 shows that if $\mathbf{d}$ is an $\epsilon(n)$-smooth degree sequence where $\epsilon(n) \to 0$, then for the model $G(\mathbf{x})$ with $\mathbf{x}$ chosen uniformly at random from $X(\mathbf{d})$,

$$\mathrm{Var}\left(\frac{Y}{\mathbf{E}Y}\right) = \frac{\mathbf{E}(Y^2)}{\mathbf{E}(Y)^2} - 1 \to 0.$$

Moreover,

$$\mathbf{E}Y = b_n \prod_{i=1}^{n} d_i$$

with

$$b_n = \frac{\psi(n)\psi(sm-n)}{\psi(sm)} = \frac{n!\,(sm-n)!\,m!}{\nu!\,(m-\nu)!\,(sm)!}.$$

Hence,

$$\frac{Y}{b_n \prod_i d_i} \xrightarrow{p} 1 \tag{22}$$

and thus

$$\ln Y - \ln\left(b_n \prod_i d_i\right) \xrightarrow{p} 0. \tag{23}$$

It follows by Lemma 1 that (22) and (23) hold also for $G(\mathbf{x})$ with $\mathbf{x}$ chosen uniformly at random from $[n]^{sm}$. In the remainder of the proof we will use this model, i.e. $\tilde{G}(n,m,s)$, and it suffices to show that for this model

$$\frac{m}{n^{3/2}}\left(\ln\left(b_n \prod_i d_i\right) - \alpha_n\right) \xrightarrow{d} \mathrm{N}(0, \tfrac{1}{2}s^{-2}), \tag{24}$$

or, equivalently, using again $d = sm/n$,

$$\frac{d}{n^{1/2}}\left(\ln\Big(\prod_i d_i\Big) + \ln b_n - \alpha_n\right) \xrightarrow{d} \mathrm{N}(0, \tfrac{1}{2}). \tag{25}$$

We first note that Stirling's formula yields

$$\begin{aligned}
\ln b_n &= (s-1)\nu\ln\nu + (s-1)(m-\nu)\ln(m-\nu) - (s-1)m\ln m \\
&\qquad + \tfrac{1}{2}\ln s + o(1) \\
&= (s-1)\nu((d-1)\ln(d-1) - d\ln d) + \tfrac{1}{2}\ln s + o(1) \\
&= (1-s^{-1})n\Big(-\ln d - 1 + \frac{1}{2d} + \frac{1}{6d^2} + O(d^{-3})\Big) + \tfrac{1}{2}\ln s + o(1).
\end{aligned} \tag{26}$$

Next, by the proof of Lemma 1 **whp** $d_{\min} > d/2$, and then a Taylor expansion yields

$$\begin{aligned}
\ln\prod_{i=1}^n d_i - n\ln d &= \sum_{i=1}^n \ln\left(1 + \frac{d_i - d}{d}\right) \\
&= \sum_{i=1}^n \left(\sum_{k=1}^5 \frac{(-1)^{k-1}}{k}\left(\frac{d_i-d}{d}\right)^k + O\left(\left(\frac{d_i-d}{d}\right)^6\right)\right) \\
&= \sum_{k=1}^5 \frac{(-1)^{k-1}}{k} d^{-k} S_k + O\left(d^{-6}S_6\right),
\end{aligned} \tag{27}$$

where

$$S_k = \sum_{i=1}^{n} (d_i - d)^k.$$

It remains to analyse these random variables. Note first that

$$S_1 = 0. \tag{28}$$

We claim that, for each $k \geq 2$,

$$\begin{aligned} \mathbf{E}S_k &= O\left(nd^{\lfloor k/2 \rfloor}\right), &(29)\\ \mathrm{Var} S_k &= O\left(nd^k\right); &(30) \end{aligned}$$

more precisely

$$\begin{aligned} \mathbf{E}S_2 &= nd(1 - \frac{1}{n}) = nd + O(d), &(31)\\ \mathbf{E}S_3 &= nd + O(d), &(32)\\ \mathbf{E}S_4 &= 3nd^2 + O(nd), &(33) \end{aligned}$$

and, moreover,

$$n^{-1/2} d^{-1} (S_2 - \mathbf{E}S_2) \xrightarrow{d} \mathrm{N}(0, 2). \tag{34}$$

Postponing the proof of these claims, we find

$$\mathrm{Var}\left(dn^{-1/2} d^{-k} S_k\right) = O(d^{2-k}) = o(1), \qquad k \geq 3, \tag{35}$$

and thus

$$dn^{-1/2}\left(d^{-k} S_k - d^{-k}\mathbf{E}S_k\right) \xrightarrow{p} 0, \qquad k \geq 3. \tag{36}$$

Moreover, if $k \geq 5$, then

$$dn^{-1/2} d^{-k} \mathbf{E}S_k = O\left(n^{1/2} d^{1-k+\lfloor k/2 \rfloor}\right) = O\left(n^{1/2} d^{-2}\right) = o(1), \tag{37}$$

and thus also

$$dn^{-1/2} d^{-k} S_k \xrightarrow{p} 0, \qquad k \geq 5. \tag{38}$$

Using (28), (38) and (36) in (27), we obtain

$$dn^{-1/2}\left(\ln \prod_i d_i - \left(n \ln d - \tfrac{1}{2} d^{-2} S_2 + \tfrac{1}{3} d^{-3} \mathbf{E}S_3 - \tfrac{1}{4} d^{-4} \mathbf{E}S_4\right)\right) \xrightarrow{p} 0, \tag{39}$$

and, by (31)–(33),

$$dn^{-1/2}\left(\ln\prod_i d_i - \left(n\ln d - \tfrac12 nd^{-1} + \tfrac13 nd^{-2} - \tfrac34 nd^{-2}\right)\right) \\ + \tfrac12 n^{-1/2}d^{-1}\left(S_2 - \mathbf{E}S_2\right) \xrightarrow{p} 0, \qquad (40)$$

whence (36) yields

$$dn^{-1/2}\left(\ln\prod_i d_i - n\ln d + \tfrac12 nd^{-1} + \tfrac{5}{12}nd^{-2}\right) \xrightarrow{d} \mathrm{N}(0,\tfrac12). \qquad (41)$$

Finally, (25) follows by (41) together with (26), which easily yields

$$\ln b_n + n\ln d - \tfrac12 nd^{-1} - \tfrac{5}{12}nd^{-2} = \alpha_n + o(1). \qquad (42)$$

It remains to verify (29)–(34). We will be somewhat sketchy. First, $\mathbf{E}S_k = n\mathbf{E}(d_1-d)^k$, with $d_1 \sim \mathrm{Bi}(nd, n^{-1})$. We introduce the centered indicator variables

$$I'_{ji} = \begin{cases} 1-\frac1n, & \text{if } x_j = i,\\ -\frac1n, & \text{if } x_j \neq i,\end{cases} \qquad 1\le j\le nd,\ 1\le i\le n.$$

Then $d_i - d = \sum_j I'_{ji}$, and standard arguments yield (expanding the power and using the independence of the x_i)

$$\mathbf{E}(d_i-d)^k = \mathbf{E}\Big(\sum_j I'_{ji}\Big)^k = O(d^{\lfloor k/2\rfloor}).$$

This proves (29) and more careful estimates yield (31)–(33). (Alternatively, we may verify (31)–(33), and (29) for the values of k that we need ($k\le 6$), using formulae for the first moments of a binomial distribution.)

It may be similarly shown that

$$\mathrm{Cov}\left((d_1-d)^k,(d_2-d)^k\right) = O(n^{-1}d^k)$$

and thus

$$\mathrm{Var}S_k \le n\mathbf{E}(d_1-d)^{2k} + n(n-1)\mathrm{Cov}\left((d_1-d)^k,(d_2-d)^k\right) = O(nd^k),$$

which is (30).

Finally, the central limit theorem (34) may be proved by the method of moments. For example, if we introduce the indicator variables $J_{jk} = I(x_j = x_k)$, $j, k = 1, \dots, nd$, then

$$S_2 = \sum_{i=1}^{n} d_i^2 - nd^2 = \sum_{j,k=1}^{nd} J_{jk} - nd^2$$

and thus

$$S_2 - \mathbf{E}S_2 = \sum_{j,k=1}^{nd} (J_{jk} - \mathbf{E}J_{jk}) = 2\sum_{j<k}(J_{jk} - \mathbf{E}J_{jk}).$$

It is now fairly straightforward to estimate the moments (or, simpler, cumulants) of $S_2 - \mathbf{E}S_2$ and obtain (34). We omit the details. □

4.2 Further remarks

As stated above, Schmidt and Shamir [10] used the second moment method directly on $\mathcal{G}(n, m, s)$, thus proving the result when Y can be approximated (in L^2) by some non-random number c_n (depending on n and m), i.e. when $Y/c_n \to 1$ in L^2 for some c_n (say $c_n = \mathbf{E}Y$). The conditioning used above means that in this paper we treat the more general case when Y can be approximated by a function of the vertex degrees. The conditional expectation of Y given the vertex degees is given by (6), which shows that the appropriate approximating function is a constant times the product of the vertex degrees; the proof of Theorem 3 shows that we can write this as

$$c_n \exp\left(\sum \log d_i\right) \approx c_n' \exp\left(-\sum (d_i - d)^2/2d\right)$$

and thus we can as well approximate by a function of n and the sums of vertex degrees and squares of vertex degrees. In the case $s = 2$ (graphs) this is the same as taking a function of the numbers of vertices and edges and the number of paths of length two in the random graph, or equivalently, the numbers of trees of orders 1, 2 and 3 in the graph. This was suggested as a probable good approximation (for the number of perfect matchings as well as for some other related variables) in Janson [7]. The argument there was not rigorous, but was based on an orthogonal expansion which suggested that only these terms are important when $d \gg n^{1/3}$, just as only the total number of edges is important when $d \gg n^{1/2}$. The same argument suggests that a suitable approximation for $d \gg n^{1/4}$ can be found as an (exponential) function of the numbers of trees of orders up to 4 in the random graph. For hypergraphs the corresponding thing would be to take the numbers of pairs and triples of hyperedges that are connected in the edge graph. The same ought to apply for $d \gg n^{1/\ell}$ for any fixed ℓ, (counting trees up to order ℓ), but we see no way of performing a variance calculation to justify this rigorously.

References

[1] E.A.Bender and E.R.Canfield, The asymptotic number of labelled graphs with given degree sequences, *Journal of Combinatorial Theory, Series A*, 24 (1978) 296-307.

[2] B.Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, *European Journal on Combinatorics*, 1 (1980) 311-316.

[3] B.Bollobás and A.M.Frieze, *On matchings and hamiltonian cycles in random graphs*, Annals of Discrete Mathematics 28, (1985) 23-46.

[4] B.Bollobás, T.I.Fenner and A.M.Frieze, *Hamilton cycles in random graphs with minimal degree at least k*, in A tribute to Paul Erdős, edited by A.Baker, B.Bollobás and A.Hajnal,1990) 59 - 96.

[5] C.Cooper, A.M.Frieze, M.Molloy and B.Reed, *Perfect matchings in random r-regular, s-uniform hypergraphs*, Combinatorics, Probability and Computing, to appear.

[6] P.Erdös 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.

[7] S.Janson, *The numbers of spanning trees, Hamilton cycles and perfect matchings in a random graph*, Combinatorics, Probability and Computing, to appear.

[8] R.W.Robinson and N.C.Wormald, *Almost all cubic graphs are Hamiltonian*, Random Structures and Algorithms 3 (1992) 117-126.

[9] R.W.Robinson and N.C.Wormald, *Almost all regular graphs are Hamiltonian*, Random Structures and Algorithms to appear.

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