Repeated degrees in random uniform hypergraphs

Paul Balister

Department of Mathematical Sciences
The University of Memphis, Tennessee, USA
pbalistr@memphis.edu

Béla Bollobás*

Department of Pure Mathematics and
Mathematical Statistics
University of Cambridge Centre for Mathematical Sciences
Wilberforce Road Cambridge CB3 0WB, UK
B.Bollobas@dpmms.cam.ac.uk
and
Department of Mathematical Sciences
The University of Memphis, Tennessee, USA
bollobas@msci.memphis.edu

Jenő Lehel

Department of Mathematical Sciences
The University of Memphis, Tennessee, USA
jlehel@memphis.edu

Michał Morayne†

Institute of Mathematics and Computer Science
Wrocław University of Technology
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
michal.morayne@pwr.wroc.pl

Abstract

We prove that in a random 3-uniform or 4-uniform hypergraph of order n the probability that some two vertices have the same degree tends to one as $n \to \infty$.

1 Introduction

For $r \geq 2$, an *r-uniform hypergraph* is a pair $H = (V, E)$ consisting of a finite set V, called the set of vertices, and a set E of distinct r-element subsets of V, called the set of edges. The *degree* of a vertex $v \in V$ is $d_H(v) = |\{e \in E : v \in e\}|$. The hypergraph H is called *degree irregular* if its vertex degrees are all distinct, i.e., $d_H(v) = d_H(u)$ implies $u = v$.

Repeated degrees in graphs (the case $r = 2$) were discussed in [2] (see also [3]). The study of degree irregular hypergraphs started with [5]. In [6] it is shown that there exist degree irregular r-uniform hypergraphs of order n if and only if $r \geq 3$ and $n \geq r + 3$. Furthermore, it is proved that almost every r-uniform hypergraph is degree irregular for $r \geq 6$ when the edges emerge independently with probability $p = 1/2$. Here we discuss the cases $r = 3$, 4 when the probability $p = p_n$ depends on n.

We show that for $r = 3$ and 4, a random r-uniform hypergraph of order n has vertices of the same degree with probability approaching one as $n \to \infty$. The cases $r \geq 5$, where the same is not true, will be discussed in [1] in a slightly more general setting and using different techniques.

Here we shall use the Central Limit Theorem (actually, the rate of convergence estimate due to Berry and Esséen), and the second moment method to show that for $r = 3$, 4 the values of the vertex degrees are concentrated in a 'small' interval of integers where the repetition of degrees will occur with high probability. In the case of $r = 3$ the repetition of a degree is guaranteed by the elementary pigeonhole principle (see Theorem 5). In the case of $r = 4$ (see Theorem 13) we apply the birthday paradox (Lemma 6) which essentially says that if elements are randomly (independently and

*Research supported in part by NSF grants DMS-0906634, CNS-0721983 and CCF-0728928, ARO grant W911NF-06-1-0076, and TAMOP-4.2.2/08/1/2008-0008 program of the Hungarian National Development Agency

†Partially supported by grant MNiSW N N 206 1842 33; written in part when M.M. visited the Department of Mathematical Sciences of the University of Memphis in Spring, 2008

[resubmit_SIDMA_2march.tex] Printed on March 2, 2012

uniformly) selected from an n element set, and the selection is repeated $c_n\sqrt{n}$ times, where $c_n \to \infty$ is any sequence, then the probability of selecting some element more than once approaches 1 as $n \to \infty$ (see [7]).

2 Preliminaries

Consider the probability space of all r-uniform hypergraphs with n (labelled) vertices, and the edges chosen independently with probability p_n depending on n. This space is referred to as a random r-uniform hypergraph of order n and denoted by $H_r(n, p_n)$. With some abuse of terminology we shall refer to its elementary events as *random r-uniform hypergraphs.*

Let S_N be the number of successes in N Bernoulli trials with the probability of success in a single trial equal to p_n. Observe that for $N = \binom{n-1}{r-1}$, S_N is the degree of any fixed vertex in $H_r(n, p_n)$. In the sequel we set $q_n := 1 - p_n$ (and $q := 1 - p$).

For the rate of convergence in the Central Limit Theorem we use an estimation due to Berry and Esséen (see [4] p.542) as follows.

Theorem (Berry-Esséen)**.** *Let $Z_1, Z_2, \ldots, Z_k$ be independent identically distributed random variables with mean $\mathbb{E}[Z_1] = 0$, variance $\sigma^2 = \mathbb{E}[Z_1^2] > 0$, and $\rho = \mathbb{E}[|Z_1|^3] < \infty$. Then*

$$\left| \mathbb{P}\left[\frac{\sum_{i=1}^{k} Z_i}{\sigma\sqrt{k}} \leq x \right] - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt \right| \leq \frac{3\rho}{\sigma^3\sqrt{k}} .$$

□

For a random hypergraph $H_r(n, p)$ and integer $\Delta > 0$, let

$$Z_\Delta = |\{v \in V : Np - \Delta < d(v) \leq Np + \Delta\}| .$$

Notice that the concentration of the degrees in the interval of radius Δ around the mean Np is described by $Z_\Delta/2\Delta$. We shall need two technical lemmas about the expectations $\mathbb{E}[Z_\Delta]$ and $\mathbb{E}[Z_\Delta^2]$. For A, B depending on n, we shall use the notation $A \sim B$ provided $A/B \to 1$ as $n \to \infty$.

Lemma 1. *If $\Delta \to \infty$ as $n \to \infty$, and $\lim_{n\to\infty} \Delta/\sqrt{Np_nq_n} = 0$, then*

$$\mathbb{E}[Z_\Delta]^2 \sim \frac{2\Delta^2(r-1)!}{\pi n^{r-3} p_n q_n} .$$

Proof. For $i = 1, \dots, N$, let Z_i be the two valued random variable taking on the value $1 - p = q$ with probability $p = p_n$, and the value $-p$ with probability q. Clearly $\mathbb{E}[Z_i] = 0$, $\sigma^2 = pq$, $\rho = q^3p + p^3q \le pq$, and hence $3\rho/(\sigma^3\sqrt{N}) \le 3/\sqrt{Npq}$. Furthermore, we have $\sum_{i=1}^N Z_i = S_N - Np$. Applying the Berry-Esséen theorem with $Z_1, \dots, Z_N$ and $x = \pm\Delta/\sqrt{Npq}$ we obtain easily that

$$\left|\mathbb{P}\left[-\frac{\Delta}{\sqrt{Npq}} < \frac{S_N - Np}{\sqrt{Npq}} \le \frac{\Delta}{\sqrt{Npq}}\right] - \frac{1}{\sqrt{2\pi}}\int_{-\frac{\Delta}{\sqrt{Npq}}}^{\frac{\Delta}{\sqrt{Npq}}} e^{-t^2/2}\,dt\right| \le \frac{6}{\sqrt{Npq}}\ ,$$

The assumption $\lim\limits_{n\to\infty} \Delta/\sqrt{Npq} = 0$ implies that

$$\frac{1}{\sqrt{2\pi}}\int_{-\frac{\Delta}{\sqrt{Npq}}}^{\frac{\Delta}{\sqrt{Npq}}} e^{-t^2/2}\,dt \sim \frac{1}{\sqrt{2\pi}}\frac{2\Delta}{\sqrt{Npq}}\ .$$

Thus as $N \sim n^{r-1}/(r-1)!$ and $\Delta \to \infty$ we obtain that

$$\mathbb{E}[Z_\Delta] \sim n \cdot \frac{1}{\sqrt{2\pi}}\frac{2\Delta}{\sqrt{Npq}} \sim \sqrt{\frac{2\Delta^2(r-1)!}{\pi n^{r-3}pq}}\ , \tag{1}$$

and the lemma follows. □

Lemma 2. *If $Np_n \to \infty$ and $Nq_n \to \infty$ as $n \to \infty$, then*

$$\mathbb{E}[Z_\Delta^2] \le (1 + o(1))\left(\mathbb{E}[Z_\Delta] + \frac{2\Delta^2(r-1)!}{\pi n^{r-3}p_nq_n}\right)\ .$$

Proof. Set $A = Np_n - \Delta$, $B = Np_n + \Delta$, and let $V = \{v_1, \dots, v_n\}$ be the vertex set of the random hypergraph. For $1 \le i \le n$ and an integer $k \in (A, B]$, define

$$L_i^{(k)} = \begin{cases} 1 & \text{if } d(v_i) = k \\ 0 & \text{otherwise.} \end{cases}$$

Let $X_i = \sum_{k\in(A,B]} L_i^{(k)}$ be the random variable that is 1 if $d(v_i) \in (A, B]$ and 0 otherwise, so $Z_\Delta = \sum_{i=1}^n X_i$. Thus we have

$$\begin{aligned}\mathbb{E}[Z_\Delta^2] &= \mathbb{E}\left[\left(\textstyle\sum_{i=1}^n X_i\right)^2\right] = \sum_{i=1}^n \mathbb{E}[X_i^2] + \sum_{1\le i\ne j\le n} \mathbb{E}[X_iX_j] \\ &= \mathbb{E}[Z_\Delta] + n(n-1)\mathbb{E}[X_1X_2] = \mathbb{E}[Z_\Delta] + n(n-1)\sum_{k,\ell\in(A,B]} \mathbb{E}[L_1^{(k)}L_2^{(\ell)}] \\ &\le \mathbb{E}[Z_\Delta] + n^2(2\Delta)^2 \max_{k,\ell\in(A,B]} \mathbb{E}[L_1^{(k)}L_2^{(\ell)}]\ .\end{aligned}$$

Next we estimate $\mathbb{E}[L_1^{(k)}L_2^{(\ell)}]$. For an integer M, let $\mu(M,p)$ denote that value of k for which $\binom{M}{k}p^kq^{M-k}$ is maximal.

Let $M=\binom{n-2}{r-1}\sim\frac{n^{r-1}}{(r-1)!}$. The assumptions $Np_n\to\infty$ and $Nq_n\to\infty$ imply that $\mu(M,p_n)\sim Mp_n$ and $M-\mu(M,p_n)\sim Mq_n$. Furthermore, by applying Stirling's formula we have that

$$\binom{M}{\mu(M,p_n)}p_n^{\mu(M,p_n)}q_n^{M-\mu(M,p_n)}\sim\frac{1}{\sqrt{2\pi Mp_nq_n}}.$$

Let $D(m)$ be the event that there are exactly m edges containing both v_1 and v_2. We obtain

$$\begin{aligned}
\mathbb{E}\left[L_1^{(k)}L_2^{(\ell)}\right] &= \mathbb{P}\left[(d(v_1)=k)\wedge(d(v_2)=\ell)\right]\\
&= \sum_{m\le\min(k,\ell)}\mathbb{P}\left[(d(v_1)=k)\wedge(d(v_2)=\ell)\mid D(m)\right]\cdot\mathbb{P}(D(m))\\
&= \sum_{m\le\min(k,\ell)}\mathbb{P}\left[(d(v_1)=k)\mid D(m)\right]\cdot\mathbb{P}\left[(d(v_2)=\ell)\mid D(m)\right]\cdot\mathbb{P}(D(m))\\
&= \sum_{m\le\min(k,\ell)}\binom{M}{k-m}p_n^{k-m}q_n^{M-k+m}\cdot\binom{M}{\ell-m}p_n^{\ell-m}q_n^{M-\ell+m}\cdot\mathbb{P}(D(m))\\
&\le \sum_{m=0}^{M}\left[\binom{M}{\mu(M,p_n)}p_n^{\mu(M,p_n)}q_n^{M-\mu(M,p_n)}\right]^2\mathbb{P}(D(m))\\
&= \left[\binom{M}{\mu(M,p_n)}p_n^{\mu(M,p_n)}q_n^{M-\mu(M,p_n)}\right]^2\sim\frac{1}{2\pi Mp_nq_n}\sim\frac{(r-1)!}{2\pi n^{r-1}p_nq_n}.
\end{aligned}$$

□

3 3-uniform hypergraphs

The full description of the behavior of $H_3(n,p_n)$ concerning the existence of vertices of the same degree will be given in Theorem 5. This result will be obtained through Theorems 3 and 4 that treat particular cases. For $r=3$, we have $N=\binom{n-1}{2}\sim n^2/2$.

Theorem 3. *If $p\in(0,1)$ and $p_n\to p$, then the probability that the random 3-uniform hypergraph $H_3(n,p_n)$ has two vertices of the same degree tends to one as $n\to\infty$.*

Proof. Let $\Delta = \lceil \alpha_n n \rceil$ such that $\Delta \to \infty$ and $\alpha_n \to 0$ as $n \to \infty$. Because $\lim_{n\to\infty} \Delta/\sqrt{p_n q_n N} = \sqrt{\frac{2}{pq}} \lim_{n\to\infty} \alpha_n = 0$, we may apply Lemma 1. Then $\mathbb{E}[Z_\Delta] \sim \frac{2\Delta}{\sqrt{\pi pq}}$ and since $1/\sqrt{pq} \geq 2$, we obtain that

$$\lim_{n\to\infty} \frac{\mathbb{E}[Z_\Delta]}{2\Delta} \sim \frac{1}{\sqrt{\pi pq}} \geq \frac{2}{\sqrt{\pi}} > 1. \tag{2}$$

Applying Lemma 2 with $r = 3$, we have $\mathbb{E}[Z_\Delta^2] \leq (1 + o(1))\frac{4\Delta^2}{\pi pq} \sim \mathbb{E}[Z_\Delta]^2$. As $\mathbb{E}[Z_\Delta^2] \geq \mathbb{E}[Z_\Delta]^2$, we obtain

$$\lim_{n\to\infty} \frac{\mathbb{E}[Z_\Delta]^2}{\mathbb{E}[Z_\Delta^2]} = 1 \ . \tag{3}$$

Let $0 < \beta < 1 - \frac{\sqrt{\pi}}{2}$. By Chebyshev's inequality (see [**?**]) and by (3) we obtain

$$\mathbb{P}[|Z_\Delta - \mathbb{E}[Z_\Delta]| \geq \beta\mathbb{E}[Z_\Delta]] \leq \frac{\mathbb{E}[Z_\Delta^2] - \mathbb{E}[Z_\Delta]^2}{\beta^2\mathbb{E}[Z_\Delta]^2} \to 0 \ .$$

Hence

$$\begin{aligned}\mathbb{P}[Z_\Delta > (1-\beta)\mathbb{E}[Z_\Delta]] \geq \mathbb{P}[|Z_\Delta - \mathbb{E}[Z_\Delta]| < \beta\mathbb{E}[Z_\Delta]]& \\ = 1 - \mathbb{P}[|Z_\Delta - \mathbb{E}[Z_\Delta]| \geq \beta\mathbb{E}[Z_\Delta]] &\to 1\end{aligned}$$

follows, and by (2), this implies

$$\mathbb{P}\left[\frac{Z_\Delta}{2\Delta} > (1-\beta)\frac{2}{\sqrt{\pi}}\right] \to 1.$$

As $(1-\beta)\frac{2}{\sqrt{\pi}} > 1$, this implies that $\mathbb{P}[Z_\Delta > 2\Delta] \to 1$. Hence, with probability tending to one, some degrees will repeat by the pigeonhole principle. □

Theorem 4 below serves to complement Theorem 3 in the proof of Theorem 5. It will be applied also with $r = 4$ in the proof of Theorem 13.

Theorem 4. *Fix $t > 0$. If $r \geq 3$ and $p_n = o(n^{3-r})$, then the probability that the random r-uniform hypergraph $H_r(n, p_n)$ has more than t vertices of the same degree tends to one as $n \to \infty$.*

Proof. Let $N = \binom{n-1}{r-1}$ and $p = p_n = \alpha_n n^{3-r}$, $\alpha_n \to 0$. We distinguish two cases according to the sequence $n\alpha_n^{1/2}$.

Case 1: assume that $n\alpha_n^{1/2} \to \infty$.
Let $\Delta = \lceil n\alpha_n^{1/4} \rceil$, so $\Delta \to \infty$. We shall use Chebyshev's inequality in the following form:

$$\mathbb{P}[|S_N - Np| \geq u] \leq \frac{Npq}{u^2} \quad \text{for } u > 0. \tag{4}$$

By (4) we have

$$\begin{aligned}\mathbb{E}[n - Z_\Delta] &= \sum_{v \in V} \mathbb{P}[|d(v) - Np| \geq \Delta] = n \cdot \mathbb{P}[|S_N - Np| \geq \Delta] \leq n \cdot \frac{Npq}{\Delta^2} \\ &\leq n \cdot \frac{n^{r-1}(\alpha_n n^{3-r})}{(n\alpha_n^{1/4})^2} = n\alpha_n^{1/2}.\end{aligned}$$

Then, using Markov's inequality we obtain $\mathbb{P}[Z_\Delta \leq n - n\alpha_n^{1/4}] \leq (n\alpha_n^{1/2})/(n\alpha_n^{1/4}) = \alpha_n^{1/4} \to 0$, in other words, with probability tending to 1, more than $n - n\alpha_n^{1/4}$ vertices have degree from an interval of length $2\Delta = 2\lceil n\alpha_n^{1/4} \rceil$. Since $(n - n\alpha_n^{1/4})/(2n\alpha_n^{1/4}) \to \infty$ as $n \to \infty$, the ratio $Z_\Delta/2\Delta$ exceeds t. Hence, by the pigeonhole principle, there exist more than t vertices with the same degree.

Case 2: assume that $n\alpha_n^{1/2} < K$ for some constant K.
If X is the number of edges of $H_r(n,p)$, then $\mathbb{E}[X] = \binom{n}{r}p < n^r p$.

If d_k is the number of vertices of degree k, then $\sum_{k=0}^{N} d_k = n$. Assuming that there are not more than t vertices of the same degree, we have the following lower bound for the number of edges:

$$X = \frac{1}{r}\sum_{k=0}^{N} k d_k \geq \frac{1}{r}\sum_{k=0}^{\lfloor n/t \rfloor - 1} kt = \frac{t}{r} \cdot \frac{(\lfloor n/t \rfloor - 1)\lfloor n/t \rfloor}{2} \geq \frac{(n-2t)^2}{2rt}.$$

Meanwhile using Markov's inequality we obtain

$$\mathbb{P}[X \geq (n-2t)^2/2rt] \leq \frac{n^r p}{(n-2t)^2/(2rt)} \sim 2rtn\alpha_n \leq \frac{2K^2rt}{n} \to 0.$$

Therefore the probability that for some k, $0 \leq k \leq N$, there are more than t vertices of degree k tends to 1.

Assume now that $p_n = o(n^{3-r})$ and the conclusion of the theorem does not follow. Then there is an $\varepsilon > 0$, a constant t, and an infinite subsequence $n_1, n_2, \ldots$ such that for each i the probability that the hypergraph $H_r(n_i, p_{n_i})$ has more than t vertices of the same degree is smaller than $1-\varepsilon$. Then we can choose a subsequence $(n_{i_j})_j$ for which either $n_{i_j}\alpha_{n_{i_j}}^{1/2} < K$ holds, with some constant K, or $n_{i_j}\alpha_{n_{i_j}}^{1/2} \to \infty$. Then, Case 2 or Case 1 yields a contradiction. □

For $r = 3$ the condition in Theorem 4 becomes $p_n \to 0$. Note that, by symmetry, the conclusion remains true if $p_n \to 1$.

Theorem 5. *The probability that the random 3-uniform hypergraph of order n has two vertices of the same degree tends to one as $n \to \infty$.*

Proof. Let $p_n \in [0, 1]$, and assume that the conclusion is not true. Then there exists $\varepsilon > 0$ and an infinite subsequence $n_1, n_2, \ldots$ such that the probability that there are two vertices of the same degree in $H_3(n_i, p_{n_i})$ is smaller than $1-\varepsilon$ for every i. Then choose a convergent subsequence $p_{n_{ij}} \to p$. If $p \in (0, 1)$, then a contradiction follows from Theorem 3. If $p = 0$ or 1, a contradiction follows from Theorem 4 with $r = 3$. □

4 4-uniform hypergraphs

The general birthday problem, originated in the birthday paradox (see [7]), is as follows: given t random integers drawn independently and uniformly from the interval $[1, s]$, what is the probability $p(t; s)$ that at least two integers are the same? It is well-known and can be checked easily that as $t \to \infty$,

$$p(t; s) = 1 - \prod_{k=1}^{t-1}\left(1 - \frac{k}{s}\right) \sim 1 - e^{-(t(t-1))/2s} .$$

This implies the following lemma that will be used in the proof of Theorem 13.

Lemma 6. *If $t = c_s\sqrt{s}$ elements are randomly selected independently and uniformly from an s element set, where $c_s \to \infty$ is any sequence, then the probability of selecting some element more than once tends to one as $s \to \infty$.* □

We shall show that the number of the degree values in a suitable set of Δ possible values is much larger than $\sqrt{\Delta}$, just as required by the birthday paradox. However, Lemma 6 does not apply directly since the distribution of the degrees of the vertices is not uniform, and the degrees are not independent. Thus we shall need further results to be used in the proof of Theorem 13.

First we deal with the problem of independence. Choose a fixed set S of vertices with $s := |S| = o(n)$. For definiteness, let $\alpha_n \to 0$ and assume $s = s_n \sim \alpha_n^3 n \to \infty$. Split the edge set of the hypergraph H into two hypergraphs $H_1 = (V, E_1)$ and $H_2 = (V, E_2)$ with edge sets

$$\begin{aligned} E_1 &:= \{e \in E : |e \cap S| \leq 1\}, \text{ and} \\ E_2 &:= \{e \in E : |e \cap S| \geq 2\}. \end{aligned}$$

Recall that the maximum possible degree of a vertex in H is $N := \binom{n-1}{3}$, and write $N = N_1 + N_2$ where $N_1 = \binom{n-s}{3}$ and $N_2 = N - N_1 \sim sn^2/2$ are the maximum possible vertex degrees in H_1 and H_2 respectively. Write $d_i(v)$ for the degree of the vertex v in H_i, $i = 1, 2$, and set

$$S' := \{v \in S : N_2 p_n - \Delta < d_2(v) \leq N_2 p_n + \Delta\}$$

where $\Delta := \alpha_n \sqrt{N p_n q_n}$ is assumed to be an integer. Our first aim is to show that S' is large with probability approaching to 1.

Lemma 7. *Let $\Delta = \alpha_n \sqrt{N p_n q_n}$ and $|S| \sim \alpha_n^3 n$ with $\alpha_n \to 0$. Then $\mathbb{P}[|S'| > |S|/2] \to 1$ as $n \to \infty$.*

Proof. Fix a vertex $v \in S$. Then $d_2(v)$ is distributed as a Binomial $B(N_2, p_n)$ random variable. Thus by Chebyshev's inequality,

$$\mathbb{P}[d_2(v) \notin (N_2 p_n - \Delta, N_2 p_n + \Delta]] \leq \frac{N_2 p_n q_n}{\Delta^2} = \frac{N_2}{\alpha_n^2 N} \sim \frac{|S| \cdot n^2/2}{\alpha_n^2 n^3/6} \sim 3\alpha_n.$$

Thus

$$\mathbb{E}[|S| - |S'|] \lesssim 3\alpha_n |S|,$$

and so by Markov's inequality,

$$\mathbb{P}[|S'| \leq |S|/2] = \mathbb{P}[|S| - |S'| \geq |S|/2] \lesssim 6\alpha_n \to 0$$

as $n \to \infty$. The result follows. □

Now define

$$Z'_\Delta = |\{v \in S' : N p_n - 2\Delta < d(v) < N p_n - \Delta\}|.$$

The reason for the rather asymmetric definition of Z'_Δ will become apparent later when we consider the non-uniformity of the degree distribution in the proof of Theorem 12.

We note that when conditioning on H_2 (and hence also on the set S'), $d(v) = d_1(v)+d_2(v)$, where $d_2(v)$ is determined by H_2 and the $d_1(v)$ is a Binomial $B(N_1,p_n)$ random variable. We also note that the $d_1(v)$ are independent for $v \in S$ as no edge from H_1 can contribute to more than one of these degrees.

Lemma 8. *If $\Delta = \alpha_n\sqrt{Np_nq_n} \to \infty$, with $\alpha_n \to 0$ and $|S| = o(n)$, then*

$$\mathrm{Var}[Z'_\Delta \mid H_2] \le \mathbb{E}[Z'_\Delta \mid H_2] \sim \frac{\alpha_n|S'|}{\sqrt{2\pi}}\ .$$

Proof. Set $p = p_n$ and $q = q_n$. Let I_v be the indicator function of the event that $Np - 2\Delta < d(v) < Np - \Delta$. Now conditioned on H_2 and assuming $v \in S'$, we can write $d_2(v) = N_2p + \lambda_v$ where $|\lambda_v| \le \Delta$. Thus I_v is also the indicator function of the event that

$$N_1p - (2\Delta + \lambda_v) < d_1(v) \le N_1p - (\Delta + \lambda_v)$$

and $d_1(v)$ has a Binomial $B(N_1,p)$ distribution. The proof for the expectation now follows from the same argument as in Lemma 1 with $r = 4$ and only a few minor differences. By the Berry-Esséen theorem we have

$$\begin{aligned}\mathbb{E}[I_v \mid H_2] &= \mathbb{P}(N_1p - (2\Delta + \lambda_v) < d_1(v) \le N_1p - (\Delta + \lambda_v)]\\ &\sim \frac{1}{\sqrt{2\pi}}\int_{-\frac{2\Delta+\lambda_v}{\sqrt{N_1pq}}}^{-\frac{\Delta+\lambda_v}{\sqrt{N_1pq}}} e^{-t^2/2}dt \sim \frac{\Delta}{\sqrt{2\pi N_1pq}}\end{aligned}$$

with additive error at most $6/\sqrt{N_1pq}$ in the first approximation (as in Lemma 2). The second approximation follows as $2\Delta + \lambda_v = o(\sqrt{N_1pq})$. As $\Delta \to \infty$ we obtain

$$\mathbb{E}[I_v \mid H_2] \sim \frac{\Delta}{\sqrt{2\pi N_1pq}}.$$

Now as $s = |S| = o(n)$ we have $N_1 = \binom{n-s}{3} \sim \binom{n-1}{3} = N$, so

$$\mathbb{E}[Z'_\Delta \mid H_2] \sim |S'| \cdot \frac{\Delta}{\sqrt{2\pi N_1pq}} \sim \frac{\alpha_n|S'|}{\sqrt{2\pi}}\ .$$

For the variance we recall that, conditioned on H_2, the $d_1(v)$ are independent for distinct v, so writing $p_v = \mathbb{E}[I_v \mid H_2]$ we have

$$\mathrm{Var}[Z'_\Delta \mid H_2] = \sum_{v\in S'} \mathrm{Var}[I_v \mid H_2] = \sum_{v\in S'} p_v(1-p_v) \le \sum_{v\in S'} p_v = \mathbb{E}[Z'_\Delta \mid H_2].$$

$\square$

Lemma 9. *Assume that* $|S| \sim \alpha_n^3 n$ *and* $\Delta = \alpha_n\sqrt{Np_nq_n} \to \infty$ *where* $\alpha_n \to 0$, $\alpha_n n^{1/18} \to \infty$. *Then*

$$\lim_{n\to\infty} \mathbb{P}[Z'_\Delta \geq \alpha_n^{-1}\sqrt{2\Delta} \mid H_2] = 1$$

provided that $|S'| > |S|/2$.

Proof. For this proof we shall assume all probabilities and expectations are conditioned on H_2 and that $|S'| \geq |S|/2$. We may apply Lemma 8 because $\Delta/\sqrt{Np_nq_n} = \alpha_n \to 0$ and $\Delta \to \infty$. Thus we have

$$\mathrm{Var}[Z'_\Delta] \leq \mathbb{E}[Z'_\Delta] \sim \frac{\alpha_n|S'|}{\sqrt{2\pi}}. \tag{5}$$

Thus for sufficiently large n we may assume

$$\mathbb{E}[Z'_\Delta] \geq \tfrac{1}{6}\alpha_n^4 n \to \infty$$

as $\alpha_n \gg n^{-1/18}$. Applying Chebyshev's inequality we obtain

$$\mathbb{P}\left[|Z'_\Delta - \mathbb{E}[Z'_\Delta]| > \tfrac{1}{2}\mathbb{E}[Z'_\Delta]\right] \leq \frac{4\mathrm{Var}[Z'_\Delta]}{\mathbb{E}[Z'_\Delta]^2} \leq \frac{4}{\mathbb{E}[Z'_\Delta]} \to 0\,.$$

Hence $\mathbb{P}[Z'_\Delta \geq \frac{1}{2}\mathbb{E}[Z'_\Delta]] \to 1$ follows. Finally we note that $\Delta = \alpha_n\sqrt{Np_nq_n} \leq \alpha_n n^{3/2}$, so

$$\frac{\frac{1}{2}\mathbb{E}[Z'_\Delta]}{\alpha_n^{-1}\sqrt{\Delta}} \geq \frac{\alpha_n^4 n/6}{\alpha_n^{-1/2} n^{3/4}} = \tfrac{1}{6}(\alpha_n n^{1/18})^{9/2} \to \infty.$$

Thus $\frac{1}{2}\mathbb{E}[Z'_\Delta] \geq \alpha_n^{-1}\sqrt{\Delta}$ for sufficiently large n, and so the result follows. □

We now need to address the issue that the degrees of the vertices contributing to Z'_Δ are not uniform in $(Np_n - 2\Delta, Np_n - \Delta]$.

First we state and solve an elementary matrix optimization problem. For a $t \times s$ matrix M, the product of $q = \min\{t, s\}$ elements no two of which are in the same row or in the same column will be called a *permanent term*. The *full permanent* of M is the total sum $F(M)$ of all permanent terms.

A matrix (x_{ij}) is called *row monotonic* if $x_{ij} \leq x_{ik}$, for every i and $j < k$. Let $a_i \geq 0$ be given real numbers, $i = 1, \ldots, t$, and let $\mathcal{M}$ be the family of all $t \times s$ row monotonic matrices (x_{ij}) with entries $x_{ij} \geq 0$ and with fixed row sums $x_{i1} + x_{i2} + \cdots + x_{is} = a_i$, $i = 1, \ldots, t$.

Lemma 10. *The full permanent function* F *attains its maximum on* $\mathcal{M}$ *at the matrix with constant rows.*

Proof. For $i = 1,\dots,t$, let $\overline{a}_i = a_i/s$ be the average of the elements in row i. Let $\beta(M)$ be the number of elements of $M \in \mathcal{M}$ that differ from the average of their row. We show that the matrix $M_0 \in \mathcal{M}$ with $\beta(M_0) = 0$ maximizes F.

Let $M \in \mathcal{M}$ be optimal, and assume that row ℓ of M is not constant. Let $x_{\ell j} < \overline{a}_\ell < x_{\ell k}$ such that j is largest possible and k is smallest possible. Set $\delta = \min\{\overline{a}_\ell - x_{\ell j}, x_{\ell k} - \overline{a}_\ell\}$. Let M' be the matrix obtained from M by replacing $x_{\ell j}$ and $x_{\ell k}$ with $x'_{\ell j} = x_{\ell j} + \delta$ and $x'_{\ell k} = x_{\ell k} - \delta$, respectively. Observe that $M' \in \mathcal{M}$.

When computing the difference $F(M') - F(M)$ all those permanent terms cancel that contain none of $x_{\ell j}$, $x_{\ell k}$, $x'_{\ell j}$, and $x'_{\ell k}$. Assume that a permanent term T in $F(M')$ contains just one factor y from columns j and k, let $y = x'_{\ell j}$ or $x'_{\ell k}$, and set $T' = T/y$. Obviously

$$(x'_{\ell j} + x'_{\ell k} - x_{\ell j} - x_{\ell k})T' = 0 \ ,$$

hence in the difference $F(M') - F(M)$ all those permanent terms cancel that contain just one factor from columns j and k.

Let $M(\ell, i; j, k)$ be the $(t-2)\times(s-2)$ matrix that is obtained from M by removing rows ℓ, i, and columns j, k. Now the remaining permanent terms can be grouped as follows:

$$\begin{aligned} F(M') - F(M) &= \sum_{i\neq\ell} (x'_{\ell j}x_{ik} + x'_{\ell k}x_{ij} - x_{\ell j}x_{ik} - x_{\ell k}x_{ij})F(M(\ell, i; j, k)) \\ &= \delta \sum_{i\neq\ell} (x_{ik} - x_{ij})F(M(\ell, i; j, k)) \ . \end{aligned}$$

Because M is a non-negative row monotonic matrix, $F(M') - F(M) \geq 0$. Therefore, $M' \in \mathcal{M}$ is optimal as well. Obviously $\beta(M') \leq \beta(M) - 1$, because $x'_{\ell j} = \overline{a}_\ell$ or $x'_{\ell k} = \overline{a}_\ell$, or both.

Repeating the procedure above, by taking non-constant rows in any order, we obtain an optimal matrix M_0 with $\beta(M_0) = 0$ in at most $\beta(M)$ steps. This matrix M_0 has obviously constant rows. $\square$

We will say that an integer-valued random variable $Y \in [1, s]$ has *monotonic distribution* if the sequence $(\mathbb{P}[Y = k])_{k=1}^{s}$ is non-decreasing.

Lemma 11. *For $t \leq s$, let $Y_1, Y_2, \dots, Y_t \in [1, s]$ be independent random variables with monotonic distributions. Let $W_1, W_2, \dots, W_t \in [1, s]$ be independent random variables with the same uniform distribution. Then*

$$\mathbb{P}[Y_i = Y_\ell \ \textit{for some}\ 1 \leq i < \ell \leq t] \geq \mathbb{P}[W_i = W_\ell \ \textit{for some}\ 1 \leq i < \ell \leq t] \ .$$

Proof. We have

$$\mathbb{P}[W_i = W_\ell \text{ for some } 1 \le i < \ell \le t] = 1 - \sum_{1 \le i_1, \dots, i_t \le s} \mathbb{P}[(W_1 = i_1) \wedge \dots \wedge (W_t = i_t)]$$

and

$$\mathbb{P}[Y_i = Y_\ell \text{ for some } 1 \le i < \ell \le t] = 1 - \sum_{1 \le i_1, \dots, i_t \le s} \mathbb{P}[(Y_1 = i_1) \wedge \dots \wedge (Y_t = i_t)] \,,$$

where the summations are taken over all distinct integers $i_1, \dots, i_t \in [1, s]$. For every $1 \le i \le t$, $1 \le j \le s$, let $x_{ij} = \mathbb{P}[Y_i = j]$, and define the $t \times s$ matrix $M = (x_{ij})$. By the independence,

$$\sum_{1 \le i_1, \dots, i_t \le s} \mathbb{P}[(Y_1 = i_1) \wedge \dots \wedge (Y_t = i_t)] = \sum_{1 \le i_1, \dots, i_t \le s} \mathbb{P}[Y_1 = i_1] \dots \mathbb{P}[Y_t = i_t] = F(M).$$

Notice that M is a row monotonic matrix with row sums equal to one. Then by applying Lemma 10, we obtain that $F(M) \le F(M_0)$ where M_0 is the $t \times s$ matrix with constant rows $(1/s, 1/s, \dots, 1/s)$ given by the distributions of the random variables W_i. Therefore

$$\begin{aligned} \mathbb{P}[Y_i = Y_\ell \text{ for some } 1 \le i < \ell \le t] &= 1 - F(M) \\ &\ge 1 - F(M_0) = \mathbb{P}[W_i = W_\ell \text{ for some } 1 \le i < \ell \le t] \end{aligned}$$

follows. □

Theorem 12. *Assume $p_n, q_n \ge cn^{-1}$ for some $c > 0$. Then the probability that a random 4-uniform hypergraph $H_4(n, p_n)$ has two vertices of the same degree tends to one as $n \to \infty$.*

Proof. We adopt the notation of Lemma 9 by selecting $\alpha_n \to 0$ sufficiently slowly so that $\alpha_n n^{1/18} \to \infty$, setting $\Delta = \alpha_n \sqrt{N p_n q_n}$, and choosing S so that $|S| \sim \alpha_n^3 n$. By Lemma 7, we may condition on H_2 and assume $|S'| > |S|/2$.

Since $p_n, q_n \ge cn^{-1}$, $\Delta = \alpha_n \sqrt{N p_n q_n} \ge c' \alpha_n n^{1/2}$ for some $c' > 0$, so $\Delta \to \infty$. Thus by Lemma 9 it follows that, with probability tending to one, the random 4-uniform hypergraph $H = H_4(n, p_n)$ has a set T of at least $t := \alpha_n^{-1} \sqrt{\Delta}$ vertices in S' with degree values in $(N p_n - 2\Delta, N p_n - \Delta]$. Moreover, conditioned on T and H_2, the degrees of the vertices in T are independent, and are given by $d_2(v) = N_2 p_n + \lambda_v$ plus

$d_1(v)$, where $d_1(v)$ is a Bernoulli $B(N_1, p_n)$ random variable conditioned so that it lies in the interval

$$(N_1 p_n - (2\Delta + \lambda_v), N_1 p_n - (\Delta + \lambda_v)].$$

As $v \in S'$, $|\lambda_v| \leq \Delta$, and so this interval lies entirely below the mean $N_1 p_n$ of the Binomial. Thus the distribution of $d_1(v)$, and hence $d(v)$, is monotonic. The result follows from Lemmas 11 and 6. □

Theorem 13. *The probability that the random 4-uniform hypergraph of order n has two vertices of the same degree tends to one as $n \to \infty$.*

Proof. Assume that the conclusion is not true. By symmetry we may assume $p_n \in [0, \frac{1}{2}]$ for all n. Then there is a $\varepsilon > 0$ and a subsequence $n_1, n_2, \ldots$ such that the probability that there are two vertices of the same degree in $H_4(n_i, p_{n_i})$ is smaller than $1 - \varepsilon$ for every i. Now either there is a subsequence $p_{n_{i_j}}$ with $p_{n_{i_j}} = o(n_{i_j}^{-1})$ or there is a constant $c > 0$ and a subsequence $p_{n_{i_j}}$ with $p_{n_{i_j}} \geq c n_{i_j}^{-1}$. In the first case we obtain a contradiction from Theorem 4 with $r = 4$, and in the second case we obtain a contradiction from Theorem 12. □

References

[1] P. Balister, B. Bollobás, J. Lehel, and M. Morayne, Hypergraph irregularity, (in preparation).

[2] B. Bollobás: Vertices of given degree in a random graph, Journal of Graph Theory, 6 (1982) 147–155.

[3] B. Bollobás: *Random Graphs*, Academic Press Inc. (London) Ltd. 1985.

[4] W. Feller: *An Introduction to Probability Theory and Its Applications*, Vol.2, John Wiley and Sons, Inc., 1971, New York, London, Sydney.

[5] R. J. Faudree, R. H. Schelp, M. S. Jacobson, and J. Lehel, Irregular networks, regular graphs and integer matrices with distinct row and column sums, Discrete Math., 76 (1989) 223–240.

[6] A. Gyárfás, M. S. Jacobson, L. Kinch, J. Lehel, and R. H. Schelp, Irregularity strength of uniform hypergraphs, J. Combin. Math. Combin. Comput., 11 (1992) 161–172.

[7] `http://en.wikipedia.org/wiki/Birthday_paradox`