Intersections of random hypergraphs and tournaments

Béla Bollobás [*] Alex Scott [†]

Abstract

Given two random hypergraphs, or two random tournaments of order n, how much (or little) can we make them overlap by placing them on the same vertex set? We give asymptotic answers to this question.

1 Introduction

Let G and H be two random hypergraphs or two random tournaments of order n. If we place G and H on the same vertex set, how much can we make the two graphs of tournaments agree (or disagree)? The aim of this paper is to give asymptotic answers to this question.

1.1 Tournaments

Let T and T' be two tournaments of order n. If we place T, T' randomly on the same vertex set then the expected number of common edges (i.e. edges with the same orientation) is $(1/2)\binom{n}{2}$. The *positive discrepancy* $\mathrm{disc}^+(T, T')$ measures how much more we can get the two tournaments to agree, and the

[*]Department of Pure Mathematics and Mathematical Statistics, Wilberforce Road, Cambridge CB3 0WB, UK; *and* Department of Mathematical Sciences, University of Memphis, Memphis TN38152, USA; *and* London Institute for Mathematical Sciences, 35a South St, Mayfair, London W1K 2XF, UK; email: bb12@cam.ac.uk. Research supported in part by MULTIPLEX no. 317532.

[†]Mathematical Institute, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, UK; email: scott@maths.ox.ac.uk.

negative discrepancy $\mathrm{disc}^-(T,T')$ measures how much more we can get them to disagree. Formally, we define:

$$\mathrm{disc}^+(T,T') := \max_\phi |E(\phi(T)) \cap E(T')| - \frac{1}{2}\binom{n}{2},$$
$$\mathrm{disc}^-(T,T') := \frac{1}{2}\binom{n}{2} - \min_\phi |E(\phi(T)) \cap E(T')|,$$

where the maximum is taken over all bijections ϕ from the vertex set of T to the vertex set of T'. We also define the (unsigned) *discrepancy* $\mathrm{disc}(T,T') = \max\{\mathrm{disc}^+(T,T'), \mathrm{disc}^-(T,T')\}$.

The *transitive tournament* TT_n *of order* n is the tournament with vertex set $[n]$ and directed edges $\{ij : i < j\}$. Note that positive and negative discrepancy are the same when one tournament is transitive: $\mathrm{disc}^+(T,TT_n) = \mathrm{disc}^-(T,TT_n)$, as we can reverse all edges of TT_n by reversing the order of the vertices. The *random tournament of order* n is the tournament with vertex set $[n]$, where independently for each pair $\{i,j\}$ the tournament contains the edges ij or ji with probability $1/2$ each.

The minimal value of $\mathrm{disc}^+(T,TT_n)$ has been extensively studied. Let

$$f(n) = \min_{|T|=n} \mathrm{disc}^+(T,TT_n),$$

where the minimum is taken over all tournaments of order n. After being studied by several authors (see Erdős and Moon [11], Reid [18] and Jung [14]), the order of magnitude of $f(n)$ was determined by Spencer ([20], [21]; see also Fernandez de la Vega [13]), who showed that

$$f(n) = \Theta(n^{3/2}).$$

In fact, Spencer showed that with high probability a random tournament T satisfies

$$\mathrm{disc}^+(T,TT_n) = \Theta(n^{3/2}).$$

Here, we will consider the discrepancy $\mathrm{disc}(T,T')$ when *both* tournaments T and T' are random. We will show in Section 3 that for a pair of random tournaments the discrepancy is much larger than $\Theta(n^{3/2})$: in fact, with exponentially small failure probability, we have

$$\mathrm{disc}(T,T') = \Theta(n^{3/2}\sqrt{\log n}).$$

We note that the discrepancy of tournaments has been considered by a number of authors (see, for instance, Rödl and Spencer [17], Berger and Shor [1] and Czygrinow Poljak and Rödl [9] for algorithmic results). Discrepancy has also been studied in social choice theory, and is equivalent to determining the Slater index (see Slater [19], Bermond [2], Laslier [15], Charon and Hudry [8]).

1.2 Hypergraphs

Let G and H be two k-uniform hypergraphs of order n, with densities p and q respectively. If we place G and H randomly on the same vertex set then the expected number of common edges is $pq\binom{n}{k}$. The *positive discrepancy* $\mathrm{disc}^+(G,H)$ measures the extent to which we can get the two graphs to overlap, and the *negative discrepancy* $\mathrm{disc}^-(G,H)$ measures the extent to which we can get them to be disjoint. Formally, we define:

$$\mathrm{disc}^+(G,H) := \max_{\phi} |\phi(E(G)) \cap E(H)| - pq\binom{n}{k}$$
$$\mathrm{disc}^-(G,H) := pq\binom{n}{k} - \min_{\phi} |\phi(E(G)) \cap E(H)|,$$

where the maximum is taken over all bijections ϕ from $V(G)$ to $V(H)$. We also define the *discrepancy* $\mathrm{disc}(G,H) = \max\{\mathrm{disc}^+(G,H), \mathrm{disc}^-(G,H)\}$. We note that a related measure for the discrepancy of a *single* hypergraph was introduced by Erdős and Spencer [12], and further investigated by Erdős, Goldberg, Pach and Spencer [10] and Bollobás and Scott [5] (who introduced the signed versions of discrepancy).

The *random k-uniform hypergraph* $\mathcal{G}^{(k)}(n,p)$ is the k-uniform hypergraph with vertex set $[n]$, where each of the possible $\binom{n}{k}$ edges is present independently with probability p. It follows easily from the results of [5] (see also [6]) that if G is (for instance) a complete k-uniform hypergraph on $n/2$ vertices, together with $n/2$ isolated vertices, and H is a random k-uniform hypergraph with density p, then (for a large range of p) with high probability

$$\mathrm{disc}(G,H) = \Theta(\sqrt{p(1-p)}n^{(k+1)/2}). \tag{1}$$

In the case $k=2$, a bound of form $\Omega(n^{(k+1)/2}$ holds much more generally: it was shown in [6] that if G and H are graphs of order n, with densities p

and q respectively, then

$$\text{disc}^+(G,H)\text{disc}^-(G,H) \geq c(p,q)n^3. \tag{2}$$

In particular, if p and q are bounded away from 1 then $\text{disc}(G,H) = \Omega(n^{3/2})$.

In light of the results above, it seems plausible that a version of (2) should extend to k-uniform hypergraphs along the lines of (1) (indeed, we conjectured this in [6]). However we showed in [7] that such a straightforward extension does not hold even for k=3, as there is a pair of nontrivial 3-uniform hypergraphs G, H with $\text{disc}(G,H) = 0$. More generally, if we allow edge weights, then for every $k \geq 2$ we construct a collection of k nontrivial weighted hypergraphs such that every pair has discrepancy 0 (here, a hypergraph is nontrivial if its weight function is not constant). On the other hand, if we take on additional hypergraph we do get a version of (2): every set of $k+1$ nontrivial weighted hypergraphs has some pair that has discrepancy at least $\Omega(n^{(k+1)/2})$, up to normalization. Further discusssion, as well as more detailed positive results, can be found in [7].

In this paper, we will consider the discrepancy $\text{disc}(G,H)$ when *both* G and H are random k-uniform hypergraphs. We will show in Section 4, for a large range of p and q, that if $G \in \mathcal{G}^{(k)}(n,p)$ and $H \in \mathcal{G}^{(k)}(n,q)$ then with failure probability $\exp(-n^{1-\epsilon})$ both positive and negative discrepancies have order

$$\Theta(n^{(k+1)/2}\sqrt{p(1-p)q(1-q)\log n}).$$

Note that this beats (1) by a factor $\sqrt{\log n}$, similarly to the tournament case. In Section 4, we also investigate the behaviour of positive and negative discrepancies for p and q in the range where this does not happen.

2 Tools

We shall need some standard bounds on the binomial distribution. For simplicity we gather these together in this section.

We use standard Chernoff bounds: if X is a sum of Bernoulli random variables and $\mu = \mathbb{E}X$ then

$$\mathbb{P}[X > \mathbb{E}X + t] \leq \exp\left(\frac{-t^2}{2\mu + 2t/3}\right) \tag{3}$$

and

$$\mathbb{P}[X < \mathbb{E}X - t] \leq \exp\left(\frac{-t^2}{2\mu}\right) \tag{4}$$

It follows from (3) and (4) that

$$\mathbb{P}[|X - \mathbb{E}X| > t] \leq 2\exp(-\min\{t^2/4\mu, 3t/4\}). \tag{5}$$

We will use the elementary fact that if $X \sim B(n,p)$, where $np(1-p) = \Omega(1)$, then $\mathbb{P}[X = \lfloor np \rfloor] = \Omega(1/\sqrt{np(1-p)})$ uniformly in n and p. This follows immediately from the facts: (1) $\mathbb{P}[X = t]$ is maximized at $t = \lfloor np \rfloor$ or $t = \lceil np \rceil$; (2) $\mathbb{P}[|X - np| \leq 2\sqrt{np(1-p)}] \geq 3/4$, by Chebyshev's Inequality; and (3) if np is not an integer then, with $t = \lfloor np \rfloor$, $\mathbb{P}[X = t+1]/\mathbb{P}[X = t] = p(n-t)/(1-p)(t+1) = \Theta(1)$, as $np = \Omega(1)$.

The following version of De Moivre-Laplace (see [3]) will also be useful.

Lemma 1. *Suppose $p = p(n)$ and $h = h(n)$ satisfy $p(1-p)n \to \infty$ and $|h| = o((p(1-p)n)^{2/3})$. Suppose that $X \sim B(n,p)$. Then*

$$\mathbb{P}[X \geq np + h] \sim 1 - \Phi(h/\sqrt{p(1-p)n}).$$

In particular, if $x = x(n)$ is bounded away from 0 and ∞, and $p(1-p)n/(\log n)^3 \to \infty$, then

$$\mathbb{P}[X \geq np + x\sqrt{p(1-p)n\log n}] \sim \frac{1}{x\sqrt{2\pi\log n}} n^{-x^2/2}.$$

Proof. The first assertion is a version of de Moivre-Laplace given in [3]. For the second, we have $x\sqrt{p(1-p)n\log n} = o((p(1-p)n)^{2/3})$ and so

$$\mathbb{P}[X \geq \mathbb{E}X + x\sqrt{p(1-p)n\log n}] \sim 1 - \Phi(t) \sim e^{-t^2/2}/t\sqrt{2\pi},$$

where $t = x\sqrt{p(1-p)n\log n}/(\sqrt{p(1-p)n}) = x\sqrt{\log n}$. □

We also note the following simple bound.

Lemma 2. *Suppose $n \geq 1$ and $p = p(n)$ are such that $p(1-p)n \geq 10$, and suppose*

$$|K| \leq p(1-p)n/2.$$

Then, for $X \sim B(n,p)$,

$$\mathbb{P}[X = \lceil np + K \rceil] = \Omega(e^{-K^2/p(1-p)n}/\sqrt{np(1-p)}),$$

uniformly in n, p, K.

Proof. Suppose that $X \sim B(n,p)$ and $p(1-p)n \geq 10$. Then as noted above $\mathbb{P}[X = \lceil np \rceil] = \Omega(1/\sqrt{np(1-p)})$, uniformly in n, p. Now if $np+k$ is an integer then

$$\frac{\mathbb{P}[X = np+k+1]}{\mathbb{P}[X = np+k]} = \frac{p}{1-p}\frac{n-(np+k)}{np+k+1} = \frac{1-k/(1-p)n}{1+(k+1)/pn}.$$

If $0 \leq k \leq p(1-p)n/2$ then $k/(1-p)n \leq 1/2$ and so

$$\frac{1-k/(1-p)n}{1+(k+1)/pn} \geq e^{-2k/(1-p)n-(k+1)/pn} \geq e^{-2(k+1)/p(1-p)n},$$

where we have used the fact that $1-t \geq \exp(-2t)$ for $t \in [0, 3/4]$. Similarly, if $0 \geq k \geq -p(1-p)n/2 - 1$ then

$$\frac{1-k/(1-p)n}{1+(k+1)/pn} = \frac{1+|k|/(1-p)n}{1-(|k|-1)/pn} \leq e^{|k|/(1-p)n+2(|k|-1)/pn} \leq e^{2|k|/p(1-p)n},$$

It follows that, for $|K| \leq p(1-p)n/2$,

$$\begin{aligned}
\mathbb{P}[X = \lceil np+K \rceil] &\geq \mathbb{P}[X = \lceil np \rceil] \cdot \prod_{0 \leq k \leq |K|+1} e^{-2(k+1)/p(1-p)n} \\
&\geq \Omega(1/\sqrt{np(1-p)}) \cdot e^{-(K+3)^2/p(1-p)n} \\
&= \Omega(e^{-K^2/p(1-p)n}/\sqrt{np(1-p)}),
\end{aligned}$$

uniformly in n, p, K. □

Lemma 2 transfers straightforwardly to a bound on intersections of random subsets of $[n]$ with fixed size.

Lemma 3. *There is a constant c such that the following holds. Suppose $n \geq 1$ and $n_1, n_2 \leq n$, and define $p = n_1/n$, $q = n_2/n$, $\sigma = \sqrt{p(1-p)q(1-q)n}$. Let A and B be random subsets of $[n]$ chosen independently and uniformly at random, with $|A| = n_1$ and $|B| = n_2$. If $\sigma^2 \geq c$, and L is a real number such that $pqn + L$ is an integer and*

$$|L| \leq \sigma^2/5, \tag{6}$$

then

$$\mathbb{P}[|A \cap B| = pqn + L] = \Omega(e^{-3L^2/\sigma^2}/\sigma), \tag{7}$$

uniformly in p, q, n.

Proof. We will take c to be sufficiently large for our estimates below to hold. We may assume that A is fixed, and B is chosen at random. So the probability that $|A \cap B| = pqn + L$ is equal to

$$\binom{pn}{pqn+L}\binom{(1-p)n}{q(1-p)n-L}\Big/\binom{n}{qn}.$$

From Lemma 2 we know

$$q^{qpn+L}(1-q)^{(1-q)pn-L}\binom{pn}{pqn+L} = \Omega(e^{-L^2/q(1-q)pn}/\sqrt{pnq(1-q)})$$

and

$$\begin{aligned} q^{q(1-p)n-L}(1-q)^{(1-q)(1-p)n+L}&\binom{(1-p)n}{q(1-p)n-L} \\ &= \Omega(e^{-L^2/q(1-q)(1-p)n}/\sqrt{(1-p)nq(1-q)}). \end{aligned}$$

Taking the product of the previous two equations, and dividing through by $q^{qn}(1-q)^{(1-q)n}\binom{n}{qn} = \Theta(1/\sqrt{q(1-q)n})$, the result follows. $\square$

Note that, for suitable L and σ, the bound in Lemma 3 can be summed over an interval of length σ to obtain an inequality of form

$$\mathbb{P}[|A \cap B| \geq pqn + L] = \Omega(e^{-4L^2/\sigma^2}). \tag{8}$$

3 Random tournaments

The aim of this section is to prove that, for a pair of random tournaments T_1, T_2, we have with high probability

$$\mathrm{disc}^+(T_1, T_2) = \Theta(n^{3/2}\sqrt{\log n}).$$

It follows immediately that (with high probability) $\mathrm{disc}^-(T_1, T_2)$ has the same order of magnitude: if we let $\overline{T}_1$ be the tournament obtained from T_1 by reversing all edges, it is clear that T_1 and $\overline{T}_1$ have the same distribution, while $\mathrm{disc}^-(T_1, T_2) = \mathrm{disc}^+(\overline{T}_1, T_2)$.

Theorem 4. *For every $\epsilon > 0$ there are constants $\alpha, \beta > 0$ such that the following holds. Let T_1, T_2 be random tournaments of order n. Then, with failure probability* $\exp(-\Omega(n^{1-\epsilon}))$,

$$\alpha n^{3/2}\sqrt{\log n} \leq \mathrm{disc}^+(T_1, T_2) \leq \beta n^{3/2}\sqrt{\log n}.$$

Proof. The upper bound is straightforward. For any bijection $\phi : V(T_1) \to V(T_2)$, the number of common edges is distributed as $B(\binom{n}{2}, 1/2)$. By (3), for fixed $\beta > 0$, the probability that the number of common edges exceeds its expectation by $\beta n^{3/2}\sqrt{\log n}$ is at most $\exp(-\beta^2 n^3 \log n/(2+o(1))\frac{1}{2}\binom{n}{2}) < \exp(-(2+o(1))\beta^2 n \log n) = o(e^{-n}/n!)$, provided $\beta > 1/\sqrt{2}$. Since there are $n!$ possible mappings ϕ, we then have $\mathrm{disc}^+(T_1, T_2) < \beta n^{3/2}\sqrt{\log n}$ with failure probability $\exp(-\Omega(n))$.

For the lower bound, we will construct a bijection $\phi : V(T_1) \to V(T_2)$ in three rounds.

We begin by setting $V = V(T_1) = \{v_1, \ldots, v_n\}$ and $W = V(T_2) = \{w_1, \ldots, w_n\}$. We set $r = \lfloor n/2 \rfloor$ and write $V_0 = \{v_1, \ldots, v_r\}$, $V_1 = V \setminus V_0$, $W_0 = \{w_1, \ldots, w_r\}$, $W_1 = W \setminus W_0$. We also define the induced tournaments $T_1' = T_1[V_0]$ and $T_2' = T_2[W_0]$. We write $\Gamma_1^{\pm}(\cdot)$ and $\Gamma_2^{\pm}(\cdot)$ for the in-neighbourhood/out-neighbourhood of vertices in T_1 and T_2 respectively.

In the first round, we take an arbitrary bijection between V_0 and W_0: define $\phi : V_0 \to W_0$ by setting $\phi(v_i) = w_i$ for $i = 1, \ldots, r$. Let $X_1 = |\phi(E(T_1')) \cap E(T_2')|$ be the number of edges on which the two orientations (of edges in W_0) agree. Then $X_1 \sim B(\binom{r}{2}, 1/2)$, and so by (4) we have

$$X_1 \geq \frac{1}{2}\binom{r}{2} - n^{3/2}, \tag{9}$$

with failure probability $e^{-\Omega(n)}$.

Let $V_2 \subset V_1$ be an arbitrary set of $s = \lfloor n/6 \rfloor$ vertices. In the second round, we construct an injection $\phi : V_2 \to W_1$, so that we gain significantly more than the expected number of common edges in the bipartite digraph between $\phi(V_2)$ and W_0 (we do not examine the edges inside $\phi(V_2)$ at this point).

For each $v \in V_2$ and $w \in W_1$, we let X_{vw} be the number of edges between v and V_0 that would have the same orientation as their image if we mapped v to w:

$$X_{vw} = |\phi(\Gamma_1^+(v) \cap V_0) \cap \Gamma_2^+(w) \cap W_0| + |\phi(\Gamma_1^-(v) \cap V_0) \cap \Gamma_2^-(w) \cap W_0|.$$

Then $X_{vw} \sim B(r, 1/2)$, so by Lemma 1 we can pick a constant $\eta > 0$ such that, for all sufficiently large n,

$$\mathbb{P}[X_{vw} \geq n/4 + \eta\sqrt{n \log n}] = n^{-\alpha}$$

for some $\alpha = \alpha(\eta, n) < \epsilon/2$.

We define a bipartite graph B with vertex classes V_2 and W_1, with $v \in V_2$ joined to $w \in W_1$ if $X_{vw} \geq n/4 + \eta\sqrt{n \log n}$. We shall show that with high probability this graph contains a matching, and then use this to construct our mapping from V_2 to W_1. Note that the edges of B are not independent. However, for each $v \in V_2$, the random variables $\{X_{vw} : w \in W_1\}$ are independent, and for each $w \in W_1$, the random variables $\{X_{vw} : v \in V_2\}$ are independent; this will be enough for us to bound the degrees of vertices in B, which will in turn be enough to prove the existence of the required matching.

For fixed $v \in V_2$, let $N_v = d_B(v) = |\{w \in W_1 : X_{vw} \geq n/4 + \eta\sqrt{n \log n}\}|$. The random variables $\{X_{vw} : w \in W_1\}$ are independent, so $N_v \sim B(n - r, n^{-\alpha})$. Since $\mathbb{E}N_v = n^{-\alpha}(n-r) \sim n^{1-\alpha}/2$, it follows from (5) that $\mathbb{P}[N_v < n^{1-\alpha}/3] < \exp(-\Omega(n^{1-\alpha}))$. Thus, with failure probability $\exp(-\Omega(n^{1-\alpha}))$ we have

$$N_v \geq n^{1-\alpha}/3 \tag{10}$$

for all $v \in V_0$.

Similarly, for $w \in W_1$, we let $N_w = |\{v \in V_2 : X_{vw} \geq n/4 + \eta\sqrt{n \log n}\}|$. The random variables $\{X_{vw} : v \in V_2\}$ are independent, so $N_w \sim B(s, n^{-\alpha})$. Now $\mathbb{E}N_w = n^{-\alpha}s \sim n^{1-\alpha}/6$, so by (5), $\mathbb{P}[N_w > n^{1-\alpha}/3] < \exp(-\Omega(n^{1-\alpha}))$. Thus, with failure probability $\exp(-\Omega(n^{1-\alpha}))$ we have

$$N_w \leq n^{1-\alpha}/3 \tag{11}$$

for all $w \in W_1$.

If (10) and (11) hold, then every vertex in V_2 has degree at least $n^{1-\alpha}/3$ in B, while every vertex in W_1 has degree at most $n^{1-\alpha}/3$ in B. It follows that $|\Gamma_B(S)| \geq |S|$ for every subset S of V_2, and so by Hall's Theorem there is a matching M in B from V_2 to W_1. Let us define $\phi : V_2 \to W_1$ by mapping each vertex of V_2 to its partner in M. The number of edges between W_0 and $\phi(V_2)$ that are oriented in the same direction in both T_2 and the image of T_1 is then at least

$$(n/4 + \eta\sqrt{n \log n})s = \frac{1}{2}rs + \Omega(n^{3/2}\sqrt{\log n}). \tag{12}$$

Finally, in the third round, we extend ϕ to a bijection between V and W by choosing a random bijection between the remaining vertices in each set. Let X_2 be the number of edges in common between $\phi(T_1)$ and T_2 that lie either inside W_1 or between $W_1 \setminus \phi(V_2)$ and W_0. Then X_2 is binomial with mean $\frac{1}{2}(\binom{n-r}{2} + r(n-r-s))$, and so by (4) we have

$$X_2 \geq \frac{1}{2}\binom{n-r}{2} + \frac{1}{2}r(n-r-s) - n^{3/2} \tag{13}$$

with failure probability $e^{-\Omega(n)}$ (note that we have not looked at these edges in T_1 before this step of the argument).

Finally, we note that with failure probability $\exp(-\Omega(n^{1-\alpha}))$ all of (9), (12) and (13) hold. Summing these, we see that the number of common edges between $\phi(T_1)$ and T_2 is at least $\frac{1}{2}\binom{n}{2} + \Omega(n^{3/2}\sqrt{\log n})$. □

4 Random hypergraphs

We now turn to the discrepancy of pairs of random hypergraphs.

We note first that there are trivial upper bounds on the positive and negative discrepancies. Let G_1, G_2 be k-uniform hypergraphs of order n with densities p and q respectively. The maximum possible positive discrepancy over such pairs occurs when we can nest one inside the other, so that G_1 and G_2 have $\min\{p,q\}\binom{n}{k}$ common edges. Subtracting the expected intersection of $pq\binom{n}{k}$, we get

$$\mathrm{disc}^+(G_1, G_2) \leq \min\{p(1-q)\binom{n}{k}, (1-p)q\binom{n}{k}\}. \tag{14}$$

Similarly, the maximum negative discrepancy occurs when $\overline{G}_1$ and G_2 have maximum possible overlap, and so are nested; in this case $\overline{G}_1$ and G_2 share $\min\{q, 1-p\}\binom{n}{k}$ edges and so G_1 and G_2 share $q\binom{n}{k} - \min\{q, 1-p\}\binom{n}{k}$ edges. Subtracting this from $pq\binom{n}{k}$ gives

$$\mathrm{disc}^-(G_1, G_2) \leq \min\{(1-p)(1-q)\binom{n}{k}, pq\binom{n}{k}\}. \tag{15}$$

We can also deduce (15) from (14), as replacing one of the hypergraphs G_1, G_2 by its complement exchanges $\mathrm{disc}^+(G_1, G_2)$ and $\mathrm{disc}^-(G_1, G_2)$.

Our aim here is to show that, if G_1 and G_2 are random hypergraphs with densities p and q, we get a similar phenomenon to the tournament case. In particular, we will first show that for a wide range of densities the positive and negative discrepancies are both (with high probability) of order

$$\Theta(n^{(k+1)/2}\sqrt{p(1-p)q(1-q)\log n}). \tag{16}$$

We will then turn, in the final part of this section, to the sparse case, where the behaviour is rather different.

We will need a little notation. For a k-uniform hypergraph G, a vertex $v \in V(G)$ and a set $S \subset V(G)$ we define

$$\Gamma(v, S) = \{T \subset S : |T| = k-1, T \cup \{v\} \in E(G)\}.$$

Note that if G is a graph then $\Gamma(v, S) = \Gamma(v) \cap S$; more generally, if G is a k-uniform hypergraph then $\Gamma(v, S)$ is the edge set of a $(k-1)$-uniform hypergraph on S.

For sets S and T of vertices in a k-uniform hypergraph G, we also define $e_{(i,k-i)}(S, T)$ to be the number of edges that have i vertices in S and $k-i$ vertices in T.

4.1 Dense hypergraphs

The trivial bounds (14) and (15) imply that

$$\begin{aligned}
&\min\{\mathrm{disc}^+(G_1, G_2), \mathrm{disc}^-(G_1, G_2)\} \\
&\qquad\qquad \le \binom{n}{k} \min\{p(1-q), (1-p)q, pq, (1-p)(1-q)\} \\
&\qquad\qquad = \binom{n}{k} \min\{p, 1-p\} \cdot \min\{q, 1-q\} \\
&\qquad\qquad = O(p(1-p)q(1-q)n^k).
\end{aligned}$$

If positive and negative discrepancies both behave as in (16), we must have

$$n^{(k+1)/2}\sqrt{p(1-p)q(1-q)\log n} = O(p(1-p)q(1-q)n^k),$$

and so $p(1-p)q(1-q) = \Omega(\log n/n^{k-1})$. We will show that if p and q satisfy this constraint then, with high probability, the positive and negative discrepancies do indeed both have order $\Theta(n^{(k+1)/2}\sqrt{p(1-p)q(1-q)\log n})$.

Theorem 5. *Fix $k \ge 2$ and $\epsilon > 0$. Let $p = p(n)$ and $q = q(n)$ satisfy $p, q \in (0, 1)$ and $p(1-p)q(1-q) = \Omega(\log n/n^{k-1})$. Let $G_1 \in \mathcal{G}^{(k)}(n, p)$ and $G_2 \in \mathcal{G}^{(k)}(n, q)$ be random k-uniform hypergraphs. Then, with failure probability* $\exp(-\Omega(n^{1-\epsilon}))$,

$$\mathrm{disc}^+(G_1, G_2) = \Theta(n^{(k+1)/2}\sqrt{p(1-p)q(1-q)\log n}) \tag{17}$$

and

$$\mathrm{disc}^-(G_1, G_2) = \Theta(n^{(k+1)/2}\sqrt{p(1-p)q(1-q)\log n}). \tag{18}$$

Proof. The proof will follow a similar strategy to that of Theorem 4; however, there are some additional complications.

We may assume that $p, q \leq 1/2$, or else replace one or both of G_1, G_2 by its complement (recall that replacing one of the graphs by its complement exchanges positive and negative discrepancies). We may also assume $p \leq q$, or else exchange G_1 and G_2. So $q(n) = \Omega(\sqrt{\log n/n^{k-1}})$ and $p(n) = \Omega(\log n/n^{k-1})$.

The upper bounds in (17) and (18) are straightforward. For a fixed bijection $\phi : V(G_1) \to V(G_2)$, the number X_ϕ of common edges is distributed as $B(\binom{n}{k}, pq)$. Let $\alpha \geq 1$ and set $t = \alpha n^{(k+1)/2}\sqrt{pq \log n}$ and $\mu = pq\binom{n}{k} \leq pqn^k$. We use (5): since $t^2/\mu = \Omega(\alpha^2 n \log n)$ and $t = \alpha\sqrt{pqn^{k+1}\log n} = \Omega(\alpha n \log n)$, we have $\mathbb{P}[|X_\phi - \mathbb{E}X_\phi| > t] = n^{-\Omega(\alpha n)}$. There are $n!$ choices for ϕ, so for sufficiently large α it follows that, with failure probability $\exp(-\Omega(n))$,

$$
\begin{aligned}
&\max\{\mathrm{disc}^+(G_1, G_2), \mathrm{disc}^-(G_1, G_2)\} \\
&\qquad \leq \mathrm{disc}^+(G_1, G_2) + \mathrm{disc}^-(G_1, G_2) \\
&\qquad = \max_\phi |\phi(E(G_1)) \cap E(G_2)| - \min_\phi |\phi(E(G_1)) \cap E(G_2)| \\
&\qquad = O(n^{(k+1)/2}\sqrt{pq \log n}).
\end{aligned}
$$

As $p, q \leq 1/2$, this is $O(n^{(k+1)/2}\sqrt{p(1-p)q(1-q)\log n})$.

For the lower bounds, we will as before construct a bijection $\phi : V(G_1) \to V(G_2)$ in three rounds. We will prove (17), and then note that (18) follows with straightforward changes to the argument.

We begin as before by setting $V = V(G_1) = \{v_1, \ldots, v_n\}$ and $W = V(G_2) = \{w_1, \ldots, w_n\}$. We set $r = \lfloor n/2 \rfloor$ and write $V_0 = \{v_1, \ldots, v_r\}$, $V_1 = V \setminus V_0$, $W_0 = \{w_1, \ldots, w_r\}$, $W_1 = W \setminus W_0$. We write $\Gamma_1(\cdot, \cdot)$ and $\Gamma_2(\cdot, \cdot)$ for neighbourhoods in G_1 and G_2 respectively.

For convenience, we will refer to edges in G_1 that have i vertices in V_1 and edges in G_2 that have i vertices in W_1 as *i-crossedges* (so for instance edges inside V_0 or W_0 are 0-crossedges).

In the first round, we define $\phi : V_0 \to W_0$ by setting $\phi(v_i) = w_i$ for $i = 1, \ldots, r$. The number of common edges in W_0 (i.e. $|\phi(E(G_1')) \cap E(G_2')|$) has distribution $B(\binom{r}{k}, pq)$, which has expectation $\mu = \Theta(pqn^k)$. For $\alpha > 0$, and $t = \alpha n^{(k+1)/2}\sqrt{p(1-p)q(1-q)}$, we have $t^2/\mu = \Omega(\alpha^2 n)$ and $t = \alpha n\sqrt{pqn^{k-1}} = \Omega(\alpha n\sqrt{\log n})$, as $p, q \leq 1/2$ and $pqn^{k-1} = \Omega(\log n)$. So by (5), we have

$$pq\binom{r}{k} + O(n^{(k+1)/2}\sqrt{p(1-p)q(1-q)}) \tag{19}$$

common edges, with failure probability $\exp(-\Omega(n))$. Note that (19) depends only on the edges inside V_0 and W_0.

We now concentrate on 1-crossedges: we show that we can get many common crossedges, and examine other types of crossedge later. We choose a subset V_2 of V_1 and construct an injection $\phi : V_2 \to W_1$, so that we gain significantly more than the expected number of common 1-crossedges in the bipartite graph between $\phi(V_2)$ and W_0. However, we have to be a little careful here. As in the tournament case, it is natural to map v to w if the image of $\Gamma_1(v, V_0)$ has a large overlap with $\Gamma_2(w, W_0)$. But this could happen because we have picked vertices in W_1 that have many crossedges: the remaining vertices of W_1 will have fewer 1-crossedges (on average), and so we would expect to lose when we pair them with vertices from V_1. We must also be careful to preserve sufficient independence between edges, and to ensure that we can control the degree sequence in the bipartite graph B (of pairs (v, w) with large common neighbourhood) so as to guarantee Hall's condition.

We therefore proceed as follows (we will give an informal sketch, and then a formal algorithm). We start by choosing subsets $V_2 \subset V_1$ and $W_2 \subset W_1$, putting aside the remaining vertices to use later. We examine the 1-crossedges from V_2 and from W_2, and drop to subsets V_3 and W_3 such that $|W_3| \sim 2|V_3|$ and all vertices in V_3 and W_3 have roughly the expected number of 1-crossedges. We next adjust the neighbourhoods of vertices in V_3 and W_3 by randomly removing edges so that every vertex in V_3 has a neighbourhood of size exactly $\lfloor p\binom{r}{k-1}\rfloor$ in V_0 and every vertex in W_3 has a neighbourhood of size exactly $\lfloor q\binom{r}{k-1}\rfloor$ in W_0 (this is not essential for the algorithm, but simplifies the analysis). We then argue, as is the tournament case, that with high probability there is a matching from V_3 to W_3 such that every pair creates many additional common 1-crossedges. Finally we clean up: we pair off the unused vertices of V_2 and W_2 at random with vertices that were put aside earlier (we have not previously examined 1-crossedges from these), and pair off any leftover vertices at random. As we shall show, with high probability the gain in the matching step outweighs any loss from the unexamined 1-crossedges and the crossedges of other types.

More formally, let $R = \binom{r}{k-1}$. We choose a small constant $\eta > 0$, and apply the following algorithm.

1. Let $V_2 \subset V_1$ be an arbitrary set of $\lfloor n/8 \rfloor$ vertices, and let $W_2 \subset W_1$ be an arbitrary subset of $\lfloor n/4 \rfloor$ vertices.

2. Let $V_3 \subset V_2$ be a set of $\lfloor n/40 \rfloor$ vertices v such that $|\Gamma_1(v, V_0)| \in (pR, pR + \sqrt{p(1-p)R})$ and let $W_3 \subset W_2$ be a set of $\lfloor n/20 \rfloor$ vertices w such that $|\Gamma_2(w, W_0)| \in (qR, qR + \sqrt{q(1-q)R})$. [If these cannot be found, the algorithm fails.]

3. For each $v \in V_3$, choose uniformly at random a set $A_v \subset \Gamma_1(v, V_0)$ such that $|A_v| = \lfloor Rp \rfloor$. For each $w \in W_3$, choose uniformly at random a set $B_w \subset \Gamma_2(w, W_0)$ such that $|B_w| = \lfloor Rq \rfloor$.

4. Define a bipartite graph B with vertex classes V_3 and W_3 such that $v \in V_3$ is adjacent to $w \in W_3$ if
$$|\phi(A_v) \cap B_w| \geq pqR + \eta\sqrt{p(1-p)q(1-q)R\log n}. \tag{20}$$

5. Find a perfect matching M in B from V_3 to W_3, and use this to define ϕ on V_3. [If this is cannot be done, the algorithm fails.]

6. Extend the domain of ϕ to include the rest of V_2 by taking a random injection from $V_2 \setminus V_3$ to $W_1 \setminus W_2$.

7. Extend the range of ϕ to include the rest of W_2 by taking a random injection from $W_2 \setminus \phi(V_3)$ to $V_1 \setminus V_2$ (and let this be ϕ^{-1} on $W_2 \setminus W_3$).

8. Finally, extend the domain to the whole of V_1 by picking a random bijection from the remaining vertices of V_1 to the remaining vertices of W_1.

We will show that with high probability the algorithm succeeds, and gives a mapping demonstrating (17). Note that the algorithm can only fail at Step 2 and Step 5.

In Step 2, we know that $pR \to \infty$, so Lemma 1 implies that $|\Gamma_1(v, V_0)| \in (pR, pR + \sqrt{p(1-p)R})$ with asymptotic probability $\Phi(1) - \Phi(0) \approx 0.341$. Thus, provided n is sufficently large, we have $\mathbb{P}[|\Gamma(v) \cap V_0| \in (pR, pR + \sqrt{p(1-p)R})] > 1/3$, independently for each $v \in V_2$. The number of vertices in V_2 that satisfy this is therefore (stochastically) bounded below by a random variable with distribution $B(\lfloor n/8 \rfloor, 1/3)$. It follows by (5) that, with exponentially small failure probability, there are more than $n/40$ vertices available for V_3. A similar argument applies to W_3.

In Step 3, note that the collection of all sets A_v and B_w is independent, as is the collection of sets given by $\Gamma_1(v, V_0)$, $v \in V_1$, and $\Gamma_2(w, W_0)$, $w \in W_1$.

In Step 4, let E_{vw} be the event that the edge vw is in B. We bound $\mathbb{P}[E_{vw}]$ from below using Lemma 3: note that we are applying the lemma with parameters $\widetilde{p} = \lfloor Rp \rfloor / R = (1 + O(1/n^{k-1}))p$ and $\widetilde{q} = \lfloor Rq \rfloor / R = (1 + O(1/n^{k-1}))q$; and then $L \sim \eta\sqrt{\widetilde{p}(1-\widetilde{p})\widetilde{q}(1-\widetilde{q})R\log n} = \Theta(\eta\sqrt{pqn^{k-1}\log n})$. Now since $pqn^{k-1} = \Omega(\log n)$, we have $L = O(\eta pqn^{k-1})$, which satisfies (6) if η is sufficiently small; we also have $\sigma^2 \sim \widetilde{p}(1-\widetilde{p})\widetilde{q}(1-\widetilde{q})R$, so $L^2/\sigma^2 \sim \Theta(\eta^2 \log n)$. It follows from Lemma 3 as in (8) that, provided η is sufficiently small, $\mathbb{P}[E_{vw}] = \alpha$ for some $\alpha = \alpha(p, q, n) \geq n^{-\epsilon}$.

For each $v \in V_3$, the sets B_w, $w \in W_3$, are independent from each other and from A_v, and so the events $\{E_{vw} : w \in W_3\}$ are independent. Thus $d_B(v) \sim B(\lfloor n/20 \rfloor, \alpha)$, and so $\mathbb{E}d_B(v) \sim \alpha n/20 = \Omega(n^{1-\epsilon})$. It follows from (5) that $\mathbb{P}[d_B(v) < \alpha n/30] < \exp(-\Omega(n^{1-\epsilon}))$. Thus, with failure probability $\exp(-\Omega(n^{1-\epsilon}))$ we have

$$d_B(v) \geq \alpha n/30 \tag{21}$$

for all $v \in V_3$.

Similarly, for each $w \in W_3$, the events $\{E_{vw} : v \in V_3\}$ are independent, so $d_B(w) \sim B(\lfloor n/40 \rfloor, \alpha)$. By (5), we have $\mathbb{P}[d_B(w) > \alpha n/30] < \exp(-\Omega(n^{1-\epsilon}))$. Thus, with failure probability $\exp(-\Omega(n^{1-\epsilon}))$ we have

$$d_B(w) \leq \alpha n/30 \tag{22}$$

for all $w \in W_3$.

If (21) and (22) hold, then every vertex in V_3 has degree at least $\alpha n/30$ in B, while every vertex in W_3 has degree at most $\alpha n/30$. As before, it follows by Hall's Theorem there is a matching M in B from V_3 to W_3. We define $\phi : V_3 \to W_3$ by mapping each vertex of V_3 to its partner in M.

We have shown that the algorithm succeeds in running with high probability. We now examine the number of common edges between $\phi(G_1)$ and G_2. We have already controlled the number of edges inside W_0 by (19); we next consider edges between W_0 and W_1.

For each $v \in V_3$, it follows from (20) that we obtain at least

$$pqR + \eta\sqrt{p(1-p)q(1-q)R\log n}$$

common 1-crossedges (note that we may obtain more than $|\phi(A_v) \cap B_{\phi(v)}|$ edges, as we deleted some edges in Step 3). Summing over V_3, this gives a total of at least

$$pqR|V_3| + \Omega(n^{(k+1)/2}\sqrt{p(1-p)q(1-q)\log n}) \qquad (23)$$

common crossedges.

The next three steps are very similar. In Step 6, we take a random mapping from $V_2 \setminus V_3$ to $W_1 \setminus W_2$. We have already examined the 1-crossedges between V_3 and V_0: let

$$m_1 = e_1^{(1,k-1)}(V_2 \setminus V_3, V_0) = e_1^{(1,k-1)}(V_2, V_0) - e_1^{(1,k-1)}(V_3, V_0).$$

Now $e_1^{(1,k-1)}(V_2, V_0) \sim B(R|V_2|, p)$. Applying (5) with $\mu = pR|V_2| = \Theta(pn^k)$ and $t = \alpha n^{(k+1)/2}\sqrt{p(1-p)} = \Omega(n \log n)$, for a suitable constant α, we see that with failure probability $\exp(-\Omega(n))$ we have $e_1^{(1,k-1)}(V_2, V_0) = pR|V_2| + O(n^{(k+1)/2}\sqrt{p(1-p)})$. On the other hand, by our choice of V_3, we have $e_1^{(1,k-1)}(V_3, V_0) = pR|V_3| + O(|V_3|\sqrt{p(1-p)R}) = pR|V_3| + O(n^{(k+1)/2}\sqrt{p(1-p)})$. So with failure probability $\exp(-\Omega(n))$,

$$m_1 = pR(|V_2| - |V_3|) + O(n^{(k+1)/2}\sqrt{p(1-p)}) = \Theta(pn^k),$$

as $pn^{k-1} \to \infty$. Since we have not yet examined the 1-crossedges in G_2 between $\phi(V_1 \setminus V_3)$ and W_0, the number of common crossedges has distribution $B(m_1, q)$. By (5) again, with $\mu = m_1 q = \Theta(pqn^k)$ and $t = \alpha n^{(k+1)/2}\sqrt{pq} = \Omega(\log n)$, we see that with failure probability $\exp(-\Omega(n))$ the number of common crossedges is $qm_1 + O(n^{(k+1)/2}\sqrt{pq})$, which equals

$$pqR(|V_2| - |V_3|) + O(n^{(k+1)/2}\sqrt{p(1-p)q(1-q)}). \qquad (24)$$

Applying the same argument (with p and q reversed) in Step 7 to $W_2 \setminus \phi(V_3)$, with failure probability $\exp(-\Omega(n))$, the number of common 1-crossedges between W_0 and $W_2 \setminus \phi(V_3)$ is

$$pqR(|W_2| - |V_3|) + O(n^{(k+1)/2}\sqrt{p(1-p)q(1-q)}). \qquad (25)$$

Moving to Step 8, we argue as in (19): the number of common 1-crossedges we gain in this step has distribution $B(N, pq)$, where $N = R(|V_1| - |V_2| - |W_2| + |V_3|)$, and so by (5) is

$$pqR(|V_1| - |V_2| - |W_2| + |V_3|) + O(n^{(k+1)/2}\sqrt{p(1-p)q(1-q)}), \qquad (26)$$

with failure probability $\exp(-\Omega(n))$.

Adding (23), (24), (25) and (26) together, we get that the total number of common 1-crossedges is, with failure probability $\exp(-\Omega(n^{1-\epsilon}))$, at least

$$pqR|V_1| + \Omega(n^{(k+1)/2}\sqrt{p(1-p)q(1-q)\log n}). \tag{27}$$

Note that the event (27), and the preceding algorithm, depend on 1-crossedges.

Finally, we count the number of common edges inside $\phi(V_1) = W_1$ and the number of common i-crossedges for $i = 2, \ldots, k-1$. As we have not previously looked at these edges, the number of common edges has distribution $B(\binom{n}{k} - \binom{r}{k} - (n-r)\binom{r}{k-1}, pq)$, and so as in (19) is within

$$O(n^{(k+1)/2}\sqrt{p(1-p)q(1-q)}) \tag{28}$$

of its expectation, with failure probability $\exp(-\Omega(n))$.

Finally, we note that (19), (27) and (28) all hold with failure probability $\exp(-\Omega(n^{1-\epsilon}))$, and so with failure probability $\exp(-\Omega(n^{1-\epsilon}))$ the number of common edges between $\phi(G_1)$ and G_2 is at least

$$pq\binom{n}{k} + \Omega(n^{(k+1)/2}\sqrt{p(1-p)q(1-q)\log n}).$$

The argument for negative discrepancy is the same, except that we look for degrees in the intervals $(pR - \sqrt{p(1-p)R}, pR)$ and $(qR - \sqrt{q(1-q)R}, qR)$ in Step 2, choose supersets in Step 3, and adjust (20) to: $|\phi(A_v) \cap B_w| \leq pqR - \eta\sqrt{p(1-p)q(1-q)R\log n}$. □

4.2 Sparse hypergraphs

The bounds in Theorem 5 hold as long as $p(1-p)q(1-q) = \Omega(\log n/n^{k-1})$. As noted above, these bounds can no longer hold for very sparse or dense pairs of graphs: for instance, if $pq = o(\log n/n^{k-1})$, we expect $\mathrm{disc}^-(G_1, G_2) = O(pq\binom{n}{k}) = O(pqn^k) = o(n^{(k+1)/2}\sqrt{pq\log n})$, so the bound (18) on the negative discrepancy cannot hold. On the other hand, there is no such constraint on the positive discrepancy. In this section, we investigate this regime.

As usual, we may assume that $p, q \leq 1/2$ (as we can always complement either graph and exchange positive and negative discrepancies). Thus the negative discrepancy must be at most pqn^k, while the positive discrepancy can be much larger.

Theorem 6. *Fix $k \geq 2$. Suppose $p, q \leq 1/2$, $pn^k \to \infty$ and $qn^k \to \infty$. Suppose that $pqn^{k-1} = \log n/\beta$, where $\beta = \beta(n) \to \infty$. Then, with high probability, for $G_1 \in \mathcal{G}^{(k)}(n,p)$ and $G_2 \in \mathcal{G}^{(k)}(n,q)$, we have*

$$\mathrm{disc}^-(G_1, G_2) = \Theta(pqn^k) \tag{29}$$

and

$$\mathrm{disc}^+(G_1, G_2) = \Theta(\min\{pn^k, qn^k, n\log n/\log\beta\}). \tag{30}$$

Proof. We begin with the positive discrepancy. We first prove the lower bound. Define $K = K(n,p,q)$ by

$$K = \min\{pn^{k-1}, qn^{k-1}, \log n/\log\beta\}/(8 \cdot k!).$$

Thus our aim is to show that $\mathrm{disc}^+(G_1, G_2) = \Omega(Kn)$.

Note that, since $\min\{p,q\}n^k \to \infty$, it follows from (3) and (4) that with high probability $e(G_1) = (1+o(1))p\binom{n}{k}$ and $e(G_2) = (1+o(1))q\binom{n}{k}$. Thus if we write p^*, q^* for the density of G_1, G_2 respectively, we have with high probability $pq\binom{n}{k} - p^*q^*\binom{n}{k} = o(pqn^k) = o(Kn)$.

Suppose first that $\min\{p,q\} \leq 10/n^{k-1}$ and $\max\{p,q\} \leq 1/(2000 \cdot k!)$. With high probability there are matchings of size at least $\min\{p,q\}\binom{n}{k}/100$ in both G_1 and G_2 (this is easly shown by choosing the edges of G_1 one at a time, and taking a greedy matching). Picking a mapping $\phi : V(G_1) \to V(G_2)$ for which two such matchings coincide, we may ensure that G_1 and G_2 have at least $\min\{p,q\}\binom{n}{k}/100$ common edges. On the other hand, $e(G_1)e(G_2)/\binom{n}{k} = (1+o(1))pq\binom{n}{k} \leq (1+o(1))\min\{p,q\}\binom{n}{k}/200$, so we get $\mathrm{disc}^+(G_1, G_2) = \Omega(\min\{p,q\}n^k) = \Omega(Kn)$, as required.

Next, suppose that $\min\{p,q\} \geq 10/n^{k-1}$ and $\max\{p,q\} \leq 1/(2000 \cdot k!)$. We have $pqn^k = n\log n/\beta = o(n\log n/\log\beta)$ and $pq\binom{n}{k} \leq \min\{p,q\}n^k/(2000 \cdot k!)$; so, for sufficiently large n, $pq\binom{n}{k} \leq Kn/200$, and in order to show that $\mathrm{disc}^+(G_1, G_2) = \Omega(Kn)$ it is therefore enough to find a placement of G_1 and G_2 so that they have at least $Kn/100$ common edges.

Let $r = \lfloor n/2 \rfloor$ and $R = \binom{r}{k-1}$. We follow a slightly simplified version of the algorithm in the proof of Theorem 5. The first round is as before: we select the partitions $V(G_1) = V_0 \cup V_1$ and $V(G_2) = W_0 \cup W_1$, with $|V_0| = |W_0| = r$, and a random bijection $\phi : V_0 \to W_0$. In the second round, we follow as far as Step 5 of the algorithm in Theorem 5, with some adjustments as follows.

1. Let $V_3 \subset V_1$ be a set of $\lfloor n/20 \rfloor$ vertices v such that $|\Gamma_1(v, V_0)| \geq pR$ and let $W_3 \subset W_1$ be a set of $\lfloor n/10 \rfloor$ vertices w such that $|\Gamma_2(w, W_0)| \geq qR$. [If this is cannot be done, the algorithm fails.]

2. For each $v \in V_3$, choose uniformly at random a set $A_v \subset \Gamma_1(v, V_0)$ such that $|A_v| = \lfloor Rp \rfloor$. For each $w \in W_3$, choose uniformly at random a set $B_w \subset \Gamma_2(w, W_0)$ such that $|B_w| = \lfloor Rq \rfloor$.

3. Define a bipartite graph B^* with vertex classes V_3 and W_3 such that $v \in V_3$ is adjacent to $w \in W_3$ if
$$|\phi(A_v) \cap B_w| \geq K.$$

4. Find a perfect matching M in B^* from V_3 to W_3, and use this to define ϕ on V_3. [If this is cannot be done, the algorithm fails.]

5. Extend the domain of ϕ to include the rest of V by taking a random injection between $V \setminus V_3$ and $W \setminus W_3$.

If the algorithm succeeds then we have found a suitable placement.

It is easily seen that Step 1 fails with exponentially small probability (by Lemma 1, each vertex in V_1 or W_1 is available for V_3 or W_3 independently with probability at least 1/3). Using ϕ, we may identify $V_0 = W_0 = [r]$. Let us bound from below the probability that an edge vw is present in B. Let A be a fixed pR-set in $[r]^{(k-1)}$, and let B be a random qR-set (we shall omit floors and ceilings from now on). We select elements for B from $[r]^{(k-1)}$ one at a time, without replacement. We shall say initially that a choice is *successful* if it belongs to A; after we have had K successful choices, we say that each subsequent choice is successful with probability $p/2$ (regardless of whether it belongs to A). Thus we have $|A \cap B| \geq K$ if and only if we have K or more successful choices. Note that if we have had fewer than K successes, then a choice is successful with probability at least $(pR - K)/R \geq p/2$. So the number of successes stochastically dominates a binomial distribution with parameters qR and $p/2$. Since $\binom{qR}{K} \geq (qR)_K/K^K \geq (qR/2K)^K$, we have

$$\begin{aligned}
\mathbb{P}[|A \cap B| \geq K] &\geq \binom{qR}{K}(p/2)^K(1-p/2)^{qR-K} \\
&\geq \left(\frac{qR}{2K}\right)^K (p/2)^K e^{-pqR} \\
&\geq \left(\frac{pqn^{k-1}}{10K \cdot (k-1)!}\right)^K n^{-o(1)} \\
&= \left(\frac{\log n}{10\beta K \cdot (k-1)!}\right)^K n^{-o(1)} \\
&\geq \left(\frac{\log \beta}{10\beta}\right)^{\log n/8\log\beta} n^{-o(1)} \\
&\geq n^{-1/4},
\end{aligned}$$

for sufficiently large n and β. It follows, as in the proof of Theorem 5, that with high probability there is a matching in B from V_3 to W_3, as required.

Finally, suppose that $\max\{p,q\} \geq 1/(2000 \cdot k!)$, say $p \geq 1/(2000 \cdot k!)$. As $pqn^{k-1} = \log n/\beta$, we have $q = O(\log n/\beta n^{k-1})$ and hence $K = \Omega(qn^{k-1}/(8 \cdot k!))$. We therefore want to show that $\mathrm{disc}^+(G_1, G_2) = \Omega(qn^k)$.

Let H_1 be a random subgraph of G_1 where we keep each edge with probability $1/(2000 \cdot k!)$. Then, by the arguments above, with high probability we have $\mathrm{disc}^+(H_1, G_2) = \Omega(Kn)$. Choose a placement of H_1 and G_2 onto the same vertex set such that this discrepancy is achieved. We now add back the other edges of G_1: the expected intersection is now $pq\binom{n}{k} + \Omega(qn^k)$; but as $qn^k \to \infty$, it follows from (5) that with high probability the same holds for $\mathrm{disc}^+(G_1, G_2)$.

We have proved the lower bound. We now turn to the upper bound on $\mathrm{disc}^+(G_1, G_2)$. As noted already, with high probability we have $e(G_1) = \Theta(pn^k)$ and $e(G_2) = \Theta(qn^k)$, and so (14) implies that $\mathrm{disc}^+(G_1, G_2) = O(\min\{pn^k, qn^k\})$. So we only need to show that $\mathrm{disc}^+(G_1, G_2) = O(n\log n/\log\beta)$. Let $N = \binom{n}{k}$ and $L = 4n\log n/\log\beta$. The probability that (for a fixed placement) G_1 and G_2 have at least L common edges is at most

$$\binom{N}{L}(pq)^L \leq (eNpq/L)^L \leq \left(\frac{en^{k-1}pq\log\beta}{2\log n}\right)^L = \left(\frac{e\log\beta}{2\beta}\right)^L,$$

which is at most $\beta^{-L/2} = e^{-2n\log n}$, provided β is sufficiently large. The same holds for all $n!$ placements of G_2, so with high probability we have $\operatorname{disc}^+(G, H) \leq L$, as required.

We now turn to the negative discrepancy (since this is a relatively weak result, we sketch the argument here). The upper bound follows from (15), so we need only prove the lower bound. Note that with high probability we have $e(G_1) = (1 + o(1))p\binom{n}{k}$ and $e(G_2) = (1 + o(1))\binom{n}{k}$.

Suppose first that $pqn^k = O(1)$, so $\max\{p, q\} \to 0$. Choose $\lambda = \lambda(n) \to \infty$ such that $\max\{p, q\} = o(1/\lambda)$. Placing G_1 and G_2 at random on the same vertex set, the expected number of common edges is $O(1)$ and so with high probability we have at most λ common edges, say $e_1, \ldots, e_t$. With high probability, the e_i are vertex-disjoint. So we can pick vertices $v_i \in e_i$, for $i = 1, \ldots, t$; and vertices $w_1, \ldots, w_t$ that do not lie in any of the edges. Now, for each i, exchange the vertices v_i and w_i in G_1: the expected number of common edges is then at most $\lambda(p + q) + O(pq\lambda n^{k-1}) = o(1)$, and so with high probability there are no common edges, and we have found a suitable placement.

Next suppose that $pqn^k \to \infty$, but $\min\{p, q\}\binom{n}{k} \leq n$, say $p\binom{n}{k} \leq n$. We first choose G_1: with high probability this contains a set of at least $p\binom{n}{k}/(10 \cdot k!)$ vertex-disjoint edges. We now generate G_2, initially picking *non-edges* without replacement until we have a matching of size at least $p\binom{n}{k}/(10 \cdot k!)$. This succeeds with high probability, so we can choose a random mapping for which the two matchings coincide: with high probability, adding the edges of G_2 will now give negative discrepancy $\Omega(pqn^k)$.

Finally, suppose that $pqn^k \to \infty$ and $\min\{p, q\}n^k = \Omega(n)$. Choose $\lambda = \lambda(n)$ so that $\lambda^2 pqn^{k-1} = o(\log n)$. We follow the algorithm above with some changes. In Step 1, we greedily choose V_3 to be the first $\lfloor n/20 \rfloor$ vertices $v \in V_2$ with $|\Gamma_1(v, V_0)| \leq \lambda pR$; and W_3 to be the first $\lfloor n/10 \rfloor$ vertices $w \in W_2$ with $|\Gamma_1(w, W_0)| \leq \lambda qR$. It follows that, with high probability, $e^{(1,k-1)}(V_3, V_0) = (1 + o(1))|V_3|R$ and $e^{(1,k-1)}(W_3, W_0) = (1 + o(1))|W_3|R$. In Step 2, we simply take $A_v = \Gamma_1(v, V_0)$ and $B_v = \Gamma_2(w, W_0)$. In Step 3, we join v to w if $\phi(A_v) \cap B_w = \emptyset$. Note that $\mathbb{E}[|A_v \cap B_w|] = o(\log n)$, so writing $\alpha = |A_v|/R$ and $\beta = |B_w|/R$ (and considering the effects of fixing B_w and choosing the elements of A_v one at a time) we have $\mathbb{P}[A_v \cap B_w = \emptyset] \geq (1 - 2\beta)^{\alpha R} \geq \exp(-3\alpha\beta R) = n^{-o(1)}$. The argument can now be completed as in Theorem 5. $\square$

In the diagonal case ($p = q$), Theorem 5 and Theorem 6 give the following corollary.

Corollary 7. *Fix $k \geq 2$ and $\epsilon > 0$. Suppose $p = p(n) \leq 1/2$ satisfies $pn^k \to \infty$, and let G_1, G_2 be random hypergraphs chosen independently from $\mathcal{G}^{(k)}(n,p)$. If $p = \Omega(\sqrt{\log n/n^{k-1}})$ then, with failure probability* $\exp(-n^{1-\epsilon})$*,*

$$\mathrm{disc}^+(G_1, G_2) = \Theta(n^{(k+1)/2} p(1-p)\sqrt{\log n})$$

and

$$\mathrm{disc}^-(G_1, G_2) = \Theta(n^{(k+1)/2} p(1-p)\sqrt{\log n}).$$

If $p = \sqrt{\log n/\beta n^{k-1}}$, where $\beta = \beta(n) \to \infty$, then with high probability

$$\mathrm{disc}^+(G_1, G_2) = \Theta\left(\min\left\{pn^k, \frac{n \log n}{\log \beta}\right\}\right)$$

and

$$\mathrm{disc}^-(G_1, G_2) = \Theta(p^2 n^k).$$

5 Conclusion

We have determined to within a constant factor the positive and negative discrepancies of a pair of random hypergraphs or tournaments. A number of interesting questions still remain, and we mention a few here.

- As noted in the introduction, Spencer has shown that with high probability a random tournament T of order n satisfies $\mathrm{disc}(T, TT_n) = \Theta(n^{3/2})$. Can more be said about the distribution of $\mathrm{disc}(T, TT_n)$? What is the behaviour of the upper tail?

- The *Slater index* $i(T)$ of a tournament T is the minimum number of arcs that must be reversed to make T transitive. If T has order n, then $i(T) = \frac{1}{2}\binom{|T|}{2} - \mathrm{disc}(T, TT_n)$. What is the maximum value of the Slater index, or equivalently the minimum value of $\mathrm{disc}(T, TT_n)$, over tournaments T of order n? It was conjectured by Bermond [2] that perhaps a regular tournament is extremal (see Charon and Hudry [8] for further results and discussion).

- Following the results of [6], we should expect that for any pair T, T' of tournaments of order n we have $\mathrm{disc}(T, T') \geq cn^{3/2}$. Is this true? And what pair of tournaments minimizes this quantity?

- What can we say about $\mathrm{disc}(G, H)$, and about signed discrepancies, if G and H are *pseudorandom* graphs?

- What is the threshold $p = p(n)$ for the property that, for $G_1, G_2 \in \mathcal{G}(n, p)$, there is almost surely a packing of G_1 and G_2 into K_n? More generally, what is the range of p and q for which this holds almost surely for $G_1 \in \mathcal{G}(n, p)$ and $G_2 \in \mathcal{G}(n, q)$? The same questions arise for hypergraphs. We will return to this in another paper [4].

Note. As we were completing this paper, we discovered that Ma, Naves and Sudakov [16] had independently and simultaneously proved results very similar to our Theorems 5 and 6.

References

[1] Bonnie Berger and Peter W. Shor, Tight bounds for the maximum acyclic subgraph problem, *J. Algorithms* **25** (1997), 1–18.

[2] J.-C. Bermond, Ordres à distance minimum dun tournoi et graphes partiels sans circuits maximaux, *Math. Sci. Hum.* **37** (1972), 5-25

[3] B. Bollobás, Random graphs. Second edition. Cambridge Studies in Advanced Mathematics, 73, Cambridge University Press, Cambridge, 2001. xviii+498 pp.

[4] B. Bollobás, S. Janson and A.D. Scott, Packing random graphs and hypergraphs, *manuscript*

[5] B. Bollobás and A. D. Scott, Discrepancy in graphs and hypergraphs, In: More Sets, Graphs and Numbers, Bolyai Society Mathematical Studies **15** (E. G. Gyula, O. H. Katona and L. Lovász, Eds.), Springer, Berlin, 2006, pp. 33–56.

[6] B. Bollobás and A.D. Scott, Intersections of graphs, *J. Graph Theory* **66** (2011), 261–282.

[7] B. Bollobás and A.D. Scott, Intersections of hypergraphs, *Journal of Combinatorial Theory, Series B, to appear.*

[8] Irène Charon and Olivier Hudry, Links between the Slater index and the Ryser index of tournaments, *Graphs Combin.* **19** (2003), 309–322.

[9] A. Czygrinow, S. Poljak and V. Rödl, Constructive quasi-Ramsey numbers and tournament ranking *SIAM J. Discrete Math.* **12** (1999), 48–63.

[10] P. Erdős, M. Goldberg, J. Pach and J. Spencer, Cutting a graph into two dissimilar halves, *J Graph Theory* **12** (1988), 121–131.

[11] P. Erdős and J. W. Moon, On sets of consistent arcs in a tournament, *Canad. Math. Bull.* **8** (1965), 269–271.

[12] P. Erdős and J. Spencer, Imbalances in k-colorations, *Networks* **1** (1971/1972), 379–385.

[13] W. Fernandez de la Vega, On the maximum cardinality of a consistent set of arcs in a random tournament, *J. Combin. Theory Ser. B* **35** (1983), 328–332.

[14] H. A. Jung, On subgraphs without cycles in a tournament. Combinatorial theory and its applications, II (Proc. Colloq., Balatonfüred, 1969), pp. 675–677. North-Holland, Amsterdam, 1970.

[15] J.-F. Laslier, Tournament solutions and majority voting, *Springer-Verlag, Berlin*, 1997. xiv+255 pp.

[16] J. Ma, H. Naves and B. Sudakov, Discrepancy of random graphs and hypergraphs, submitted.

[17] Svatopluk Poljak, Vojtěch Rödl and Joel Spencer, Tournament ranking with expected profit in polynomial time, *SIAM J. Discrete Math.* **1** (1988), 372–376.

[18] K. B. Reid, On sets of arcs containing no cycles in a tournament, *Canad. Math. Bull.* **12** (1969), 261–264.

[19] P. Slater, Inconsistencies in a schedule of paired comparisons, *Biometrika* **48** (1961), 303–312

[20] J. Spencer, Optimal ranking of tournaments, *Networks* **1** (1971/72), 135–138.

[21] J. Spencer, Optimally ranking unrankable tournaments, *Period. Math. Hungar.* **11** (1980), 131–144.