

Discrepancy in graphs and hypergraphs

B. Bollobás ^{*†} A.D. Scott [‡]

Abstract

Let G be a graph with n vertices and $p\binom{n}{2}$ edges, and define the discrepancies $\text{disc}_p^+(G) = \max_{Y \subset V(G)} \{e(Y) - p\binom{|Y|}{2}\}$ and $\text{disc}_p^-(G) = \max_{Y \subset V(G)} \{p\binom{|Y|}{2} - e(Y)\}$. We prove that if $p(1-p) \geq 1/n$ then $\text{disc}_p^+(G)\text{disc}_p^-(G) \geq p(1-p)n^3/6400$. We also prove a similar inequality for k -uniform hypergraphs, and give related results concerning 2-colourings of k -uniform hypergraphs. Our results extend those of Erdős, Goldberg, Pach and Spencer [6] and Erdős and Spencer [7].

1 Introduction

The *discrepancy* of a graph G is $\text{disc}(G) = \max_{Y \subset V(G)} |e(Y) - \frac{1}{2}\binom{|Y|}{2}|$, where we write $e(Y) = e(G[Y])$ for the number of edges of G spanned by Y . If G has edge density $1/2$ then the discrepancy can be seen as a measure of how uniformly the edges are distributed among the vertices; see Sós [11] and Beck and Sós [1]) for more discussion and a general account of discrepancy. Erdős and Spencer [7] showed that for some constant $c > 0$ every graph G of order n satisfies $\text{disc}(G) \geq cn^{3/2}$. More generally, they showed that for every $k \geq 3$ there is a constant $c_k > 0$ such that if H is a k -uniform hypergraph of order n then $\text{disc}(H) \geq c_k n^{(k+1)/2}$, where $\text{disc}(H) = \max_{Y \subset V(H)} |e(Y) - \frac{1}{2}\binom{|Y|}{k}|$. By considering random graphs they showed that this bound is sharp up to the value of the constant.

^{*}Trinity College, Cambridge CB2 1TQ and Department of Mathematical Sciences, University of Memphis, Memphis TN38152; email: bollobas@msci.memphis.edu

[†]Research supported in part by NSF grant DSM 9971788 and DARPA grant F33615-01-C-1900

[‡]Department of Mathematics, University College London, Gower Street, London WC1E 6BT; email: scott@math.ucl.ac.uk

Now suppose that G is a graph with $e(G) = m = p\binom{|G|}{2}$, where $p < 1/2$, so that we expect a random subset $Y \subset V(G)$ to span a subgraph with $p\binom{|Y|}{2}$ edges. Then a more appropriate measure of edge distribution is given by the quantity $\text{disc}_p(G) = \max_{Y \subset V(G)} |e(Y) - p\binom{|Y|}{2}|$. Erdős, Goldberg, Pach and Spencer [6] showed that in this case $\text{disc}_p(G) \geq c\sqrt{mn}$, where c is an absolute constant.

A subset of vertices with large discrepancy can clearly be either more or less dense than the whole graph. Let us define the *positive discrepancy* by $\text{disc}^+(G) = \max_{Y \subset V(G)} \left\{ e(Y) - \frac{1}{2}\binom{|Y|}{2} \right\}$ and the *negative discrepancy* by $\text{disc}^-(G) = \max_{Y \subset V(G)} \left\{ \frac{1}{2}\binom{|Y|}{2} - e(Y) \right\}$. Then a random graph $G \in \mathcal{G}(n, 1/2)$ shows that it is possible to have $\max\{\text{disc}^+(G), \text{disc}^-(G)\} \leq cn^{3/2}$. The one-sided discrepancy can be smaller: for instance, the complete bipartite graph $K_{n/2, n/2}$ has positive discrepancy $O(n)$, although its negative discrepancy is cn^2 . Similarly, the graph $2K_{n/2}$ has positive discrepancy $O(n)$ but negative discrepancy cn^2 . These examples show that we can guarantee small discrepancy on one side provided we allow large discrepancy on the other. In this paper we shall prove that positive discrepancy substantially smaller than $n^{3/2}$ guarantees negative discrepancy substantially larger than $n^{3/2}$; indeed, we shall quantify the trade-off between positive and negative discrepancies. Surprisingly, the correct measure turns out to be the product $\text{disc}^+(G)\text{disc}^-(G)$.

We remark that a different type of negative discrepancy was considered by Erdős, Faudree, Rousseau and Schelp [5] with the idea of showing that graphs with small negative discrepancy contain complete subgraphs of fixed size. For further recent results in this direction see Krivelevich [9] and Keevash and Sudakov [8].

We begin with some definitions. For a k -uniform hypergraph G , a real $p \in [0, 1]$ and $X \subset V(G)$ let

$$d_p(X) = e(X) - p\binom{|X|}{k}.$$

For disjoint sets of vertices X and Y , let

$$d_p(X, Y) = e(X, Y) - p|X||Y|.$$

Then we define

$$\text{disc}_p^+(G) = \max_{X \subset V(G)} d_p(X) \tag{1}$$

and

$$\text{disc}_p^-(G) = - \min_{X \subset V(G)} d_p(X), \quad (2)$$

and set

$$\text{disc}_p(G) = \max_{X \subset V(G)} |d_p(X)| = \max\{\text{disc}_p^+(G), \text{disc}_p^-(G)\}.$$

If p is not specified we assume $p = 1/2$, so for instance $\text{disc}(G) = \text{disc}_{1/2}(G)$. Note that the cases $p = 0$ and $p = 1$ are trivial, and that if $e(G) = p_0 \binom{|G|}{2}$ we have $\text{disc}_p(G) \geq |d_p(V(G))| = |p - p_0| \binom{|G|}{2}$. We will therefore usually take p with $e(G) = p \binom{|G|}{2}$. Note that, for any p , $\text{disc}_p^+(G) = \text{disc}_{1-p}^-(\overline{G})$ and $\text{disc}_p^-(G) = \text{disc}_{1-p}^+(\overline{G})$. We shall usually assume $p \leq 1/2$, since if $p > 1/2$ we may replace G by \overline{G} and p by $1 - p$.

We remark that it does not make much difference if we restrict the definitions in (1) and (2) to sets X of size $n/2$ (or some other size cn): as noted by Erdős, Goldberg, Pach and Spencer [6], this would change the resulting discrepancy by at most a constant factor.

We shall frequently refer to a *random bipartition* $V = X \cup Y$. Unless otherwise stated, this means a random bipartition in which each vertex is assigned independently to X or Y with equal probability. Throughout the paper we shall use ϵ_i and ρ_j for sequences of independent Bernoulli random variables, with $\epsilon_i \in \{+1, -1\}$ and $\rho_j \in \{0, 1\}$, each taking either value with probability $1/2$.

The rest of the paper is organized as follows. In section 2 we give lower bounds on $\text{disc}_p(G)$ for graphs; in section 3 we turn our attention to hypergraphs. Finally, in section 4, we consider some related results concerning subgraphs of a fixed graph or hypergraph.

2 Discrepancy of graphs

In this section we prove our results on graph discrepancy. Let G be a graph of order n and size $p \binom{n}{2}$. If G is very sparse, say $0 < p \leq 1/(n-1)$, then taking the union of $p \binom{n}{2} / 2$ edges from G gives a subgraph with at most $p \binom{n}{2}$ vertices, so $\text{disc}_p^+(G) \geq p \binom{n}{2} / 2 - p(p \binom{n}{2})^2 / 2 \geq pn^2 / 5 - p^3 n^4 / 8 > pn^2 / 20$ for sufficiently large n , while since G has average degree at most 1 it contains an independent set of size at least $n/2$, and so $\text{disc}^-(G) \geq p \binom{n/2}{2} > pn^2 / 9$ for sufficiently large n . On the other hand, $\max\{\text{disc}^+(G), \text{disc}^-(G)\} \leq e(G) < pn^2 / 2$.

Thus $\text{disc}_p^+(G)$ and $\text{disc}_p^-(G)$ are both $\Theta(pn^2)$. A similar argument applies if G is very dense, with $p \geq 1 - 1/(n-1)$. (More precise bounds are given by Erdős, Goldberg, Pach and Spencer [6].)

We therefore restrict our attention to graphs with $p(1-p) \geq 1/n$. Our main result is the following.

Theorem 1. *Let G be a graph of order n and size $p\binom{n}{2}$, where $p(1-p) \geq 1/n$. Then*

$$\text{disc}_p^+(G)\text{disc}_p^-(G) \geq p(1-p)n^3/6400. \quad (3)$$

As an immediate corollary we get the following result of Erdős, Goldberg, Pach and Spencer [6].

Corollary 2. *Let G be a graph of order n and size $p\binom{n}{2}$, where $p(1-p) \geq 1/n$. Then*

$$\text{disc}_p(G) \geq \sqrt{p(1-p)}n^{3/2}/80.$$

We remark that the result of Erdős and Spencer for graphs can easily be deduced from Theorem 1: if $1/3 \leq p \leq 2/3$ then $\text{disc}(G) \geq \frac{1}{2}(\text{disc}_p^+(G) + \text{disc}_p^-(G)) \geq n^{3/2}/160$, while otherwise $\text{disc}(G) \geq |e(G) - \frac{1}{2}\binom{n}{2}| \geq \binom{n}{2}/6 \geq n^{3/2}/12$.

We also remark that, for $r \geq 2$, the Turán graph $T_r(n)$ gives a bound on the optimal constant in (3). Defining p by $t_r(n) = e(T_r(n)) = p\binom{n}{2}$, we have $p \sim 1 - \frac{1}{r}$. A little calculation shows that

$$\text{disc}_p^+(T_r(n)) = \frac{pn}{8} + O(r), \quad (4)$$

and, for r even,

$$\text{disc}_p^-(T_r(n)) = (1 + o(1))\frac{(1-p)n^2}{8},$$

which implies

$$\text{disc}_p^+(T_r(n))\text{disc}_p^-(T_r(n)) \leq (1 + o(1))\frac{p(1-p)n^3}{64}.$$

Before turning to the proof of Theorem 1, we make some comments about one-sided discrepancies. Since every graph with n vertices and $t_r(n)$ edges contains a subgraph of order u and size at least $t_r(u)$ for every $1 \leq u \leq n$,

the Turán graphs $T_r(n)$ have minimal positive p -discrepancy among graphs of order n and size $t_r(n)$. Thus (4) gives an optimal bound in these cases, which have density $p \sim 1 - \frac{1}{r}$. To obtain a similar bound for arbitrary densities, we define an extension of the Turán numbers for non-integral r . Given an integer $n \geq 1$ and a real number $r \geq 1$, we can write $n = qr + s$, where q is an integer and $0 \leq s < r$. We define the fractional Turán number $t_r(n)$ by

$$t_r(n) = \binom{n}{2} - \bar{t}_r(n),$$

where

$$\bar{t}_r(n) = s \binom{q+1}{2} + (r-s) \binom{q}{2}.$$

Note that this is consistent with the definition of Turán numbers when r is integral; it is convenient to work with the quantity $\bar{t}_r(n)$ instead of $t_r(n)$.

A bound matching (4) will follow from the following result.

Lemma 3. *Suppose that $n \geq 1$ is an integer and $1 \leq r \leq n$. Let G be a graph with n vertices and at least $t_r(n)$ edges. Then, for $2 \leq u \leq n$, G contains a subgraph with u vertices and at least $t_r(u)$ edges.*

Proof. It is enough to prove the theorem when $u = n - 1$. Taking complements, this is equivalent to showing that if $e(G) \leq \bar{t}_r(n)$ then there is a vertex v such that $e(G \setminus v) \leq \bar{t}_r(n - 1)$. We may also assume $r > 1$, or else G is empty.

Adding edges if necessary, we may assume that

$$e(G) = \lfloor \bar{t}_r(n) \rfloor = \bar{t}_r(n) - \eta, \tag{5}$$

where $0 \leq \eta < 1$. Thus if $n = qr + s$,

$$\begin{aligned} e(G) &= s \binom{q+1}{2} + (r-s) \binom{q}{2} - \eta \\ &= \frac{1}{2} (rq^2 + (2s-r)q) - \eta. \end{aligned}$$

A short calculation shows that

$$\Delta(G) \geq \left\lceil \frac{2e(G)}{n} \right\rceil = q - 1 + \left\lceil \frac{qs + s - 2\eta}{qr + s} \right\rceil. \tag{6}$$

By (5) and (6), it is sufficient to show that

$$q - 1 + \left\lceil \frac{qs + s - 2\eta}{qr + s} \right\rceil + \eta \geq \bar{t}_r(n) - \bar{t}_r(n-1). \quad (7)$$

If $q = 0$ then we have a complete graph and are done immediately. Thus we may assume that $q \geq 1$. Now if $s \geq 1$, then it is easily seen that

$$\bar{t}_r(n) - \bar{t}_r(n-1) = q,$$

while if $0 \leq s < 1$, then a simple calculation shows that

$$\bar{t}_r(n) - \bar{t}_r(n-1) = q - 1 + s. \quad (8)$$

Now if $s > \eta$ then $qs + s > 2\eta$, and so the left side of (7) is at least $q + \eta$, and thus (7) is satisfied. If $s \leq \eta$, however, then $0 \leq s < 1$, so (8) holds. It is then sufficient by (7) to show that

$$\left\lceil \frac{qs + s - 2\eta}{qr + s} \right\rceil + \eta \geq s,$$

which holds provided

$$\frac{qs + s - 2\eta}{qr + s} > -1.$$

But $qr + s = n$ and $qs + s - 2\eta \geq -2\eta > -2$, so this holds for $n \geq 2$. \square

Calculating as in (4), we obtain the following result.

Corollary 4. *For $0 \leq p \leq 1$, every graph G with n vertices and $p\binom{n}{2}$ edges satisfies*

$$\text{disc}_p^+(G) \geq \frac{pn}{8} + O\left(\frac{1}{1-p}\right).$$

We now turn to the proof of Theorem 1. We shall need two simple inequalities (these follow easily from the Littlewood-Khinchin inequality, see [10], [12], [13]; however, we give short proofs at the end of the section). Recall that ϵ_i and ρ_i are i.i.d. Bernoulli with $\epsilon_i \in \{+1, -1\}$ and $\rho_i \in \{0, 1\}$.

Lemma 5. *For $n \geq 1$,*

$$\mathbb{E} \left| \sum_{i=1}^n \epsilon_i \right| \geq \sqrt{n/2}.$$

Lemma 6. Let $\mathbf{a} = (a_i)_{i=1}^n$ be a sequence of real numbers, and A a real number. Then

$$\mathbb{E} \left| \sum_{i=1}^n \epsilon_i a_i - A \right| \geq \|\mathbf{a}\|_1 / \sqrt{2n}$$

and

$$\mathbb{E} \left| \sum_{i=1}^n \rho_i a_i - A \right| \geq \|\mathbf{a}\|_1 / \sqrt{8n}.$$

Our main tool in the proof of Theorem 1 is the following lemma, which shows that in a random bipartition of a graph G , we do not expect the vertex neighbourhoods to split too evenly.

Lemma 7. Let G be a graph of order n and size $p\binom{n}{2}$, where $p(1-p) \geq 1/n$. Let $V(G) = X \cup Y$ be a random bipartition. Then

$$\mathbb{E} \sum_{x \in X} \left| |\Gamma(x) \cap Y| - p|Y| \right| \geq \sqrt{p(1-p)} n^{3/2} / 20.$$

Proof. We may assume $p \leq 1/2$ since we may take complements and replace p by $1-p$. Suppose $x \in V(G)$ has degree $d = d(x) = p(n-1) + r(x)$. For $v \neq x$, define $e_v = 1$ if $xv \in E(G)$ and $e_v = 0$ otherwise. Then

$$\begin{aligned} \mathbb{E} \left| |\Gamma(x) \cap Y| - p|Y \setminus \{x\}| \right| &= \mathbb{E} \left| \sum_{v \neq x} \rho_v (e_v - p) \right| \\ &= \mathbb{E} \left| \frac{1}{2} \sum_{v \neq x} (e_v - p) + \frac{1}{2} \sum_{v \neq x} \epsilon_v (e_v - p) \right| \\ &\geq \max \left\{ \frac{1}{2} |d - (n-1)p|, \frac{1}{2} \left| \mathbb{E} \sum_{v \neq x} \epsilon_v (e_v - p) \right| \right\}, \end{aligned}$$

since $\sum_{v \neq x} (e_v - p) = d - (n-1)p$ and the distribution of $\sum_{v \neq x} \epsilon_v (e_v - p)$ is symmetric about 0. Now, by Lemma 5,

$$\begin{aligned} \mathbb{E} \left| \sum_{v \neq x} \epsilon_v (e_v - p) \right| &= \mathbb{E} \left| \sum_{i=1}^d \epsilon_i (1-p) + \sum_{i=d+1}^{n-1} \epsilon_i (-p) \right| \\ &\geq \mathbb{E} \left| \sum_{i=1}^d \epsilon_i (1-p) \right| \\ &\geq (1-p) \sqrt{d/2} \end{aligned}$$

and so

$$\mathbb{E} \left| |\Gamma(x) \cap Y| - p|Y \setminus \{x\}| \right| \geq \frac{1}{2} \max\{|r(x)|, (1-p)\sqrt{d(x)/2}\}.$$

Now for $x \in V = V(G)$ let $I(x) = 1$ if $x \in X$ and $I(x) = 0$ otherwise. Then, since $I(x)$ and $|\Gamma(x) \cap Y|$ are independent random variables,

$$\begin{aligned} \mathbb{E} \sum_{x \in X} \left| |\Gamma(x) \cap Y| - p|Y \setminus \{x\}| \right| &= \mathbb{E} \sum_{x \in V} I(x) \left| |\Gamma(x) \cap Y| - p|Y \setminus \{x\}| \right| \\ &= \frac{1}{2} \sum_{x \in V} \mathbb{E} \left| |\Gamma(x) \cap Y| - p|Y \setminus \{x\}| \right| \\ &\geq \frac{1}{4} \max\left\{ \sum_{x \in V} |r(x)|, \sum_{x \in V} (1-p)\sqrt{d(x)/2} \right\} \\ &\geq \frac{1}{8} \sum_{x \in V} (|r(x)| + (1-p)\sqrt{d(x)/2}). \end{aligned}$$

Note that the first equality holds as $Y = Y \setminus \{x\}$ if $I(x) = 1$. Furthermore, $|r(x)| + (1-p)\sqrt{d(x)/2}$ is minimized when $r(x) = 0$ and so $d(x) = p(n-1)$. Thus

$$\begin{aligned} \frac{1}{8} \sum_{x \in V} \left(|r(x)| + (1-p)\sqrt{d(x)/2} \right) &\geq \frac{1}{8} n(1-p)\sqrt{p(n-1)/2} \\ &\geq \sqrt{p(1-p)}n^{3/2}/20, \end{aligned}$$

since $p \leq 1/2$ and we may assume $n \geq 3$. □

After this preparation, we are ready to prove Theorem 1.

Proof of Theorem 1. Since (3) is symmetric in p and $1-p$, we may replace G by its complement \overline{G} , and so we may assume that $\text{disc}_p^+(G) \leq \text{disc}_p^-(G)$. If $\text{disc}_p^+(G) \geq \sqrt{p(1-p)}n^{3/2}/80$ we are done. Otherwise, suppose $\text{disc}_p^+(G) = \sqrt{p(1-p)}n^{3/2}/80\alpha$, where $\alpha \geq 1$. We shall show that

$$\text{disc}_p^-(G) \geq \alpha\sqrt{p(1-p)}n^{3/2}/80, \tag{9}$$

so $\text{disc}_p^+(G)\text{disc}_p^-(G) \geq p(1-p)n^3/6400$.

Let $V(G) = X \cup Y$ be a random bipartition. Then since $p(1-p) \geq 1/n$, it follows from Lemma 7 that

$$\mathbb{E} \sum_{x \in X} \left| |\Gamma(x) \cap Y| - p|Y| \right| \geq \sqrt{p(1-p)}n^{3/2}/20. \quad (10)$$

Now let $X^+ = \{x \in X : |\Gamma(x) \cap Y| \geq p|Y|\}$ and $X^- = X \setminus X^+$; so

$$d_p(X, Y) = \sum_{x \in X^+} (|\Gamma(x) \cap Y| - p|Y|) + \sum_{x \in X^-} (|\Gamma(x) \cap Y| - p|Y|).$$

Since $\mathbb{E}d_p(X, Y) = 0$, we have

$$\mathbb{E} \sum_{x \in X^+} \left| |\Gamma(x) \cap Y| - p|Y| \right| = \mathbb{E} \sum_{x \in X^-} \left| |\Gamma(x) \cap Y| - p|Y| \right|$$

and so by (10)

$$\mathbb{E} d_p(X^+, Y) = \mathbb{E} \sum_{x \in X^+} (|\Gamma(x) \cap Y| - p|Y|) \geq \sqrt{p(1-p)}n^{3/2}/40. \quad (11)$$

Now $\mathbb{E}d_p(Y) = 0$, so (11) implies

$$\mathbb{E}(d_p(X^+, Y) + \alpha d_p(Y)) \geq \sqrt{p(1-p)}n^{3/2}/40. \quad (12)$$

Let X^+, Y be a pair of sets achieving at least the expectation in (12) and let Z be a random subset of X^+ , where each vertex of X^+ is chosen independently with probability $1/\alpha$. Then it follows from (12) that

$$\begin{aligned} \mathbb{E} d_p(Z \cup Y) &= \mathbb{E}(d_p(Z) + d_p(Z, Y) + d_p(Y)) \\ &= \frac{1}{\alpha^2} d_p(X) + \frac{1}{\alpha} d_p(X, Y) + d_p(Y) \\ &\geq \frac{1}{\alpha^2} d_p(X) + \frac{1}{\alpha} \sqrt{p(1-p)}n^{3/2}/40. \end{aligned}$$

Since $\text{disc}_p^+(G) = \sqrt{p(1-p)}n^{3/2}/80\alpha$, this implies

$$d_p(X)/\alpha^2 \leq -(1/\alpha)\sqrt{p(1-p)}n^{3/2}/80$$

and so $d_p(X) \leq -\alpha\sqrt{p(1-p)}n^{3/2}/80$, which gives the desired lower bound on $\text{disc}_p^-(G)$. \square

Finally in this section we give the proofs of Lemmas 5 and 6, postponed from earlier.

Proof of Lemma 5. A simple calculation shows that for $n = 2k$ we have $\mathbb{E}|\sum_{i=1}^n \epsilon_i| = 2^{1-2k} k \binom{2k}{k}$ and for $n = 2k + 1$ we have $\mathbb{E}|\sum_{i=1}^n \epsilon_i| = 2^{-2k} (2k + 1) \binom{2k}{k} = \mathbb{E}|\sum_{i=1}^{n+1} \epsilon_i|$. Let $s_n = \mathbb{E}|\sum_{i=1}^n \epsilon_i|/\sqrt{n}$. Then, for $k \geq 1$, $s_{2k+2}/s_{2k} = (k + \frac{1}{2})/\sqrt{k(k+1)} > 1$ and, for $k \geq 0$, $s_{2k+3}/s_{2k+1} = \sqrt{(k + \frac{3}{2})(k + \frac{1}{2})}/(k + 1) < 1$. Thus $(s_{2k})_{k=1}^\infty$ is increasing and $(s_{2k+1})_{k=0}^\infty$ is decreasing; both converge to $\mathbb{E}|N(0, 1)| = \sqrt{2/\pi}$. Therefore $s_n \geq s_2 = 1/\sqrt{2}$ for all n . \square

Proof of Lemma 6. We may clearly assume that all a_i are nonnegative. Since $\sum_{i=1}^n \epsilon_i a_i$ is symmetric about 0, the expectation is minimized for a given \mathbf{a} when $A = 0$. Now if $a_i \neq a_j$ then let $a'_i = a'_j = (a_i + a_j)/2$; it is easily checked that $\mathbb{E}|B + \epsilon_i a_i + \epsilon_j a_j| \geq \mathbb{E}|B + \epsilon_i a'_i + \epsilon_j a'_j|$ for every real B . It follows that $\mathbb{E}|\sum_{i=1}^n \epsilon_i a_i| \geq \mathbb{E}|\sum_{i=1}^n \epsilon_i a|$, where $a = \sum_{i=1}^n a_i/n$. Thus, by Lemma 5,

$$\mathbb{E}|\sum_{i=1}^n \epsilon_i a_i - A| \geq a \mathbb{E}|\sum_{i=1}^n \epsilon_i| \geq a \sqrt{n/2} = \|\mathbf{a}\|_1/\sqrt{2n}.$$

The second inequality follows directly from the first. \square

Note that in fact proof of Lemma 5 implies the inequalities $\mathbb{E}|\sum_{i=1}^n a_i| \geq \sqrt{2/\pi n} \|\mathbf{a}\|_1$ if n is odd and $\mathbb{E}|\sum_{i=1}^n a_i| \geq (1 + o(1))\sqrt{2/\pi n} \|\mathbf{a}\|_1$ for general n .

3 Hypergraph discrepancy

In this section we turn our attention to hypergraphs. After defining a little notation, we begin with a result for weighted hypergraphs; we then turn to the consideration of unweighted hypergraphs.

If G is the complete k -uniform hypergraph with edge-weighting w and $X \subseteq V(G)$, we define

$$d(X) = \sum_{K \in X^{(k)}} w(K).$$

As in definitions (1) and (2) we define $\text{disc}^+(G) = \max_{X \subseteq V(G)} d(X)$ and $\text{disc}^-(G) = -\min_{X \subseteq V(G)} d(X)$; we also define

$$\text{disc}(G) = \max\{\text{disc}^+(G), \text{disc}^-(G)\}.$$

Note that this is consistent with the definitions for an unweighted hypergraph G by taking $w(e) = 1$ if $e \in E(G)$ and $w(e) = -1$ otherwise.

For disjoint sets X_1, \dots, X_t and integers k_1, \dots, k_t such that $\sum_{i=1}^t k_i = k$, we define

$$d_{k_1, \dots, k_t}(X_1, \dots, X_t) = \sum' w(e),$$

where the sum is over edges e with $|e \cap X_i| = k_i$ for every i .

We can now state the first result of the section.

Theorem 8. *Let G be the complete k -uniform hypergraph of order n with edge-weighting w such that $\sum w(e) = 0$ and $\sum |w(e)| = \binom{n}{k}$. Then*

$$\text{disc}^+(G)\text{disc}^-(G) \geq 2^{-14k^2} n^{k+1}.$$

We shall need three lemmas. In the first lemma we use the fact that if $P(x)$ is a polynomial of degree k with $\sup_{x \in [0,1]} |P(x)| \leq 1$ then every coefficient of $P(x)$ has absolute value at most $2^k k^{2k}/k!$. (Tamás Erdélyi [4] pointed out to us that this is an elementary consequence of Markov's Inequality; see [3].)

Lemma 9. *If G is a complete k -uniform hypergraph with edge-weighting w and $\text{disc}(G) \leq M$ then for disjoint subsets X, Y of $V(G)$ and $0 \leq i \leq k$,*

$$|d_{i, k-i}(X, Y)| \leq 2^{2k^2} M.$$

Proof. Let Z be a random subset of X , where each vertex is chosen independently with probability p . Then

$$\mathbb{E}(d(Z \cup Y)) = \sum_{i=0}^k p^i d_{i, k-i}(X, Y).$$

Since $\text{disc}(G) \leq M$, it follows that $\max_{0 \leq p \leq 1} |\sum_{i=0}^k p^i d_{i, k-i}(X, Y)| \leq M$ and so $\max_{0 \leq i \leq k} |d_{i, k-i}(X, Y)| \leq 2^k k^{2k} M/k! \leq 2^{2k^2} M$. \square

We also need an analogue of Lemma 7.

Lemma 10. *Let G be a complete k -uniform hypergraph of order n with edge-weighting w . Let $V(G) = U \cup W$ be a random bipartition. Then*

$$\mathbb{E} \sum_{K \in U^{(k-1)}} |d_{k-1, 1}(K, W)| \geq k2^{-k} \sum_{L \in V(G)^{(k)}} |w(L)|/\sqrt{2n}.$$

Proof. Let $V = V(G) = U \cup W$ be a random bipartition. Given $K \in V^{(k-1)}$, it follows from Lemma 6 that

$$\mathbb{E}|d_{k-1,1}(K, W \setminus K)| \geq \sum_{v \in V \setminus K} |w(K \cup \{v\})|/\sqrt{8n}.$$

Since the event $\{K \subset U\}$ and the random variable $d_{k-1,1}(K, W \setminus K)$ are independent, and each edge $L \in V^{(k)}$ occurs k times as $K \cup \{v\}$, we have

$$\begin{aligned} \mathbb{E} \sum_{K \in U^{(k-1)}} |d_{k-1,1}(K, W)| &= \sum_{K \in V^{(k-1)}} \mathbb{P}(K \subset U) \mathbb{E}|d_{k-1,1}(K, W \setminus K)| \\ &\geq \sum_{K \in V^{(k-1)}} 2^{-k+1} \sum_{v \in V \setminus K} |w(K \cup \{v\})|/\sqrt{8n} \\ &= k2^{-k} \sum_{L \in V^{(k)}} |w(L)|/\sqrt{2n}. \end{aligned}$$

□

The following lemma will be useful several times.

Lemma 11. *Let G be a k -uniform hypergraph of order n with edge-weighting w . Suppose that $\alpha \geq 1$ and X, Y are disjoint subsets of $V(G)$ with*

$$d_{1,k-1}(X, Y) + \alpha d(Y) = M \geq 0. \quad (13)$$

Then either

$$\text{disc}^+(G) \geq 2^{-3k^2} M/\alpha$$

or

$$\text{disc}^-(G) \geq 2^{-3k^2} M\alpha.$$

Proof. If $|d_{i,k-i}(X, Y)| \geq 2^{-k^2} \alpha M$ for some $0 \leq i \leq k$ then we are done by Lemma 9. Otherwise, let Z be a random subset of X , obtained by choosing

each vertex of X independently with probability $1/\alpha$. Then

$$\begin{aligned}
\mathbb{E} d(Z \cup Y) &= \mathbb{E} \sum_{i=0}^k d_{i,k-i}(Z, Y) \\
&= \sum_{i=0}^k d_{i,k-i}(X, Y)/\alpha^i \\
&\geq d(Y) + d_{1,k-1}(X, Y)/\alpha - \sum_{i=2}^k 2^{-k^2} \alpha M/\alpha^i \\
&\geq M/\alpha - (k-1)2^{-k^2} M/\alpha \\
&\geq 2^{-3k^2} M/\alpha.
\end{aligned}$$

Since some set Z must achieve this bound, we obtain the desired bound on $\text{disc}^+(G)$. \square

We can now prove the main theorem of this section.

Proof of Theorem 8. As in the proof of Theorem 1, we may assume that $\text{disc}^+(G) \leq \text{disc}^-(G)$. If $\text{disc}^+(G) \geq 2^{-7k^2} n^{(k+1)/2}$ we are done. Otherwise, suppose $\text{disc}^+(G) = 2^{-7k^2} n^{(k+1)/2}/\alpha$ for some $\alpha > 1$: we shall show $\text{disc}^-(G) \geq 2^{-7k^2} \alpha n^{(k+1)/2}$.

Note first that for disjoint sets $X, Y \subset V(G)$, if

$$d_{1,k-1}(X, Y) + \alpha d(Y) \geq 2^{-4k^2} n^{(k+1)/2} \quad (14)$$

then we are done by Lemma 11. It is therefore enough to find disjoint X, Y satisfying (14).

Let $V(G) = X_k \cup W_{k-1}$ be a random bipartition and let $W_{k-1} = X_{k-1} \cup W_{k-2}, \dots, W_2 = X_2 \cup W_1$ be random bipartitions where, as usual, in each bipartition each vertex is assigned independently to either vertex class with probability $1/2$. We define weightings w_i on the i -sets in W_i for each i by

$$w_i(K) = d_{i,1,\dots,1}(K, X_{i+1}, \dots, X_k). \quad (15)$$

Let $W_k = V(G)$ and define $w_k = w$. Then for $1 \leq i < k$ and $K \in W_i^{(i)}$,

$$\begin{aligned}
w_i(K) &= d_{i,1,\dots,1}(K, X_{i+1}, \dots, X_k) \\
&= \sum_{x \in X_{i+1}} d_{i+1,1,\dots,1}(K \cup \{x\}, X_{i+2}, \dots, X_k) \\
&= \sum_{x \in X_{i+1}} w_{i+1}(K \cup \{x\}).
\end{aligned}$$

It therefore follows from Lemma 10 that given W_{i+1} and w_{i+1} ,

$$\mathbb{E} \sum_{K \in W_i^{(i)}} |w_i(K)| \geq (i+1)2^{-(i+1)} \sum_{L \in W_{i+1}^{(i+1)}} |w_{i+1}(L)|/\sqrt{2n}. \quad (16)$$

It follows that

$$\begin{aligned} \mathbb{E} \sum_{x \in W_1} |d_{1,\dots,1}(\{x\}, X_2, \dots, X_k)| &= \mathbb{E} \sum_{x \in W_1} |w_1(x)| \\ &\geq k!2^{-\binom{k+1}{2}} \sum_{K \in W_k^{(k)}} |w_k(K)|/(\sqrt{2n})^{k-1} \\ &= k!2^{-\binom{k+1}{2}} \binom{n}{k}/(2n)^{(k-1)/2}. \end{aligned}$$

Let $X_1^+ = \{x \in W_1 : d_{1,\dots,1}(x, X_2, \dots, X_k) > 0\}$. Then, as in (11),

$$\mathbb{E} d_{1,\dots,1}(X_1^+, X_2, \dots, X_k) \geq \frac{1}{2} k!2^{-\binom{k+1}{2}} \binom{n}{k}/(2n)^{(k-1)/2} \geq 2^{-2k^2} n^{(k+1)/2}. \quad (17)$$

We partition the edges in $V_0 = X_1^+ \cup \bigcup_{i=2}^k X_i$ that meet X_1^+ in exactly one vertex as follows. For a nonempty $S \subset \{2, \dots, k\}$, let $V_S = \bigcup_{i \in S} X_i$ and $E_S = \{K \cup \{x\} : x \in X_1^+, K \in V_S^{(k-1)}, |K \cap X_i| > 0 \forall i \in S\}$. Let $d_S = \sum_{K \in E_S} w(K)$ and note that $d_{1,k-1}(X_1^+, V_S) = \sum_{\emptyset \neq T \subset S} d_T$ and $d_{\{2,\dots,k\}} = d_{1,\dots,1}(X_1^+, X_2, \dots, X_k)$. Let S_0 be minimal with $|d_{S_0}| \geq (2k)^{-k+|S_0|} d_{\{2,\dots,k\}}$. Then

$$\begin{aligned} \max_{S \subset \{2,\dots,k\}} |d_{1,k-1}(X_1^+, S)| &\geq |d_{1,k-1}(X_1^+, V_{S_0})| \\ &\geq |d_{S_0}| - \sum_{\emptyset \neq T \subsetneq S_0} |d_T| \\ &\geq ((2k)^{-k+|S_0|} - \sum_{i=1}^{|S_0|-1} k^{|S_0|-i} (2k)^{-k+i}) d_{\{2,\dots,k\}} \\ &\geq d_{\{2,\dots,k\}}/2(2k)^{k-1} \\ &\geq 2^{-k^2} d_{\{2,\dots,k\}}. \end{aligned}$$

Thus it follows from (17) that

$$\mathbb{E} \max_{S \subset \{2,\dots,k\}} |d_{1,k-1}(X_1^+, V_S)| \geq 2^{-3k^2} n^{(k+1)/2}$$

and so there is some $S \subset \{2, \dots, k\}$ with

$$\mathbb{E}|d_{1,k-1}(X_1^+, V_S)| \geq 2^{-3k^2} n^{(k+1)/2} / 2^k.$$

Now let $Y = V_S$ and $X_S^+ = \{x \in W_1 : d_{1,k-1}(\{x\}, V_S) > 0\}$. Then, since $\mathbb{E} d_{1,k-1}(W_1, V_S) = 0$, we have

$$\mathbb{E} d_{1,k-1}(X_S^+, V_S) \geq 2^{-3k^2} n^{(k+1)/2} / 2^{k+1} \geq 2^{-4k^2} n^{(k+1)/2}.$$

Finally, since $\mathbb{E} d(V_S) = 0$, we have

$$\mathbb{E} d_{1,k-1}(X_S^+, V_S) + \alpha d(V_S) \geq 2^{-4k^2} n^{(k+1)/2}.$$

It follows that there are sets X, Y satisfying (14). \square

We note that Theorem 8 implies the following bound on $\text{disc}_p^+(G)\text{disc}_p^-(G)$ for unweighted hypergraphs G .

Corollary 12. *Let G be a k -uniform hypergraph with n vertices and $p\binom{n}{k}$ edges. Then*

$$\text{disc}_p^+(G)\text{disc}_p^-(G) \geq 2^{-14k^2+2} p^2 (1-p)^2 n^{k+1}.$$

Proof. The result is trivial if $p = 0$ or $p = 1$. Otherwise, let H be the complete k -uniform hypergraph on the same vertex set as G with edge-weighting w defined by $w(e) = 1/2p$ if $e \in E(G)$ and $w(e) = -1/2(1-p)$ otherwise. Then $\sum w(e) = 0$ and $\sum |w(e)| = \binom{n}{k}$, and so, by Theorem 8,

$$\text{disc}^+(H)\text{disc}^-(H) \geq 2^{-14k^2} n^{k+1}.$$

Now for $Y \subset V(G)$,

$$\begin{aligned} d_p^{(G)}(Y) &= e(Y) - p \binom{|Y|}{2} \\ &= \sum_{K \in Y^{(k)}} (\mathbf{1}_{K \in E(G)} - p) \\ &= \sum_{K \in Y^{(k)}} 2p(1-p)w(K) \\ &= 2p(1-p)d^{(H)}(Y). \end{aligned}$$

Thus

$$\text{disc}_p^+(G)\text{disc}_p^-(G) = 4p^2(1-p)^2 \text{disc}^+(H)\text{disc}^-(H), \quad (18)$$

which implies the required bound. \square

We can, however, improve upon the $p^2(1-p)^2$ term in Corollary 12 (at the cost of a slightly worse constant) to obtain a bound similar to that in Theorem 1. First, however, we need a version of Lemma 7 for unweighted hypergraphs.

Lemma 13. *Let G be a k -uniform hypergraph of order n with $p\binom{n}{k}$ edges, where $p(1-p) \geq 1/n$ and $n \geq 2k$. Let $V(G) = X \cup Y$ be a random bipartition. Then*

$$\mathbb{E} \sum_{K \in X^{(k-1)}} |d_{k-1,1}(K, Y) - p|Y|| \geq 2^{-2k^2} \sqrt{p(1-p)} n^{k-\frac{1}{2}}.$$

Proof. We follow the argument of Lemma 7. As before, we may assume $p \leq 1/2$. Let $V = V(G) = X \cup Y$ be a random bipartition. For $K \in V^{(k-1)}$, let $d(K)$ be the number of edges of G containing K and define $r(K)$ by $d(K) = p(n-k+1) + r(K)$. Let $d = p(n-k+1)$. Then, as in Lemma 7,

$$\mathbb{E} |d_{k-1,1}(K, Y \setminus K) - p|Y \setminus K|| \geq \frac{1}{2} \max\{|r(K)|, (1-p)\sqrt{d(K)/2}\}.$$

For $K \in V^{(k-1)}$, we define $I(K) = 1$ if $K \subset X$ and $I(K) = 0$ otherwise. Then $I(K)$ and $d_{k-1,1}(K, Y \setminus K)$ are independent random variables, so

$$\begin{aligned} \mathbb{E} \sum_{K \in X^{(k-1)}} |d_{k-1,1}(K, Y) - p|Y|| &= \mathbb{E} \sum_{K \in V^{(k-1)}} I(K) |d_{k-1,1}(K, Y \setminus K) - p|Y \setminus K|| \\ &= 2^{-k+1} \sum_{K \in V^{(k-1)}} \mathbb{E} |d_{k-1,1}(K, Y \setminus K) - p|Y \setminus K|| \\ &\geq 2^{-k} \max\left\{ \sum_{K \in V^{(k-1)}} |r(K)|, \sum_{K \in V^{(k-1)}} (1-p)\sqrt{d(K)/2} \right\} \\ &\geq 2^{-(k+1)} \sum_{K \in V^{(k-1)}} |r(K)| + (1-p)\sqrt{d(K)/2}. \end{aligned}$$

Since $|r(K)| + (1-p)\sqrt{d(K)/2}$ is minimized when $r(K) = 0$ and $d(K) = p(n-k+1)$,

$$\begin{aligned} \mathbb{E} \sum_{K \in X^{(k-1)}} |d_{k-1,1}(K, Y) - p|Y|| &\geq \binom{n}{k-1} 2^{-(k+1)} (1-p) \sqrt{p(n-k+1)/2} \\ &> 2^{-2k^2} \sqrt{p(1-p)} n^{k-\frac{1}{2}}. \end{aligned}$$

□

Theorem 14. *Let G be a k -uniform hypergraph of order n with $p\binom{n}{k}$ edges, where $p(1-p) \geq 1/n$. Then*

$$\text{disc}_p^+(G)\text{disc}_p^-(G) \geq 2^{-18k^2} p(1-p)n^{k+1}.$$

Proof. Let H be the complete k -uniform hypergraph on $V(G)$ with weighting $w(e) = 1-p$ if $e \in E(G)$ and $w(e) = -p$ otherwise. Then $\text{disc}^+(H) = \text{disc}_p^+(G)$ and $\text{disc}^-(H) = \text{disc}_p^-(G)$. Note that $w(H) = 0$. As usual we may assume $p \leq 1/2$ and $\text{disc}^-(H) \geq \text{disc}^+(H) = 2^{-9k^2} \sqrt{p(1-p)}n^{(k+1)/2}/\alpha$. If $\alpha \leq 1$ we are done, so we may assume $\alpha \geq 1$. We will show that $\text{disc}^-(H) \geq 2^{-9k^2} \sqrt{p(1-p)}\alpha n^{(k+1)/2}$. If there are disjoint $X, Y \subset V(H)$ with

$$d_{1,k-1}(X, Y) + \alpha d(Y) \geq 2^{-6k^2} \sqrt{p(1-p)}n^{(k+1)/2} \quad (19)$$

then we are done by Lemma 11. Thus it is enough to find disjoint X, Y satisfying (19).

As in the proof of Theorem 8, we define random sets $W_k = X \supset W_{k-1} \supset \dots \supset W_1$, where the i -sets in W_i are weighted as in equation (15). Then by Lemma 13,

$$\mathbb{E} \sum_{K \in W_{k-1}^{\binom{k-1}{k-1}}} |w_{k-1}(K)| \geq 2^{-2k^2} \sqrt{p(1-p)}n^{k-\frac{1}{2}}, \quad (20)$$

while W_1, \dots, W_{k-2} satisfy (16). We have

$$\mathbb{E} \sum_{x \in W_1} |w_1(x)| \geq (k-1)! 2^{-\binom{k}{2}} \sum_{K \in W_{k-1}^{\binom{k-1}{k-1}}} |w_{k-1}(K)| / (\sqrt{2n})^{k-2},$$

and so, defining X_1^+ as before, we can replace (17) by

$$\mathbb{E} d_{1,\dots,1}(X_1^+, X_2, \dots, X_k) \geq 2^{-4k^2} \sqrt{p(1-p)}n^{(k+1)/2}. \quad (21)$$

The argument is completed as before (with all bounds changed by a factor $2^{-2k^2} \sqrt{p(1-p)}$). □

The following corollary is immediate.

Corollary 15. *Let G be a k -uniform hypergraph of order n with $p\binom{n}{k}$ edges, where $p(1-p) \geq 1/n$. Then*

$$\text{disc}_p(G) \geq 2^{-9k^2} \sqrt{p(1-p)}n^{(k+1)/2}.$$

We note that Corollaries 2 and 15 are best possible up to the value of the constant 2^{-9k^2} . To see this, let $G \in \mathcal{G}^{(k)}(n, p)$ be a random k -uniform hypergraph, where each possible edge is present independently with probability p , and let $S \subset V(G)$. Let $N = \binom{n}{k}$ and $h = (1 + \epsilon)k!^{-1/2} \sqrt{2p(1-p) \ln 2n^{(k+1)/2}}$. Then by standard bounds on the tail of the binomial distribution (see [2], Theorem 1.3), provided $p(1-p) \geq c_k n^{1-k}$, for any subset S of $V(G)$ we have

$$\mathbb{P}(|d_p(S)| \geq h) \leq \mathbb{P}(|B(N, p) - Np| \geq h) < 2^{-n}$$

for sufficiently large n . Thus there is some k -uniform hypergraph G of order n with $\text{disc}_p(G) \leq h$.

Let us also note that the gain from $p^2(1-p)^2$ to $p(1-p)$ between Corollary 12 and Theorem 14 comes because a “typical” vertex in G has degree $p \binom{n-1}{k-1}$: so if p is small, then the weight around a typical vertex is concentrated in fairly few edges. We remark that no similar bound is possible for the larger class of k -uniform hypergraphs with $\sum |w(e)| = \binom{n}{k}$ such that $\sum \max\{w(e), 0\} = p \binom{n}{k}$: consider a random k -uniform hypergraph $H \in \mathcal{G}^{(k)}(n, 1/2)$, and let G be the weighted hypergraph obtained by giving each edge weight $2p$ and each non-edge weight $-2(1-p)$. Then if $e(H) = \frac{1}{2} \binom{n}{k}$ (which happens with probability at least $c'_k n^{-k/2}$ if $\binom{n}{k}$ is even) we have $\sum |w(e)| = \binom{n}{k}$ and $\sum \max\{w(e), 0\} = p \binom{n}{k}$. On the other hand, it follows from (18) that $\text{disc}_p^+(G) \text{disc}_p^-(G) = 4p^2(1-p)^2 \text{disc}^+(H) \text{disc}^-(H)$, while $\text{disc}^+(H)$ and $\text{disc}^-(H)$ are both $O(n^{(k+1)/2})$ with exponentially small failure probability.

It is interesting to ask about the range in which Theorem 1 and Theorem 14 are sharp (up to the constant). For instance, in the case of graphs the remarks above show that $\text{disc}_p^+(G)$ and $\text{disc}_p^-(G)$ can both be around $c\sqrt{p(1-p)}n^{3/2}$. When p is (about) $1/2$, the complete bipartite graph and its complement show that we can have discrepancy $O(n)$ on one side (and cn^2 on the other). Thus Theorem 1 is sharp in the middle of the scale from cn to $c'n^2$, and (for $p = 1/2$) is sharp at the ends. How sharp is it at other parts of the scale, or at the ends when $p \neq 1/2$?

The constant in Theorem 14 is clearly not best possible. A more careful version of the argument should improve it to $2^{-ck \ln k}$; it would be of interest to know the correct order of magnitude. It would also be interesting to know what happens in the range $n^{1-k} \leq p \leq 1/n$.

4 Subgraph discrepancy

In previous sections we have been concerned with the discrepancy of subgraphs or, equivalently, 2-colourings of the complete graph. We begin this section by considering 2-colourings of an arbitrary graph: questions of this form were raised by Sós in [11].

For a k -uniform hypergraph G , a subgraph H of G and a real number $p \in [0, 1]$, we define

$$\text{disc}_p^+(H, G) = \max_{S \subset V(G)} e(H[S]) - pe(G[S])$$

and

$$\text{disc}_p^-(H, G) = \max_{S \subset V(G)} pe(G[S]) - e(H[S]).$$

Note that if G is the complete k -uniform hypergraph then these two definitions agree with (1) and (2). We set

$$\text{disc}_p(H, G) = \max\{\text{disc}_p^+(H, G), \text{disc}_p^-(H, G)\}.$$

We begin with a fairly straightforward analogue to Theorem 8. Note that arguing as in Corollary 12 gives a bound with $p^2(1-p)^2$ in place of $p(1-p)$.

Theorem 16. *Let G be a k -uniform hypergraph with n vertices and m edges, and let H be a subgraph of G with pm edges, where $p(1-p) \geq 1/n$. Then*

$$\text{disc}_p^+(H, G)\text{disc}_p^-(H, G) \geq 2^{-18k^2} p(1-p)m^2/n^{k-1}.$$

We first need a version of Lemma 13.

Lemma 17. *Let G be a k -uniform hypergraph with n vertices and m edges, and let $H \subset G$ be a subhypergraph of G with pm edges, where $p(1-p) \geq 1/n$. Let $V(G) = X \cup Y$ be a random bipartition. Then*

$$\mathbb{E} \sum_{K \in X^{(k-1)}} |d_{k-1,1}^{(H)}(K, Y) - pd_{k-1,1}^{(G)}(K, Y)| \geq 2^{-(k+1)} \sqrt{p(1-p)m}/\sqrt{n}.$$

Proof. For a partition $V(G) = X \cup Y$, let us write

$$f(X, Y) = \sum_{K \in X^{(k-1)}} |d_{k-1,1}^{(H)}(K, Y) - pd_{k-1,1}^{(G)}(K, Y)|.$$

As in Lemma 13, we may assume that $p \leq 1/2$ or else replace H by its complement in G . For $K \in V^{(k-1)}$, let $d_H(K)$ be the number of edges of H containing K and let $d_G(K)$ be the number of edges of G containing K . Define $r(K)$ by $d_H(K) = pd_G(K) + r(K)$. Then, as in Lemma 13,

$$\mathbb{E}|d_{k-1,1}^{(H)}(K, Y \setminus K) - pd_{k-1,1}^{(G)}(K, Y)| \geq \frac{1}{2} \max\{|r(K)|, (1-p)\sqrt{d_H(K)/2}\}.$$

Thus

$$\mathbb{E}f(X, Y) \geq 2^{-(k+1)} \sum_{K \in V^{(k-1)}} |r(K)| + (1-p)\sqrt{d_H(K)/2}.$$

Now $|r(K)| + (1-p)\sqrt{d_H(K)/2}$ is minimized when $r(K) = 0$ and so $d_H(K) = pd_G(K)$. Thus

$$\begin{aligned} \mathbb{E}f(X, Y) &\geq 2^{-(k+1)} \sum_{K \in V^{(k-1)}} (1-p)\sqrt{pd_G(K)/2} \\ &\geq 2^{-(k+1)} \sum_{K \in V^{(k-1)}} (1-p)d_G(K)\sqrt{p/2n} \end{aligned}$$

since $d_G(K) < n$. Now $\sum_{K \in V^{(k-1)}} d_G(K) = km$, so

$$\begin{aligned} \mathbb{E}f(X, Y) &\geq 2^{-(k+1)} km(1-p)\sqrt{p/2n} \\ &\geq 2^{-(k+1)} \sqrt{p(1-p)}m/\sqrt{n}. \end{aligned}$$

□

Theorem 16 now follows by a modification of the proof of Theorem 14.

Proof of Theorem 16. Let $V = V(G)$. We may assume $p \leq 1/2$ or replace H by its complement in G . We define, as in Theorem 14, an edge-weighting w on $V^{(k)}$ by $w(K) = 1-p$ if $K \in E(H)$, $w(K) = -p$ if $K \in E(G) \setminus E(H)$ and $w(K) = 0$ otherwise. Note that then $w(V) = 0$. We may assume $\text{disc}_p^-(H, G) \geq \text{disc}_p^+(H, G) = 2^{-9k^2} \sqrt{p(1-p)}e(G)/\sqrt{n}^{k-1}\alpha$. If $\alpha \leq 1$ we are done, so we may assume $\alpha \geq 1$. If there are disjoint X, Y with

$$d_{1,k-1}(X, Y) + \alpha d(Y) \geq 2^{-6k^2} \sqrt{p(1-p)}m/n^{(k-1)/2},$$

then we are done as before by Lemma 11. Once again, we define random subsets $W_k = X \supset W_{k-1} \supset \cdots \supset W_1$. Applying Lemma 17 instead of

Lemma 13 to W_{k-1} , we can replace (20) by

$$\mathbb{E} \sum_{K \in W_{k-1}^{(k-1)}} |w_{k-1}(K)| \geq 2^{-2k^2} \sqrt{p(1-p)m}/\sqrt{n}. \quad (22)$$

As before, W_1, \dots, W_{k-2} satisfy (16); applying this $k-2$ times to (22), we see that (instead of (21)) we obtain

$$\mathbb{E}d(X_1^+, X_2, \dots, X_n) \geq 2^{-4k^2} \sqrt{p(1-p)m}/n^{(k-1)/2},$$

and the argument is completed as before. \square

Corollary 18. *Let G be a k -uniform hypergraph with n vertices and m edges, and H a subgraph of G with pm edges, where $p(1-p) \geq 1/n$. Then*

$$\text{disc}_p(H, G) \geq 2^{-9k^2} \sqrt{p(1-p)m}/n^{(k-1)/2}.$$

We obtain stronger results when there is a restriction on the maximum overlap between edges of positive and negative weights.

Theorem 19. *Let G be a complete k -uniform hypergraph of order n with edge-weighting w . Suppose in addition that, for some $1 \leq s \leq r$, if $w(e) > 0$ and $w(e') < 0$ then $|e \cap e'| < s$. Let $M = \sum |w(e)|$ and $m = \sum w(e)$. If $m = (2p-1)M$, where $p(1-p) \geq 1/n$, then*

$$\text{disc}_p^+(G)\text{disc}_p^-(G) \geq 2^{-18k^2} p^2(1-p)^2 M^2/n^{s-1}.$$

Proof. Suppose first that $p = 1/2$, and let $E = \{e : w(e) \neq 0\}$. As in the proof of Theorem 8, we may assume $\text{disc}^+(G) \leq \text{disc}^-(G)$. Suppose $\text{disc}^+(G) = 2^{-9k^2} e(H)/n^{(s-1)/2} \alpha$, where $\alpha \geq 1$. If there are disjoint X, Y with

$$d_{1,k-1}(X, Y) + \alpha d(Y) \geq 2^{-6k^2} e(H)/n^{(s-1)/2}$$

then we are done by Lemma 11. Otherwise, define W_i, X_i and w_i as before, and consider W_s and w_s . Since $w(e) > 0$ and $w(e') < 0$ implies $|e \cap e'| < s$, we have, for $K \in W_s^{(s)}$,

$$|w_s(K)| = \sum_{e \cap W_s = K, |e \cap X_i| = 1 \forall i > s} |w(e)|$$

and so

$$\mathbb{E} \sum_{K \in W_s^{(s)}} |w_s(K)| = \mathbb{E} \sum_{|e \cap W_s| = s, |e \cap W_i| = 1 \forall i > s} |w(e)|.$$

Let A_e be the event that $|e \cap W_s| = s$ and $|e \cap W_i| = 1$ for all $i > s$. Then $\mathbb{P}A_e > 2^{-k^2}$ and so $\mathbb{E} \sum_{K \in W_s^{(s)}} |w_s(K)| \geq 2^{-k^2} M$. Applying Lemma 10 as in (16), we obtain that

$$\begin{aligned} \mathbb{E} \sum_{x \in W_1} |d_{1, \dots, 1}(\{x\}, X_2, \dots, X_k)| &\geq \mathbb{E} \sum_{K \in W_s^{(s)}} |w_S(K)| / (\sqrt{2n})^{s-1} \\ &\geq 2^{-2k^2} M / n^{(s-1)/2}. \end{aligned}$$

The rest of the argument follows as in the proof of Theorem 8.

Now suppose $p \neq 1/2$. As in the proof of Corollary 12, we multiply all positive edge-weights by $1/2p$ and all negative edge-weights by $-1/2(1-p)$ to obtain a new edge-weighting w' . The result follows immediately. \square

As an application of Theorem 19, let us consider the complete subgraphs of a graph and its complement. For $t \geq 2$ and a graph G , we write $k_t(G)$ for the number of copies of K_t of G . We write

$$\text{disc}_{K_k}(G) = \max_{S \subset V(G)} |k_k(G[S]) - k_k(\overline{G}[S])|.$$

For instance, $\text{disc}_{K_2}(G)$ is just $\text{disc}(G)$. Clearly, complete subgraphs of G meet complete subgraphs of its complement in at most one vertex: applying Theorem 19 to the k -uniform hypergraph of complete or independent k -sets gives the following result.

Corollary 20. *For every graph G of order n ,*

$$\text{disc}_{K_k}(G) \geq c_k n^{k-\frac{1}{2}}.$$

For instance, in some subset S ,

$$|k_3(G[S]) - k_3(\overline{G}[S])| \geq cn^{5/2}.$$

Considering random graphs shows that this result is best possible up to the constant. A similar approach yields results in some cases for $\text{disc}_H(G)$ where H is not a complete graph (and disc_H is defined in the obvious way). It would be interesting to determine the correct order of magnitude of disc_H for all graphs H . When H is fairly dense, so that copies of H and \overline{H} cannot overlap very much, we obtain a lower bound on $\text{disc}_H(G)$ using Theorem 19. However, when H is sparse this gives a much weaker bound; for instance, what can we say when H is a tree?

References

- [1] J. Beck and V. T. Sós, Discrepancy theory, in *Handbook of Combinatorics*, Vol. 2, 1405–1446, Elsevier, Amsterdam, 1995
- [2] B. Bollobás, *Random Graphs*, Second Edition, Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2001, xviii+498pp
- [3] P. Borwein and T. Erdélyi, *Polynomials and Polynomial Inequalities*, Graduate Texts in Mathematics, 161, Springer-Verlag, New York, 1995, x+480 pp.
- [4] Tamás Erdélyi, Personal communication
- [5] P. Erdős, R. Faudree, C. Rousseau and R. Schelp, A local density condition for triangles, *Discrete Math.* **127** (1994), 153–161
- [6] P. Erdős, M. Goldberg, J. Pach and J. Spencer, Cutting a graph into two dissimilar halves, *J. Graph Theory* **12** (1988), 121–131
- [7] P. Erdős and J. Spencer, Imbalances in k -Colorations, *Networks* **1** (1971/2), 379–385
- [8] P. Keevash and B. Sudakov, Local density in graphs with forbidden subgraphs, *to appear*
- [9] M. Krivelevich, On the edge distribution in triangle-free graphs, *J. Combinatorial Theory, Ser. B* **63** (1995), 245–260
- [10] J. E. Littlewood, On bounded bilinear forms in an infinite number of variables, *Quart. J. Math. Oxford* **1** (1930), 164–174
- [11] V. T. Sós, Irregularities of partitions: Ramsey theory, uniform distribution, in *Surveys in Combinatorics* (Southampton, 1983), 201–246, London Math. Soc. Lecture Note Ser., 82, Cambridge Univ. Press, Cambridge-New York, 1983
- [12] S. J. Szarek, On the best constants in the Khinchin inequality, *Studia Math.* **58** (1976), 197–208
- [13] B. Tomaszewski, A simple and elementary proof of the Khintchine inequality with the best constant, *Bull. Sci. Math.* **111** (1987), 103–109