Discrepancy of graphs and hypergraphs

Alex Scott,
University of Oxford

July 2013

Introduction

We begin with a natural question:

- How uniformly distributed can the edges of a graph be?

Questions of this type for different mathematical structures arise in various areas, especially number theory and combinatorics.

The problem for graphs and hypergraphs was introduced 40 years ago by Erdős and Spencer.

Outline

Discrepancy of a graph

Let G be a graph of order n and density p. The *discrepancy* of G is

$$\mathrm{disc}(G) := \max_{Y \subset V(G)} \left| e(Y) - p\binom{|Y|}{2} \right|.$$

The discrepancy measures how irregularly the edges are distributed among the vertices.

How small can the discrepancy be?

$p = 1/2$

Theorem (Erdős and Spencer, 1971)

Let G be a graph of order n and density $1/2$. Then

$$\mathrm{disc}(G) \geq cn^{3/2}.$$

This bound is optimal up to the constant.

Erdős, Goldberg, Pach and Spencer (1988) later extended this to other densities.

Upper bound ($p = 1/2$)

Consider a random graph $G \in \mathcal{G}(n, 1/2)$.

For any $Y \subset V(G)$ and $h > 0$ we have, by a Chernoff bound,

$$\mathbb{P}(|e(Y) - \mathbb{E}e(Y)| > h) < e^{-ch^2/|Y|^2} \leq e^{-ch^2/n^2}.$$

There are 2^n choices for Y, so if we take $h > c'n^{3/2}$ (for a suitable c') we have $\mathbb{P}(\mathrm{disc}(G) > h)$ exponentially small.

Lower bound

It is helpful to work in terms of a weight function w defined by

$$w(xy) = \begin{cases} 1/2 & \text{if } xy \in E(G) \\ -1/2 & \text{otherwise} \end{cases}$$

Note that for $Y \subset V(G)$ we have

$$w(Y) = e(Y) - \frac{1}{2}\binom{|Y|}{2},$$

so showing $\mathrm{disc}(G) \geq cn^{3/2}$ is equivalent to showing there is Y with $|w(Y)| \geq cn^{3/2}$.

Lower bound

Let $V = X \cup Y$ be a random bipartition. For a fixed vertex v, we consider $w(v, Y)$. We have

$$w(v,Y) = \frac{1}{2}|\Gamma(v) \cap Y| - \frac{1}{2}|Y \setminus \Gamma(v) \setminus v|.$$

Thus if v has degree d then $w(v, Y)$ has distribution

$$\frac{1}{2}B(d, 1/2) - \frac{1}{2}B(n-1-d, 1/2).$$

In particular, it is easily shown that, independently of d,

$$\mathbb{E}(|w(v,Y)|) \geq c\sqrt{n}.$$

Lower bound

The event $(v \in X)$ is independent of the random variable $w(v, Y)$, so

$$\mathbb{E}(|w(v, Y)| \cdot 1_{v \in X}) \geq c\sqrt{n}/2.$$

It follows by linearity of expectation that

$$\mathbb{E}(\sum_{v \in X} |w(v, Y)|) \geq cn^{3/2}/2.$$

Let (X, Y) be a partition achieving this bound.

Lower bound

Let $X^+ = \{v \in X : w(x, Y) > 0\}$ and $X^- = X \setminus X^+$. Then

$$w(X^+, Y) - w(X^-, Y) = \sum_{v \in X} |w(v, Y)| \geq cn^{3/2}/2.$$

It follows that

$$\max\{|w(X^+, Y)|, |w(X^-, Y)|\} \geq cn^{3/2}/4.$$

Without loss of generality, we have

$$w(X^+, Y) \geq cn^{3/2}/4.$$

Lower bound

But

$$w(X^+ \cup Y) = w(X^+, Y) - w(X^+) - w(Y).$$

Since $w(X^+, Y) \geq cn^{3/2}/4$, we have either

$$w(X^+ \cup Y) \geq cn^{3/2}/12,$$

or

$$\min\{w(X^+), w(Y)\} \leq -cn^{3/2}/12.$$

In either case, $\mathrm{disc}(G) = \Omega(n^{3/2})$.

General p

The result of Erdős and Spencer extends to general densities:

Theorem (Erdős, Goldberg, Pach and Spencer, 1988)

Let G be a graph of order n and density p. If $p(1-p) \geq 2/n$ then

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

We therefore know that G contains an 'atypical' induced subgraph, although it could be either denser or sparser than G.

Positive and negative discrepancy

Let G be a graph of order n and density p. The *positive discrepancy* of G is

$$\mathrm{disc}^+(G) := \max_{Y \subset V(G)} e(Y) - p\binom{|Y|}{2}$$

and the *negative discrepancy* is

$$\mathrm{disc}^-(G) := \max_{Y \subset V(G)} p\binom{|Y|}{2} - e(Y).$$

Examples

Random graphs show that we can have both positive and negative discrepancies $O(n^{3/2})$. On the other hand, the complete bipartite graph $K_{n/2,n/2}$ has positive discrepancy $O(n)$, at the cost of a very large negative discrepancy.

G	$\mathrm{disc}^+(G)$	$\mathrm{disc}^-(G)$
random	$n^{3/2}$	$n^{3/2}$
$K_{n/2,n/2}$	n	n^2
$2K_{n/2}$	n^2	n

Is there a trade-off?

Product form

Theorem (Bollobás and Scott, 2006)

Let G be a graph of order n and density p. If $p(1-p) \geq 1/n$ then

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

This is sharp (up to the constant) for the three examples above.

Proof by bullet points

Suppose that $\mathrm{disc}^+(X, Y) = \alpha \cdot c(p)n^{3/2}$, where $\alpha < 1$. We must show that $\mathrm{disc}^+(X, Y) \geq c(p)n^{3/2}/\alpha$.

- Find (X, Y) such that $w(X, Y) > cn^{3/2}$ and $w(Y)$ does not have too large a negative weight. [More precisely: control the linear combination $w(X, Y) + \alpha w(Y)$.]
- Use a random 'amplification' argument (selecting a random subset of Y) to show that $w(X)$ must be large and negative.

Hypergraphs

The definitions above extend naturally to hypergraphs.

Theorem (Erdős and Spencer, 1971)

Let G be a k-uniform hypergraph of order n and density $1/2$. Then

$$\mathrm{disc}(G) \geq cn^{k+1}.$$

This bound is optimal up to the constant for random hypergraphs with density $1/2$.

What about general $p \neq 1/2$? And positive/negative discrepancy?

Hypergraphs

Theorem (Bollobás and Scott, 2006)

Let G be a k-uniform hypergraph of order n and density p. If $p(1-p) \geq 1/n$ then

$$\mathrm{disc}^+(G)\mathrm{disc}^-(G) \geq cp(1-p)n^{k+1}.$$

For the unsigned discrepancy, this implies

$$\mathrm{disc}(G) \geq c\sqrt{p(1-p)}n^{(k+1)/2},$$

extending the Erdős-Spencer result to general densities.

Introduction

Let G and H be two graphs with the same number of vertices. It is natural to ask:

- How much or little can we make G and H overlap by placing them on the same set of vertices?

This will lead us to define the discrepancy of a *pair* of graphs.

Introduction

Two problems that can be expressed in this way by fixing H:

- Given a graph G with density p and $\alpha \in (0,1)$, is there a set of exactly $\alpha|G|$ vertices with large discrepancy?
- Consider bisections $V(G) = X \cup Y$ with $|X| = |Y|$. Ue the resulting cuts to define a notion of *bisection discrepancy*. Erdős, Goldberg, Pach and Spencer (1988) conjectured that reasonably dense graphs of order n should have bisection discrepancy $\Omega(n^{3/2})$.

Discrepancy of two graphs

For G, H with densities p, q, we write

$$\mathrm{disc}^+(G,H) = \max_\pi |E(G_\pi) \cap E(H)| - pq\binom{n}{2}$$

and

$$\mathrm{disc}^-(G,H) = pq\binom{n}{2} - \min_\pi |E(G_\pi) \cap E(H)|,$$

where π ranges over bijections $V(G) \to V(H)$. We set

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

Intersections of graphs

Theorem (Bollobás and Scott, 2011)

Let G and H be graphs of order n with densities p and q respectively. If $p, q \in (16/n, 1 - 16/n)$ then

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

In particular,

$$\mathrm{disc}(G, H) \geq c'(p, q)n^{3/2}.$$

It follows that the bisection discrepancy of G is at least $c'(p)n^{3/2}$, proving the conjecture of Erdős, Goldberg, Pach and Spencer.

A natural conjecture

What happens for hypergraphs? For dense k-uniform hypergraphs G and H, it seems natural to conjecture that

$$\mathrm{disc}(G,H) \geq cn^{(k+1)/2}.$$

This turns out to be completely wrong.

Examples

If we allow weights on our edges then we get the following example.

Example 1: Let $G = K_{n/2,n/2}$, and let H be the union of two complete graphs on $n/2$. We give the edges of H in one copy of $K_{n/2}$ weight 1 and edges in the other copy weight -1. Then

$$\operatorname{disc}(G, H) = 0.$$

Maybe this is just because we allowed weights?

Examples

Example 2: Let $V = X \cup Y$ with $|X| = |Y| = n/2$. Let H be the 3-uniform hypergraph with all edges that meet both X and Y. Let G be a Steiner Triple System (so every pair of vertices is contained in exactly one edge of G).

Then $\mathrm{disc}(G, H) = 0$.

Easy to generalize to an example in which both hypergraphs are dense.

The bad news

Theorem (Bollobás and Scott, 2013+)

For $k \geq 2$, there is a sequence $H_1, \ldots, H_k$ of nontrivial edge-weighted k-uniform hypergraphs such that for all $i \neq j$

$$\mathrm{disc}(H_i, H_j) = 0.$$

The good news

Theorem (Bollobás and Scott, 2013+)

Let $k \geq 2$, and $H_1, \ldots, H_{k+1}$ be a sequence of k-uniform hypergraphs with normalized edge-weightings. Then there are $i \neq j$ such that

$$\mathrm{disc}(H_i, H_j) \geq c_k n^{(k+1)/2}.$$

Random graphs and hypergraphs

Recall that the discrepancy of a random graph has order $\Theta(n^{3/2})$, where the Chernoff bound had to beat 2^{-n}. For a pair of random graphs, there are $n!$ bijections between the vertices, and the Chernoff bound only gives $c\sqrt{n \log n}$.

Ma, Naves and Sudakov (2013++) and Bollobás and Scott (2013++) have shown that this is the correct order. Related results on packing random graphs are proved by Bollobás, Janson and Scott (2013++). [All these results extend to k-uniform hypergraphs.]

Open problems

- Suppose G is a graph with $\mathrm{disc}^+(G)\mathrm{disc}^-(G) = O(n^3)$. What can we say about the structure of G? For instance, must it be close to one of $K_{n/2,n/2}$, $2K_{n/2}$ or a (pseudo)random graph with density $1/2$?
- What can we say about the structure of graphs with small one-sided discrepancy?
- What can be said about sparse hypergraphs ($p < 1/n$)?

Open problems

- For the discrepancy of two graphs, what is the dependence on p and q? We have a bound of form $cp^4q^4(1-p)^4(1-q)^4$, but this is probably far from best.
- What can we say about $\mathrm{disc}(G, H)$, and about signed discrepancies, if G and H are pseudorandom graphs?