Extremal and probabilistic discrete math – expository notes

Asaf Ferber *

1 A brief introduction to extremal combinatorics

These notes serve as a brief introduction to extremal and probabilistic combinatorics. It is an extended version of an expository talk that I gave in the Probability and Combinatorics seminar at UCI, and is intended to non-expert audience in the area (that is, no background is being assumed).

First, let us give some definitions and useful notation.

1.1 Some basic definitions

A graph $G = (V, E)$ is a pair with V being its set of *vertices* and $E \subseteq \binom{V}{2}$ is its set of *edges* (when I write $\binom{V}{2}$ I mean the collection of all unordered pairs of elements of V). A hypergraph is a pair $H = (V, E)$ where V is the set of vertices and the set of *edges* (or *hyperedges* to be more precise) $E \subseteq 2^V$ is any collection of subsets of V (that is, H is a *set system*). A hypergraph $H = (V, E)$ is said to be k-*uniform* if all its edges are of size k. That is, if $E \subseteq \binom{V}{k}$.

A (hyper)graph $G = (V, E)$ is said to be t-partite if there exists a partition of its vertices into t *parts*, $V = A_1 \cup \ldots \cup A_t$, such that for every edge $e \in E$ and every $1 \leq i \leq t$ we have $|e \cap A_i| \leq 1$.

A subset $X \subseteq V$ is said to be *independent* if no edge $e \in E$ is fully contained in X. In other words, let $H[X]$ be the subhypergraph of H *induced by* X (that is, its vertex-set if X and its edge-set consists of all the edges $e \in E$ for which $e \subseteq X$). The set X is *independent* if $H[X]$ has no edges.

A 2-partite graph G is referred to as a *bipartite graph*. The *chromatic number* of a graph G, denoted by $\chi(G)$, is define as the smallest t for which G is t-partite. In particular, let $V = A_1 \cup \ldots \cup A_t$ be a partition of V witnessing on G being t-partite, one can assign each $v \in V$ the unique color $1 \leq i \leq t$ for which $v \in A_i$. Then, each A_i is the the corresponding *color class* i and is an independent set. With this notation, one can define $\chi(G)$ as the minimum t for which there exists a function $f : V \to \{1, \ldots, t\}$ such that all sets $f^{-1}(\{i\})$ are independent.

A k-uniform hypergraph $H = (V, E)$ is said to be *complete* (or a *clique*) if $E = \binom{V}{k}$. In particular, if $|V| = n$, then since there is a unique complete k-uniform hypergraph on n vertices (up to isomorphism), we denote it by $K_n^{(k)}$. In case $k = 2$ (that is, H is a graph) we simply write K_n.

A bipartite graph with parts of sizes s and t, respectively, which contains all the st possible edges is said to be a *complete bipartite graph*, and is denoted by $K_{s,t}$.

A *cycle* of length n, denoted by C_n, is a graph on n vertices for which there exists a labeling of its vertices $V = \{v_1, \ldots, v_n\}$ in such a way that is edge set is $E := \{v_i v_{i+1} \mid 1 \leq i \leq n\}$ (with the rule $v_{n+1} = v_1$).

1.2 Typical problems

Extremal problems in discrete math are in general problems which consider some discrete structure which is *extreme* with respect to some properties. Among many other examples, these are problems of the following flavor:

*Department of Mathematics, University of California, Irvine. Email: asaff@uci.edu.

- **Combinatorial number theory**
  - What is the maximum size of a subset $A \subseteq [n]$ for which there is no solution to $a - a' \in PRIMES - 1$? (the -1 is just to avoid triviality. For example, if the restriction was $a - a' \notin PRIMES$ then we could take, for example, all numbers which are divisible by 4 and obtain a linear lower bound).
  - More generally, given a subset $S \subseteq [n]$, what is the maximum size of a subset $A \subseteq [n]$ for which there is no solution to $a - a' \in S$? If S is "complicated enough", then the answer is most likely unknown.
  - What is the maximum size of a subset $A \subseteq [n]$ with a k-AP? (a bound of the form $o(n)$ was obtained by Szemeredi where he introduced his famous Regularity Lemma which we will talk about in other notes...)
- **Discrete geometry** What is the maximum number of times that the unit distance can occur among n points in the plane?
- **Algebra** What is the size of the largest set of polynomials with k variables over $\mathbb{Z}_p$ for which every t of them have at most $t-1$ common zeroes?
- **Graph theory**
  - What is the maximum number of edges on an n vertex graph without a copy of K_3? without a copy of a fixed graph H? (these number are denoted by $\mathrm{ex}(n, H)$)?
  - What is the minimum degree of a graph G on n vertices that enforces a *perfect matching*? (a perfect matching in a k-uniform hypergraph on n vertices is a collection of $\lfloor \frac{n}{k} \rfloor$ disjoint edges).
- **Ramsey theory**
  - What is the size of the largest monochromatic clique one can find in any coloring of the edges of K_n using two colors?
  - For a set of numbers T, let

    $$\sum T := \{t_1 + \ldots + t_\ell \mid \ell \in \mathbb{N},\, t_i \in T, \text{ and } t_i \neq t_j \text{ for all } i \neq j\}.$$

    Given any two coloring of the integers $[n]$, what is the largest subset S for which $\sum S \cap [n]$ is monochromatic?

Note that most these problems can be modeled in the language of graphs/hypergraphs where the goal is to find a largest independent set. For example, consider the first problem. One can define a graph $G = (V, E)$ with $V = [n]$ (where $[n] := \{1, \ldots, n\}$), and its edge set E consists of all pairs $xy \in \binom{V}{2}$ for which $|x - y| \in PRIMES - 1$. Clearly, the problem is to figure out what is the size of the largest independent set in G (or at least its order of magnitude as a function of n, where n goes to infinity).

This framework lead to the development of powerful tools such as the Szemeredi Regularity Lemma [6], and the Hypergraphs Containers' method (by Balogh, Morris and Samotij [1], and independently by Saxton and Thomason [5]). Again, these two methods will be discussed in details in future notes.

1.3 Turán numbers

In this section we discuss in a bit more details the problem of finding the extremal number of a given (hyper)graph H. Namely, in understanding the asymptotic behavior of the function $\mathrm{ex}(n, H)$, as n goes to infinity.

The following theorem was proven by Turán in 1941 [7] and is a cornerstone in the area:

Theorem 1 (Turán's Theorem). $ex(n, K_{k+1}) = \left(1 - \frac{1}{k}\right)\binom{n}{2} + o(n^2)$.

To obtain the lower bound in Turán's theorem is simple: suppose that n is divisible by k, and partition $[n]$ into k sets of the same size $\frac{n}{k}$. Now let G be the complete k-partite graph on this partition, and clearly, it contains no cliques of size $k+1$. Since between any two parts we have exactly $\frac{n^2}{k^2}$ many edges, and since there are exactly $\binom{k}{2}$ many distinct pairs of parts, the number of edges in G is

$$\binom{k}{2}\frac{n^2}{k^2} = \left(1 - \frac{1}{k}\right)\binom{n}{2} + o(n^2).$$

More generally, Erdős and Stone [3] proved in the 1960s that

Theorem 2. *For every fixed graph H we have $ex(n, H) = \left(1 - \frac{1}{\chi(H)-1}\right)\binom{n}{2} + o(n^2)$.*

Note that for bipartite graphs the above bound is far from being tight as it only gives a bound of the form $o(n^2)$ and not the exact asymptotic. In fact, for bipartite graphs the extremal number is unknown even for simple graphs like $K_{4,4}$ and C_8. There is now a whole area of trying to seek for Turán numbers for bipartite graphs (the "degenerate" case). For more details you can skim through the very nice survey of Füredi and Simonovits [4].

One of the most famous conjectures about the asymptotic behavior of the extremal number of bipartite graphs, due to Erdős and Simonovits (see [2]), is

Conjecture 1. *For every nonempty bipartite graph H, there exists a rational number $\alpha \in [1, 2)$ and $c > 0$ such that*

$$\frac{ex(n, H)}{n^\alpha} \to c.$$

They also made the following conjecture (see [2]), which can be viewed as an "inverse" version of Conjecture 1:

Conjecture 2. *For every rational number $\alpha \in (1, 2)$ there exists a graph H such that $ex(n, H) = \Theta(n^\alpha)$.*

It is quite amazing that even for very simple bipartite graphs H, the asymptotic behavior of $\mathrm{ex}(n, H)$ is unknown. Apparently, for hyergraphs even much less is known. For example, consider the complete 3-uniform hypergraph on 4 vertices, $K_4^{(3)}$. Erdős offered a \$500 prize for proving

$$\mathrm{ex}(n, K_4^{(3)}) = \frac{5}{9}\binom{n}{3}.$$

As a last remark for this section, we give a proof of basically (up to some factor of $\log^C n$) the best known lower bound for $\mathrm{ex}(n, K_{s,t})$ for general s and t. The proof is probabilistic: namely, let $p \in [0, 1]$ to be determined later. Sample a graph G on n vertices where each pair $e \in \binom{[n]}{2}$ is being added to E with probability p, independently at random. Let Y be a random variable counting the number of edges in G. Clearly, we have

$$\mathbb{E}[Y] = \binom{n}{2}p \approx \frac{n^2 p}{2}.$$

Moreover, let X be a random variable counting the number of $K_{s,t}$s in G. Obviously, we have

$$\mathbb{E}[X] \leq n^{s+t}p^{st}.$$

Suppose that G is the sampled graph. In order to turn G into a $K_{s,t}$-free graph G', we can delete one edge from each copy of $K_{s,t}$ (if exists). Our goal is to show that there exists a graph G for which the corresponding $K_{s,t}$-graph G' has as many edges as possible. To this end, observe that the expected number of edges in G' is at least

$$\mathbb{E}[Y - X] \geq \frac{n^2 p}{2} - n^{s+t} p^{st}.$$

Therefore, if we choose p such that

$$n^{s+t} p^{st} \leq \frac{n^2 p}{4},$$

we conclude that there exists a $K_{s,t}$-free graph G' with at least $\frac{n^2 p}{4}$ edges.

The best choice is p of the form $Cn^{\frac{2-s-t}{st-1}}$ which leads to

$$\mathrm{ex}(n, K_{s,t}) \geq n^{2 - \frac{s+t-2}{st-1}}.$$

The conjecture is that, assuming $s \leq t$, we have $\mathrm{ex}(n, K_{s,t}) = \Theta(n^{2-1/s})$.

1.4 More problems for graphs/hypergraphs

Here we briefly describe more type of problems that one could ask on graphs/hypergraphs and many of them are still unknown:

- **Robustness** Suppose G contains "many" copies of H. How many edges can we delete without being able to make it H-free?
- **Super saturation** Suppose $e(G) \geq \mathrm{ex}(n, H) + k$. How many copies of H does it contain?
- **Counting H-free graphs** How many H-free graphs on n vertices exist?
- **Complicated "host graphs"** Suppose that G is and (say) Steiner Triple System (that is, a 3-uniform hypergraph for which evert pair of vertices is contained in **exactly** one edge). What is the size of the largest matching in G?
- **"Non algorithmic" sampling** Let $\mathcal{S}_n$ be the collection of all Steiner Triple Systems on n vertices, and let $S \in \mathcal{S}_n$ be a randomly chosen element (according to the uniform distribution). What is the typical size of the largest matching in S?
- **Packing and counting** Given two graph G and H, where H has at most as many vertices as G. Can you find "many" edge-disjoint copies of H? Can you prove that there are "many" distinct copies of H? (note that unlike the supersaturation problem above, here we also consider cases where H is large. For example, H is a perfect matching).
- **Threshold behavior** Let $G \sim G(n,p)$ (that is, G is a grpah on n vertices where each pair of vertices is being added as an edge with probability p, independently) and let H be any graph on at most n vertices. What is the smallest p for which G typically contains H?
- **Dirac threshold** Suppose p is above the threshold (that is, typically, $G \sim G(n,p)$ contains H). Is it true that typically G is such that every subgraph G' of G with at least $(1+\varepsilon)\mathrm{ex}(n,H)p$ edges must contain a copy of H?
- Same problem for hypergraphs?

References

[1] J. Balogh, R. Morris, and W. Samotij, *Independent sets in hypergraphs*, Journal of the American Mathematical Society **28** (2015), 669–709.

[2] P. Erdős, *On the combinatorial problems which I would most like to see solved*, Combinatorica **1** (1981), no. 1, 25–42.

[3] P. Erdős and A. H. Stone, *On the structure of linear graphs*, Bull. Amer. Math. Soc **52** (1946), 1087–1091.

[4] Z. Füredi and M. Simonovits, *The history of degenerate (bipartite) extremal graph problems*, In Erdős Centennial, pp. 169–264. Springer, Berlin, Heidelberg, 2013.

[5] D. Saxton and A. Thomason, *Hypergraph containers*, Inventiones Mathematicae **201** (2015), no. 3, 925–992.

[6] E. Szemerédi, *Integer sets containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 299–345.

[7] P. Turán, *Eine Extremalaufgabe aus der Graphentheorie*, Mat. Fiz. Lapok **48** (1941), 436–452.