Extremal combinatorics studies the maximum or minimum size of discrete structures under given constraints. For example, a classical question studied by Mantel in 1907 asks, “What is the maximum number of edges that a triangle-free graph can have?” The field enjoyed surprising growth in the last century fueled by its connection to other fields of mathematics and theoretical computer science that were later discovered. Naturally, numerous tools and methods have been developed in the course of advancement. One of the most successful tools among which is the so called probabilistic method. Probabilistic combinatorics on one hand refers to the study of this universal framework which can be potentially applied to any combinatorial problem, and on the other hand refers to the study of random objects such as the Erdős-Rényi random graph.

Ramsey theory refers to a large body of results in mathematics whose underlying philosophy is captured succinctly by the statement that “Every large system contains a large well-organized subsystem.” As an example, for positive integers $k$, define the Ramsey number $R(k)$ as the minimum integer $n$ in which every red, blue edge-coloring of $K_n$, the complete graph on $n$ vertices, contains a red or blue monochromatic $K_k$. Ramsey’s theorem, first proved in 1930, asserts that $R(k)$ is finite. Surprisingly, the problem of quantitatively estimating $R(k)$ remains as one of the central open problems in combinatorics, despite more than 80 years of endless efforts.

The three fields mentioned above, extremal combinatorics, probabilistic combinatorics, and Ramsey theory are three key areas in discrete mathematics that I am interested in. They share many central questions and it is often difficult to make a clear distinction between the fields. In fact, many of the problems that I have studied lies on the boundary of these areas, which is one reason that I was attracted to them.

### Generalized Ramsey numbers

One of the pillars of Ramsey theory, from which many other results follow, is the Hales–Jewett theorem. This theorem may be thought of as a statement about multiplayer, multidimensional tic-tac-toe, saying that in a high enough dimension one of the players must win (we will not give its precise technical statement here). It easily implies van der Waerden’s theorem on monochromatic arithmetic progressions in finite colorings of the integers.

The original proof of the Hales–Jewett theorem results in quantitative bounds of Ackermann type. In the late eighties, Shelah made a major breakthrough by proving the theorem with primitive recursive bounds. Shelah’s proof relied in a crucial way on a lemma now known as the Shelah cube lemma, the simplest case of which says that if we color the edges of the Cartesian product $K_n \square K_n$ in $r$ colors then, for $n$ sufficiently large, there is a rectangle with both pairs of opposite edges receiving the same color. Shelah’s proof shows that we may take $n \leq r^{(r+1)/2} + 1$. Graham, Rothschild and Spencer asked whether this bound can be improved to a polynomial in $r$, since such an improvement, if it could be generalized, would significantly improve Shelah’s upper bound from a wowzer-type to a tower-type. This question was later reiterated by Heinrich and by Faudree, Gyárfás, and Szőnyi. In a joint work with David Conlon, Jacob Fox, and Benny Sudakov we answered this question in the negative by providing a superpolynomial lower bound.
in $r$, thus ruling out the particular approach of improving the bound on Hales-Jewett theorem.

This work is closely connected to our another project on generalized Ramsey numbers. For positive integers $p$ and $q$, define the generalized Ramsey number $F(r, p, q)$ as the minimum $n$ for which every $r$-edge coloring of $K_n$ contains a $K_p$ whose edges receive less than $q$ colors. This function encodes several important problems in Ramsey theory, such as Ramsey’s theorem, Ruzsa-Szemerédi’s theorem, the $(7, 4)$-problem of Brown, Erdős, and Sós, the Shelah cube lemma mentioned above, and others as special cases. It was first defined by Erdős and Shelah about forty years ago, and then was systematically studied by Erdős and Gyárfás, who were interested in how the asymptotics of $F(r, p, q)$ in $r$ changes for fixed value of $p$ as $q$ varies from 2 to $\binom{p}{2}$. One particular problem they wanted to understand was the transition point of the function from superpolynomial to polynomial. Together with Conlon, Fox, and Sudakov, improving on previous results of Mubayi, and Eichorn and Mubayi, we completely solved the question of Erdős and Gyárfás by showing that that $F(r, p, p - 1)$ is superpolynomial, while $F(r, p, p)$ is polynomial in $r$.

**Related Work**


**Sidorenko’s conjecture**

For two graphs $H$ and $G$, a map from $V(H)$ to $V(G)$ is a homomorphism if $f(v)$ and $f(w)$ are adjacent whenever $v, w \in V(H)$ are adjacent vertices. Sidorenko’s conjecture (also made by Erdős and Simonovits in a slightly different form) asserts that for every bipartite graph $H$, the number of homomorphisms from $H$ to $G$ over graphs $G$ with a fixed edge density is asymptotically minimized when $G$ is the binomial random graph. This conjecture can be phrased in several different forms; as stated above, it is a problem in extremal graph theory, but it becomes a problem in analysis when stated as a functional inequality, and in probability theory when stated as a correlation inequality similar to that of the famous FKG inequality. Its analytical form also reveals interesting connections to Feynman integrals in quantum field theory, Mayer integrals in classical statistical mechanics, and multicenter integrals in quantum chemistry.

This beautiful conjecture is still wide open, and is known to be true only for a few families of bipartite graphs $H$ including trees, even cycles, complete bipartite graphs (Sidorenko), hypercubes (Hatami), graphs having one vertex adjacent to all vertices in the other part of the bipartition (Conlon, Fox, and Sudakov), and graphs constructed from these based on some recursive procedures (Sidorenko, and Szegedi and Li). Together with Jeong Han Kim and Joonkyung Lee, we discovered two new families of graphs for which the conjecture holds. First is the family of tree-arrangeable graphs, which are graphs admitting a specific form of a tree-decomposition. Our result generalizes the result of Conlon, Fox, and Sudakov since it includes the graphs having one vertex adjacent to all vertices in the other part. Second are graphs obtained by Cartesian products. We proved that for every tree $T$, if $H$ satisfies Sidorenko’s conjecture, then the Cartesian product $H \square T$ also does. This in particular implies Hatami’s result on hypercubes, since hypercubes can be defined as $K_2 \square \cdots \square K_2$. Furthermore, it shows that hypergrids of arbitrary dimension and side-lengths satisfy Sidorenko’s conjecture.

**Related Work**


Robustness of graphs

A typical result in graph theory says that “A graph has property $\mathcal{P}$ if it satisfies some condition $\mathcal{C}$.” For example, a classical theorem of Dirac from 1952 states that every $n$-vertex graph of minimum degree at least $n/2$ is Hamiltonian, i.e., contains a cycle passing through each vertex of the graph exactly once. Recently there has been an increasing interest in studying robustness of graph properties, aiming to strengthen classical results in extremal and probabilistic combinatorics by asking the question: “How strongly does a graph have the property $\mathcal{P}$ under the given condition $\mathcal{C}$?” For local properties, a breakthrough results independently obtained by Conlon and Gowers, and Schacht, give reasonable answers to such questions. However, the powerful tools developed there and in several subsequent work, most notably by Balogh, Morris, and Samotij, and Saxton and Thomason, do not give satisfactory answers for global properties.

I studied robustness of graphs with respect to global properties over a series of work, and would like to introduce two results in this direction related to Dirac’s theorem. First is its extension to binomial random graphs. In a joint work with Benny Sudakov, answering a question of Sudakov and Vu from 2007 and improving on several previous results, we proved that for $p \gg \log n / n$, asymptotically almost surely every subgraph of $G(n,p)$ of minimum degree at least $(1/2 + o(1))np$ contains a Hamilton cycle. Both the range of probability and the constant $1/2$ in our theorem are best possible. Second, in a joint work with Michael Krivelevich and Benny Sudakov, we proved that graphs of minimum degree at least $n/2$, which we refer to as Dirac graphs, are robustly Hamiltonian in a very strong sense. Our theorem asserts that if $p \geq C \log n / n$ for some positive constant $C$, then for every Dirac graph $G$, the random subgraph of $G$ obtained by taking each edge independently with probability $p$ contains a Hamilton cycle with high probability. In contrast, there exist graphs of minimum degree $n/2 - 1$ which does not even contain a single Hamilton cycle. Hence our theorem shows an interesting phenomenon where increasing the minimum degree by 1 makes a sudden shift in the behavior of graphs with respect to Hamiltonicity.

Related Work


Topics in extremal and probabilistic combinatorics

**Maximum union-free subfamilies**: A set $A$ of integers is sum-free if there are no $x, y, z \in A$ such that $x + y = z$. Erdős in 1965 proved that every set of $n$ nonzero integers contains a sum-free subset of size at least $\frac{n}{2}$. This is one of the first results in extremal number theory obtained by using the probabilistic method. The analogous problem in extremal set theory has also been studied for a long time. A family of sets is called union-free if no three distinct sets $X, Y, Z$ in the family satisfy $X \cup Y = Z$. An old problem of Moser asks: how large of a union-free subfamily does every family of $m$ sets have? Denote this number by $f(m)$. The study of $f(m)$ has attracted considerable interest. Riddell observed that $f(m) \geq \sqrt{m}$, and Erdős and Komlós determined the correct order of magnitude of $f(m)$ by proving that $f(m) \leq 2\sqrt{2m} + 4$. They conjectured
that \( f(m) = (c - o(1))\sqrt{m} \) for some constant \( c \), without specifying the right value of \( c \). In 1972, Erdős and Shelah improved both the upper and lower bound to \( \sqrt{2m} - 1 < f(m) < 2\sqrt{m} + 1 \) (the lower bound was also obtained independently by Kleitman), and conjectured that their upper bound is asymptotically tight. With Jacob Fox and Benny Sudakov, we proved that for all \( m \), \( f(m) = \lfloor \sqrt{4m + 1} \rfloor - 1 \). This verifies both Erdős and Komlós’s, and Erdős and Shelah’s conjectures and completely solves Moser’s problem.

**Quasi-randomness of graph balanced cut properties**: How can one tell when a given structure behaves like a random one? This natural question served as a motivation to study random-like deterministic structures, which we call quasi-random structures. One of the most popular quasi-random structures is quasi-random graphs, which, following Thomason, can be informally defined as graphs whose edge distribution closely resembles that of a truly random graph. In this context, we call an \( n \)-vertex graph \( p \)-quasi-random if for every subset of vertices \( U \), we have \( e(U) = \frac{1}{2}p|U|^2 + o(n^2) \). The cornerstone result of Chung, Graham and Wilson from 1989 established a surprising fact that many seemingly different properties are in fact equivalent to quasi-randomness. We call such graph properties as quasi-random properties.

Chung and Graham later studied the quasi-randomness of graph properties given by certain graph cuts. Their beautiful theorem says that for fixed real \( \alpha \in (0, 1) \), the property of having \( p\alpha(1 - \alpha)n^2 + o(n^2) \) edges across every cut \((U, V \setminus U)\) with \(|U| = \alpha n\) is quasi-random if and only if \( \alpha \neq 1/2 \), i.e., the cuts are not balanced. Shapira and Yuster studied its multipartite generalization by considering for \( k \geq 3 \), the property of having the “correct” number of complete graphs \( K_k \) crossing fixed ratio \( k \)-partite cuts. They were able to establish a partial analogue to Chung and Graham’s result saying that the property is quasi-random as long as the considered cuts are non-balanced. However, the problem of determining the quasi-randomness of the balanced case was left as an open problem. Janson also asked the same question in his study of quasi-randomness using graph limits.

Together with Hao Huang, we answered this question and concluded that the balanced case is quasi-random as well, thus showing that the bipartite case was an exception. Interestingly, as in many other related problems, this question turned out to be algebraic in nature, and we had to solve a system of non-linear equations to gain the key ingredient of our proof.

**Related Work**


Future plan

Various problems in extremal and probabilistic combinatorics, Ramsey theory, and discrete mathematics in general interests me. In the near future, I plan to continue working on several projects mentioned above. For the generalized Ramsey numbers, there are numerous interesting questions remaining to be answered. The following particular problem that I am currently working on with Brandon Tran, an undergraduate student at MIT, was asked during our project on Shelah’s cube lemma: “How many colors do we need in order to color $K_n$ so that the union of every $r$ color classes have chromatic number at most $2^r - 1$?” If we slightly relax the condition so that unions have chromatic number at most $2^r$, then a simple coloring based on binary expansions of integers show that $\log n$ colors suffice. On the other hand, for the original question, for $r = 2$, Conlon, Fox, Sudakov, and myself proved that at least $(\log n)^2$ colors are needed (up to some polynomial factor in $\log \log n$). The questions is, “For $r \geq 3$, is it still true that we need $\omega(\log n)$ colors?”. I find this problem particularly interesting since it asks to determine the optimality of the simple coloring based on binary expansion.

I also plan to study robustness of graphs from several different perspectives. Recall that the key question in the study of robustness was, “How strongly does a graph have the property?”. There are several ways to answer this question using different measures of robustness. For example, in the case of Dirac’s theorem, researchers have used a variety of measures to study the number of Hamilton cycles, the number of edge-disjoint Hamilton cycles, the cycle space generated by Hamilton cycles, etc. in Dirac graphs. I plan to further explore various possibilities, and use them to study classical results in extremal combinatorics in depth. The ultimate goal is to find a unified framework of studying robustness of graphs.

Discrete isoperimetric inequalities is another topic that I intend to work on in the future. The isoperimetric problem in the plane, which asks to determine a plane figure of largest possible area whose boundary has a specific length, is an old problem whose answer was known to be a circle already in Ancient Greece (although the rigorous proof was obtained only in the 19th century). For a given graph $G$ and a subset of vertices $X$, we define the vertex boundary of $X$ in $G$ as the number of vertices in $V(G) \setminus X$ that are adjacent to a vertex in $X$. The vertex-isoperimetric problem for graphs asks to determine the set of given size that minimizes the size of its vertex boundary. This problem and related problems has been extensively studied and has amazing deep connections to other famous problems. I am particularly interested in the stability of vertex-isoperimetric inequalities of hypercubes and its connection to extremal graph theoretical problems over the hypercube.

It is an exciting time to study discrete mathematics. From the theoretical point of view, connections to other fields of mathematics such as number theory, analysis, geometry, ergodic theory, theoretical computer science are bringing in intriguing open problems and exciting new developments. For example, the recent proof of the Kadison-Singer problem by Marcus, Spielman, and Srivastava shows how combinatorial methods can give new insights to problems in other areas (there are numerous other examples in the other direction as well). Also, from the applicational point of view, the need for tools that can be used in analyzing massive networks such as the internet, the synapse network of human brain, or Facebook friends graph is growing every day. Interestingly, open problems in discrete mathematics encodes these challenges in a very concrete form. These facts serve as an excellent motivation for my long term goal, which is to study intriguing open problems in the field with the hope of developing tools and methods that have great impact on various related fields as well.