Research Statement: Bill Kay

In section 1, I will propose a research project which will be of interest to the combinatorics community. This project will include relevant background and history to demonstrate the significance of the proposed research. Methods will be presented for each problem, as well as loose descriptions of some related problems whose solutions suggest that the proposed methods will be fruitful. In section 2, I will discuss my current research background along with my history of undergraduate mentorship and outreach. A sample of my 13 projects (7 accepted, 3 submitted, 1 in final coauthor revisions, and 2 in preparation) will be highlighted, and my future plans for outreach will be discussed.

1. Research Project.

My research lies at the confluence of extremal combinatorics, probabilistic combinatorics, and computer science. This project looks at the subfield of extremal combinatorics called Turán theory, which seeks to maximize one combinatorial property subject to some forbidden combinatorial substructure (for example, maximize the number of edges in a graph subject to the constraint that it contains no triangle as a subgraph). Various computational and combinatorial approaches have been successfully applied to problems in the area of Turán theory, which I will apply to a sequence of related problems.

This section contains a description of Turán theory for multigraphs with edge multiplicity at most $q$ (henceforth $q$-Turán theory for $q$-multigraphs). Then, to motivate the study of $q$-Turán theory, an overview of significant results from Turán theory for $k$-graphs (which is defined similarly, see Keevash [22]) is included.

Fix an integer $q \geq 1$. Let $F$ be a (possibly infinite) family of $q$-multigraphs, and define $ex_q(n, F)$ to be the maximum number of edges in a $q$-multigraph on $n$ vertices which contains no member of $F$. The quantity:

$$d_n := \frac{ex_q(n, F)}{\binom{n}{2}}$$

is the “edge density” of the densest $F$-free $q$-multigraph on $n$ vertices. An averaging argument shows that the $q$-Turán density of $F$:

$$\pi_q(F) := \lim_{n \to \infty} d_n$$

exists and lies on the interval $[0, q]$. Let $\Pi_q$ denote the set of all Turán densities attainable by some family of $q$-multigraphs. We will now introduce Turán theory for $k$-graphs to motivate investigation of $\Pi_q$.

A $k$-uniform hypergraph (henceforth $k$-graph) is a pair $H = (V, E)$ where $V$ is a finite set, and $E$ is a collection of subsets of $V$, each having cardinality $k$. For a family of $k$-graphs $F$, Let $\pi(F)$ be the Turán density of $F$, defined appropriately for $k$-graphs. Let $\Pi^{(k)}$ denote the set of all Turán densities attainable by some family of $k$-graphs. The computation of Turán densities for graphs has a rich history in combinatorics, culminating in the celebrated Erdős-Stone-Simonovits theorem [14], [15], which states that $\Pi^{(2)} = \{1 - \frac{1}{k}\}_{k=1}^{\infty}$, and $\pi(F)$ is determined by the member of $F$ with minimal chromatic number. In particular, $\Pi^{(2)}$ is well-ordered and has a complete description. A number $\alpha \in [0, 1)$ is a jump for an integer $k \geq 2$ if there exists $c = c(\alpha)$ such that $(\alpha, \alpha + c) \cap \Pi^{(k)} = \emptyset$. An immediate consequence of the Erdős-Stone-Simonovits theorem is that every $\alpha \in [0, 1)$ is a jump for $k = 2$.

This lead to Erdős’ so called jumping constant conjecture, a $\$1000 Erdős problem, which says that every $\alpha \in [0, 1)$ is a jump for $k \geq 3$. However, by exhibiting non-jumps for every $k \geq 3$, Rödl and Frankl showed [18] that the jumping constant conjecture is false. Since then, numerous results have emerged that demonstrate that the set of possible Turán densities is much more complex than previously conjectured.
For example, Pikhurko showed [26] that for \( k \geq 3 \), \( \Pi^{(k)} \) has uncountable cardinality and is the topological closure of the set of Turán densities attainable by forbidding only finite families.

This project aims to address open problems in \( q \)-Turán theory. Many of these problems can be stated in analogy with well known results in Turán theory for \( k \)-graphs which have been resolved using techniques from computational and extremal combinatorics. For example, \( \Pi_1 = \Pi^{(2)} = \left\{ 1 - \frac{1}{k} \right\}_{k=1}^{\infty} \) for which every \( \alpha \in [0, 1) \) is a jump. However, Rödl and Sidorenko showed [30] that \( [0, q) \) contains non-jumps for \( \Pi_q \) whenever \( q \geq 4 \), disproving the analogous jumping constant conjecture for \( q \)-multigraphs in this case. Hence, for different values of \( q \), the sets \( \Pi_q \) can be structurally diverse. My research and educational experiences focus heavily on computational and extremal combinatorics, and this project will close the gap in knowledge between the two theories using techniques known to be successful for related problems.

This project will resolve problems motivated by known results in the area of Turán theory for \( k \)-graphs. Computational techniques such as flag algebras employed by Baber and Talbot [5] as well as combinatorial techniques such as recursive constructions and blowups employed by Mubayi [24] and Pikhurko [26] will be fruitful. Recent progress in the field of \( q \)-Turán theory of Horn, La Fleur, and Rödl [21] have employed spectral and probabilistic methods. This project will apply each technique to \( q \)-Turán theory. Particular emphasis has been placed on the utility of Turán theory for \( k \)-graphs in an exploration of \( q \)-Turán theory. We will explore the reverse investigation as well; advances in computing and understanding \( q \)-Turán densities yield insight into the study of Turán densities for \( k \)-graphs. Hence, this project facilitates the study of one of the most elusive fields in extremal combinatorics without suggesting an avenue which has been exhausted.

### 1.1 Jumps in Multigraphs

Central to the study of extremal combinatorics is the notion of a blowup and the Lagrangian. Fix \( k \geq 3 \) and \( q \geq 4 \). In this section, blowups and Lagrangians are defined for multigraphs, and explanations are given as to how the analogous notions were used to disprove the jumping constant conjecture and produce jumps for \( k \)-graphs.

**Definition 1** (Blowup of a Multigraph). Let \( G \) be a multigraph on vertex set \( V = \{ v_1, v_2, \ldots, v_n \} \). Let \( \vec{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{Z}_{\geq 0}^n \). The blowup \( G(\vec{x}) \) is the graph on vertex set \( V_1 \cup V_2 \cup \ldots \cup V_n \) with \( |V_i| = x_i \), and \( e = \{v_i, v_j\} \) is an edge in \( G \) with multiplicity \( t \) if and only if for every pair \( u \in V_i \) and \( v \in V_j \), \( \{u, v\} \) is an edge in \( G(\vec{x}) \) with multiplicity \( t \).

Loosely speaking, replace each vertex \( v_i \) with an independent set of \( x_i \) vertices. Then, whenever there is an adjacency between \( v_i \) and \( v_j \), put all possible edges between \( V_i \) and \( V_j \) (with appropriate multiplicity). Tied closely to the blowup is the Lagrangian function of a multigraph:

**Definition 2** (Lagrangian of a Multigraph). Let \( G \) be a multigraph on vertex set \( V = \{ v_1, v_2, \ldots, v_n \} \). Let \( A = (a_{ij}) \) be the matrix where \( a_{ij} = t \) iff edge \( \{v_i, v_j\} \) has multiplicity \( t \). The Lagrangian of \( G \) is given by:

\[
\Lambda(G) := \max \left\{ \sum_{i<j} a_{ij} x_i x_j : x_1 + x_2 + \cdots + x_n = 1 \right\}
\]

For graphs, \( k \)-graphs, and multigraphs, the Lagrangian is the density of the densest blowup of \( G \). The following theorem is central to disproving the jumping constant conjecture for \( q \)-multigraphs:

**Theorem 3** (Rödl, Sidorenko [30]). \( \alpha \in [0, q) \) is a jump for \( \Pi_q \) if and only if there exists a finite family \( \mathcal{F} \) such that \( \pi_q(\mathcal{F}) \leq \alpha \) and \( \alpha < \min \{ \Lambda(G) : G \in \mathcal{F} \} \).
To disprove the jumping constant conjecture for $q$-multigraphs in 1995 (resp. $k$-graphs in 1984), it was shown that no family as in the second clause of Theorem 3 (or its $k$-graph analogue due to Frankl and Rödl [18]) could exist for particular choices of $\alpha$. This is where the story diverges.

For $\Pi^{(k)}$ (the $k$-graph case), Erdős showed that every $\alpha \in [0, k!/k^k)$ was a jump. Until 2011, the question remained open whether or not any values of $\alpha \in [k!/k^k, 1)$ were jumps, at which point Baber and Talbot answered the question in the affirmative. In order to apply Theorem 3, one needs to produce lower bounds on Lagrangians and upper bounds on Turán densities of families. Tools have existed to compute the former, and Razborov’s flag algebras [27] provided tools for computing the latter. Flag algebras reduce an extremal problem, such as computing bounds on Turán densities, to a semidefinite program. Using this method, Baber and Talbot produced a family of 3-graphs $F$ such that $\pi(F) \leq .2299$, while $\min_{F \in \mathcal{F}} \lambda_1(F) = .2316$, establishing that every $\alpha \in [.2299, .2316]$ is a jump.

For $\Pi_q$ (the $q$-multigraph case), Horn, La Fleur, and Rödl showed [21] that every $\alpha \in [0, 2)$ is a jump (extendin [14], [15] for $q = 2$), while Rödl and Sidorenko showed [30] that $\alpha \in \{3, 4, 5, \ldots, q - 1\}$ are all non-jumps. However, it remains open whether or not any $\alpha \in (3, q)$ is a jump. In each case, there is an initial interval of jumps, and once non-jumps are detected, jumps become elusive.

**Problem 1.** Let $q \geq 4$. Do there exist jumps on the interval $(3, q)$?

The natural method for solving this problem would be to proceed in the style of Baber and Talbot. I have written MATLAB code which computes Lagrangians for $q$-multigraphs, and an implementation which employs multigraph flag algebras is a relatively straightforward project, which I intend to undertake to complete this project. The final step would be to cleverly pick a family of $q$-multigraphs.

**1.2 Cardinality of $\Pi_q$.** Let $k \geq 3$. Pikhurko showed [26] that $\Pi^{(k)}$ has uncountable cardinality. To do this, he first picks a non-jump $\alpha \in [0, 1)$ and a sequence of $k$-graphs $\{H_i\}_{i=1}^{\infty}$ so that the $\Lambda(H_i)$ decrease to $\alpha$ “quickly.” Note: to show that this can be done, he uses a result which leans on the infinite removal lemma of Rödl and Schacht [29]. For any infinite $A \subseteq \mathbb{N}$, he defines a blowup procedure based on $A$ so that for any $A$ and $B$ infinite subsets of $\mathbb{N}$, neither contained in the other, the densities of the blowup constructions are different.

**Problem 2.** Determine for which values of $q$ the cardinality of $\Pi_q$ is countable, and for which it is uncountable.

For every $q \notin \{1, 2\}$, the cardinality of $\Pi_q$ is unknown. It is possible that for each $q$, $\Pi_q$ is countable. This would be interesting, because there is no class of Turán densities for graphs, $k$-graphs, or $q$-multigraphs that is not well-ordered but known to be countable. An argument that for some $q$, $\Pi_q$ is uncountable could follow the scheme of Pikhurko with modification. First, for any $q \geq 4$, one can still pick a non-jump $\alpha$. To produce a sequence of $q$-multigraphs $\{H_i\}_{i=1}^{\infty}$ with $\Lambda(H_i)$ decreasing to $\alpha$ is non-trivial. Pikhurko’s argument utilized the infinite removal lemma, for which there is no $q$-multigraph analogue. However, this was the most direct path to the desired result and could be circumvented. If not, it is probable that one can prove an infinite removal lemma for $q$-multigraphs in the style of Alon and Shapira [2], which is a project I am willing to undertake. The blowup construction of Pikhurko only works for $k$-graphs, but a different blowup construction for multigraphs could be fruitful. **Plainly, the main characteristic of $\Pi_q$ or $\Pi^{(k)}$ necessary to prove uncountable cardinality might be the presence of non-jumps.**

**Problem 3.** Prove that $\Pi_q$ is uncountable if and only if $[0, q)$ contains non-jumps for $\Pi_q$.
## 1.3. Structure of \( \Pi_q \)

Determining membership in \( \Pi^{(k)} \) is historically very difficult. Prior to 2006, the only known values in \( \Pi^{(3)} \) were 0, \( \frac{2}{3}, \frac{4}{5}, \frac{3}{2}, \) and 1. Mubayi showed [24] \( \left\{ \frac{(r-1)(r-2)}{r^2} \right\}_{r=4}^{\infty} \subseteq \Pi^{(3)} \) providing the first infinite subset of \( \Pi^{(3)} \). Since then, the machinery of flag algebras has been useful in producing specific values of \( \Pi^{(k)} \) (see Baber and Talbot [5] and Falgas-Ravry and Vaughan [16]). One of the largest breakthroughs in recent years is a recursive blowup construction of Pikhurko [26] which provides an infinite family of \( k \)-graphs whose edge densities are easy to compute and are members of \( \Pi^{(k)} \). As an application, he showed that there are irrational members of \( \Pi^{(k)} \) obtained by forbidding only finite families.

Substantially less is known about specific members of \( \Pi_q \). More effort has gone into establishing the order type of \( \Pi_2 \) and \( \Pi_3 \cap [0,2] \) [32], [21], and establishing that \( \Pi_q \) is not well-ordered for \( q \geq 4 \). This project will compute specific values of \( \Pi_q \), both with flag algebras and in the style of Mubayi and Pikhurko.

**Problem 4.** Provide a recursive blowup construction which produces infinitely many values of \( \Pi_q \).

**Problem 5.** Are there irrational numbers in \( \Pi_q \) for any \( q \geq 2 \)?

Brown and Simonovits showed [8] that for every \( \varepsilon > 0 \) and every family of \( k \)-graphs \( F \), there exists a finite family \( F' \subseteq F \) so that \( \pi(F') \leq \pi(F) + \varepsilon \). In some sense, this theorem says that every member of \( \Pi^{(k)} \) is approximable by forbidding finite families up to arbitrary precision. Approximations of Turán densities of \( k \)-graphs via simulation of \( k \)-graphs by infinite families of \( q \)-multigraphs with high multiplicity will be considered.

**Problem 6.** Let \( \varepsilon > 0 \). For which 3-graphs \( H \), does there exists an infinite family \( F \) of \( q \)-multigraphs so that \( \frac{1}{q} \pi_q(F) - \varepsilon < \pi(H) < \frac{1}{q} \pi_q(F) + \varepsilon \)?

## 2. Research Experience, Undergraduate Mentorship, and Outreach

### 2.1 Research Experience
I have extensive background in extremal and probabilistic combinatorics as well as computer science. In addition to coursework, I have attended workshops on methods which can be applied to these and other problems and have a research background in related fields. The most relevant workshops are the GGS [20] and the SPSACO [31], where there were courses on graph limits, flag algebras, and graph homomorphisms which are central to the results of Baber and Talbot [5], Mubayi [24], and Pikhurko [26]. I will now discuss three of my recent research projects.

**Project 1** [3]. Given sets \( X, Y \subseteq [n] := \{1,2,\ldots,n\} \) each of cardinality \( k \), we say that \( X \) and \( Y \) have order type \( \tau \in [3]^k \) if \( |X \cap Y| = \ell \) and \( \tau_i = 1, 2, \) or 3 whenever the \( i \)th element of \( X \cup Y \) is in \( X \setminus Y, Y \setminus X, \) or \( X \cap Y \) respectively. The shift graph \( G(n, \tau) \) is the graph whose vertices are all \( k \)-subsets of \( [n] \) and a pair \( X \) and \( Y \) are adjacent if and only if they have order type \( \tau \). The chromatic number of \( G(n, \tau) \) has been studied extensively by Erdős, Hajnal, Rado, and others. More recently, Avart, Luczak, and Rödl [4] asked if one could determine the chromatic number for \( G(n, \tau) \) for arbitrary \( \tau \). This question was answered in the following:

**Theorem 4** (Avart, K., Reiher, Rödl). For any type \( \tau \), there exists an (easily computable) constant \( \beta = \beta(\tau) \) so that

\[
\chi(G(n, \tau)) = \Theta(\log(\beta) n)
\]

where \( \log(t) n \) denotes the \( t \)-fold iterated logarithm of \( n \).
This paper has been tentatively accepted to the Journal of Combinatorial Theory, Series B pending revision.

**Project 2 [11].** Let $G = (V, E)$ be a graph. The line graph of $G$ (denoted $L(G)$) is the graph whose vertex set is $E$, and two vertices $e, f \in E$ are adjacent if and only if $e \cap f \neq \emptyset$. $L(G)$ is again a graph, so construct the iterated line graphs as follows:

- $L^{(0)}(G) = G$.
- $L^{(i)}(G) = L(L^{(i-1)}(G))$ for $i \geq 1$.

Define the line graph sequence of $G$ as the integer sequence $\{|L^{(i)}(G)|\}_{i=0}^{\infty}$. Finally, we say that $G$ is *Graham equivalent* to $G'$ if they have the same line graph sequence. Graham’s Tree Reconstruction Conjecture says that if two trees have the same line graph sequence, then they are isomorphic. Rephrased slightly:

**Conjecture 1** (Graham). The number of Graham equivalence classes of trees on $n$ vertices equals the number of isomorphism classes of trees on $n$ vertices.

Otter showed [25] that the number of isomorphic classes of trees on $n$ vertices is approximately $3^n$. Cooper, Swifton, and I showed the following:

**Theorem 5** (Cooper, K., Swifton). The number of Graham equivalence classes of trees on $n$ vertices is $e^{\Omega((\log n)^{3/2})}$.

While subexponential, the above bound is superpolynomial and is an application of combinatorial number theory. In particular, solutions to the Prouhet-Tarry-Escott problem are used to analyze polynomials related to the $k$-th iterated line graph for a specific family of graphs. This paper is presently submitted.

**Project 3 [1].** Let $S_n$ denote the set of permutations on symbols $[n] := \{1, 2, \ldots, n\}$. Given a permutation $\pi \in S_n$ and $\sigma \in S_{n+1}$, we say that $\sigma$ covers $\pi$ if $\sigma$ contains $\pi$ as an order isomorphic subpattern. Loosely speaking, if we can delete a symbol from $\sigma$ so that what’s left (when viewed as a permutation on $[n]$) is a copy of $\pi$. We are interested in questions concerning subsets of $S_{n+1}$ which cover the entirety of $S_n$. To this end, we have the following:

**Theorem 6** (Allison, Godbole, Hawley, K.). Let $\kappa_{n, \lambda}$ be the size of the smallest subset of $S_{n+1}$ which covers each member of $S_n$ at least $\lambda$ times. Then we have (by a probabilistic argument):

$$\frac{(n+1)!}{n^2} \leq \kappa_{n,\lambda} \leq \frac{(n+1)!}{n^2} (\log n + (\lambda - 1) \log \log n + O(1))$$

Notice that when $\lambda = 1$, the upper bound matches the lower bound with an extra factor of $\log n$. It remains open whether or not this $\log n$ factor can be removed. We also showed the following:

**Theorem 7** (Allison, Godbole, Hawley, K.). Let $A$ be a random subset of $S_{n+1}$ where each element of $S_{n+1}$ is selected for membership in $A$ with probability $p$. Then we have:

$$p \ll \frac{\log n}{n} \Rightarrow \mathbb{P}(A \text{ covers } S_n) \to 0 (n \to \infty)$$

$$p \gg \frac{\log n}{n} \Rightarrow \mathbb{P}(A \text{ covers } S_n) \to 1 (n \to \infty)$$
establishing a *probabilistic threshold* for permutation covers via Janson’s inequality. This paper has appeared in the Electronic Journal of Combinatorics.

In total I have 13 projects (7 accepted, 3 submitted, 1 in final coauthor revisions, and 2 in preparation) covering topics from extremal and probabilistic combinatorics, structural graph theory, hypergraph problems, universal cycles, and poset saturation (Accepted and submitted: [1], [3], [6], [7], [9], [10], [11], [12], [13], [17], [19])

At Emory University, I have given at least 6 talks on each of the topics of flag algebras, multigraph Turán theory, and hypergraph containers in which papers related to each topic were explained in full detail to graduate students and faculty members from Emory University and the Georgia Institute of Technology. The first two topics apply directly to this research proposal, while the third is a powerful new technique in probabilistic combinatorics, with which I am now intimately familiar. In addition to research on $q$-Turán densities, the method of hypergraph containers will be a part of my work at the next level. This project has a computational aspect; I *hold an M.S. in computer science* and have a strong background in the field of computational combinatorics. My Master’s thesis at the University of South Carolina included an implementation of the algorithmic Lovász Local Lemma of Moser and Tardos [23] to hypergraph coloring problems.

2.2 Undergraduate Mentorship and Outreach. I boast a strong record of mentoring undergraduate research across STEM fields. During each summer of 2010-2015 I participated in the East Tennessee State University Research Experience for Undergraduates as a graduate student mentor. I have provided guidance to numerous students from many different backgrounds. As a result, my total number of undergraduate coauthors is 16 which constitutes a diverse group. That is to say, facilitating undergraduate research, especially among underrepresented groups, has been a major component of my research program to this point and will continue to be going forward. An eventual career goal is to (co-)run an REU. More presently, I will make clear to interested undergraduate students that I am available to establish research collaborations. In addition to the field of mathematics, I have some experience with supporting undergraduates in computer science research. While I was the instructor of record for Introduction to Computer Science at Emory, one of my top students approached me for a letter of recommendation for her to attend a data visualization REU at Clemson University [28], to which she was subsequently accepted. She has been informed that her work will be credited in a forthcoming publication. I believe very strongly that one of the best ways to encourage young people to pursue STEM careers is to encourage early exposure to research, and while my primary research program will always aim to be competitive and modern, an undergraduate component will always be a priority.
References Cited


[17] Ferrara, Michael; Kay, Bill; Kramer, Lucas; Martin, Ryan; Reiniger, Benjamin; Smith, Heather; Sullivan, Eric, Induced Saturation in Posets. In final revisions.


