Introduction

We have worked on a wide range of topics in discrete mathematics, with such topics including spectral graph theory [33, 35], enumerative combinatorics [34, 36, 37], and elementary number theory [5, 39]. However, the bulk of our work lies in extremal and probabilistic combinatorics.

Broadly speaking, questions in *extremal combinatorics* ask how large or small a combinatorial object can be. For example, a classical theorem of Mantel's [22] states that every n-vertex triangle-free graph has at most $\frac{1}{4}n^2$ edges. More generally, Turán's problem asks for the maximum number of edges that an n-vertex F-free graph can have. Many of the recent advances towards solving special cases of Turán's problem have used tools from other areas of mathematics, such as Galois theory, algebraic geometry, and probability. In turn, solutions to Turán's problem have found applications in other areas of math such as number theory, discrete geometry, and coding theory.

Broadly speaking, *probabilistic combinatorics* studies both the application of tools from probability to solve problems in combinatorics, as well as the study of random combinatorial objects such as random graphs and random permutations. For example, Erdős [12] used random graphs to give the first exponential lower bound for diagonal Ramsey numbers. Since then, many tools from probability have been used to solve longstanding open problems in combinatorics. Probabilistic techniques have also been utilized together with tools from other areas of math to great effect. For example, Bukh [4] utilized random algebraic varieties to solve Turán's problem for certain complete bipartite graphs. A mixture of algebraic and probabilistic methods were also used by Keevash [20] in order to prove the existence of designs, and by Conlon and Ferber [7] to give an exponential improvement to bounds for multicolor Ramsey numbers. All of these results were major breakthroughs for longstanding open problems.

Below we outline some of the main subareas within extremal and probabilistic combinatorics that we are interested in, as well as various problems within each of these subareas that we plan to pursue.

Probabilistic Combinatorics

In addition to utilizing probabilistic tools throughout our work, within probabilistic combinatorics we are particularly interested in studying extremal properties of random objects. For example, we consider problems related to maximum and minimum scores in a certain card guessing game involving a randomly shuffled deck of cards. We also consider a randomized version of Turán's problem, which asks for the maximum number of edges in an F-free subgrpah of a random graph.

Card Guessing with Feedback. Consider the following one player game. We start with a deck of mn cards which consists of n card types, each appearing with multiplicity m. For example, a standard deck of playing cards corresponds to $n = 13$ and $m = 4$. The deck is shuffled uniformly at random, and then the player iteratively guesses the card type of the top card of the deck. After each guess, the top card is revealed and then discarded, with this process repeating until the deck is depleted. This game is known as the *complete feedback model*. One can also consider the *partial feedback model*, where instead of being told the card type each round, the player is only told whether their guess was correct or not. These models have been studied extensively, in part due to their applications to clinical trials [3], casino games [13], and many other real-life problems; see [11] for more information about applications.

Diaconis and Graham [9] determined the maximum and minimum expected number of correct guesses that a player can make in the complete feedback model. Further, they showed that the intuitive strategies "guess a most/least likely card type each round" give the maximum/minimum number of expected correct guesses.

The analogous problems for partial feedback are harder. This is because the strategies which achieve the maximum and minimums are unknown. Further, it is known that under partial feedback, the intuitive strategies "guess a most/least likely card type each round" do not achieve the maximum/minimum in general. Both the maximum and minimum problems remained open for nearly 40 years, but recently we essentially solved the maximum problem together with Diaconis, Graham and He:

Theorem 1 (Diaconis, Graham, He, S. [10])**.** *There exists an absolute constant $C > 0$ such that if n is sufficiently large in terms of m, then the expected number of correct guesses made in the partial feedback model is at most $m + Cm^{3/4}\log m$ regardless of the strategy used by the player.*

This bound is essentially best possible, as the player can guarantee m correct guess by guessing the same card type every round. The main obstacle in proving Theorem 1 was that the optimal strategies for this game are not known. We overcame this difficulty by using novel probabilistic arguments, as well as enumeration results which bound the number of permutations which have restricted entries.

Theorem 1 shows that the player cannot use partial feedback to get significantly more that m correct guesses, and we believe a similar phenomenon occurs when the player tries to minimize the number of correct guesses:

Problem 1. *Show that there exists an absolute constant $c > 0$ such that if n is sufficiently large in terms of m, then the expected number of correct guesses made in the partial feedback model is at least cm regardless of the strategy used by the player.*

The only lower bound for the expected number of correct guesses is due to Diaconis, Graham, and ourselves [11] who proved a lower bound of $\frac{1}{2}$, which is far from the conjectured value given in Problem 1.

The models we have described use decks which are shuffled uniformly at random, and it is natural to consider other ways of shuffling the deck. Results in this direction have been obtained for riffle shuffles [6, 21] and top to random shuffles [28]. Recently, we [38] considered the complete feedback model when the deck is shuffled "adversarially", i.e. in such a way that the maximum expected number of correct guesses that the player can obtain is minimized. We solved this problem in [38], and it is natural to consider the analogous problem under partial feedback:

Problem 2. *Determine the maximum expected number of correct guesses the player can make in the partial feedback model when the deck is shuffled adversarially.*

F-free Subgraphs of Random Hypergraphs. Szemerédi [43] famously proved that any dense subset of the integers contains arbitrarily long arithmetic progressions. Building on this, Green and Tao [16] proved that any large subset of a "psuedorandom" set of integers contains arbitrarily long progressions, which they used to prove that the primes contain arbitrarily long progressions. In a similar spirit, it was asked when the random set $[n]_p$, which is defined by including each of the first n integers $\{1, 2, \ldots, n\}$ independently and with probability p, is such that any dense subset of $[n]_p$ contains a k-term arithmetic progression with high probability. This problem was solved in breakthrough work by Conlon and Gowers [8] and Schacht [32]. The methods used in [8, 32] extend to many other probabilistic versions of classical problems, and one particular such problem that we are interested in is the problem of finding large F-free subgraphs of (random) graphs, and more generally of hypergraphs.

A *hypergraph* H is a set of vertices V together with a set E of subsets of V called *hyperedges*. A hypergraph is said to be r*-uniform* or an r*-graph* if every hyperedge has size exactly r. For example, the definition of a 2-graph is equivalent to the definition of a graph, and thus r-graphs can be viewed as a natural generalization of graphs. We define the *random r-graph* $G^r_{n,p}$ to be the r-graph on n vertices obtained by including each possible hyperedge independently and with probability p. For example, $G^2_{n,1}$ is the complete graph K_n since each possible edge is included with probability 1.

Given an r-graph F, we say that an r-graph H is F*-free* if H does not contain a subgraph isomorphic to F. Let $\mathrm{ex}(G^r_{n,p}, F)$ denote the maximum number of edges of an F-free subgraph of $G^r_{n,p}$. For example, when $p = 1$, the (deterministic) function $\mathrm{ex}(G^r_{n,1}, F)$ is the maximum number of hyperedges that an F-free r-graph on n vertices can have, which is exactly Turán's problem. Thus determining $\mathrm{ex}(G^r_{n,p}, F)$ can be viewed as a probabilistic analog to Turán's problem.

Problem 3. *Determine* $\mathbb{E}[\mathrm{ex}(G^r_{n,p}, F)]$ *for r-graphs F.*

Problem 3 has been essentially solved if F is not an r-partite r-graph due to independent work of Conlon and Gowers [8] and Schacht [32], but only sporadic results are known when F is an r-partite r-graph. One natural class of r-partite r-graphs to consider are complete r-partite r-graphs, and in this setting we proved the following[1] result with with Verstraëte:

Theorem 2 (S., Verstraëte [42])**.** *Let $K^r_{s_1,\ldots,s_r}$ denote the complete r-partite r-graph with parts of sizes $s_1, \ldots, s_r$. There exist constants $\beta_1, \beta_2, \beta_3, \gamma$ depending on $s_1, \ldots, s_r$ such that, for s_r sufficiently large in terms of $s_1, \ldots, s_{r-1}$, we have*

$$\mathbb{E}[\mathrm{ex}(G^r_{n,p}, K^r_{s_1,\ldots,s_r})] = \begin{cases} \Theta\left(pn^r\right) & 0 \le p \le n^{-\beta_1}, \\ n^{r-\beta_1+o(1)} & n^{-\beta_1} \le p \le n^{-\beta_2}(\log n)^{\gamma}, \\ \Theta(p^{1-\beta_3} n^{r-\beta_3}) & n^{-\beta_2}(\log n)^{\gamma} \le p \le 1. \end{cases}$$

This generalizes results of Morris and Saxton [25] when $r = 2$. We proved similar results with Verstraëte [40] and with Nie and Verstraëte [26] when avoiding Berge cycles and loose triangles in random hypergraphs.

[1]Throughout the text we use standard asymptotic notation: $O(f(x))$ (respectively $\Omega(f(x))$) denotes a function which is at most (respectively at least) $c \cdot f(x)$ for some constant $c > 0$, $\Theta(f(x))$ denotes a function which is both $O(f(x))$ and $\Omega(f(x))$, and $o(f(x))$ denotes a function which tends to 0 as x tends to infinity.

Extremal Combinatorics

Within extremal combinatorics, we are particularly interested in studying problems related to F-free graphs and hypergraphs. These include variants of Turán's problem, as well as newer problems such as counting maximal independent sets in K_t-free graphs.

F-free Subgraphs of General Hypergraphs If H and F are r-uniform hypergraphs, we define the *relative Turán number* $\mathrm{ex}(H,F)$ to be the maximum number of edges amongst all F-free subgraphs of H. Note that when H is K_n^r, the complete r-uniform hypergraph on n vertices, then $\mathrm{ex}(K_n^r,F)$ denotes the maximum number of edges that an F-free r-graph on n vertices can have. Because of its importance, we write $\mathrm{ex}(n,F) := \mathrm{ex}(K_n^r,F)$ and call this the *Turán number* of F.

In the previous section we discussed relative Turán numbers when H is a random hypergraph. In this section we consider $\mathrm{ex}(H,F)$ for general hypergraphs. In particular, we wish to bound $\mathrm{ex}(H,F)$ in terms of parameters of H. One particular problem of this form is the following:

Problem 4. *Given an r-graph F, determine lower bounds for $\mathrm{ex}(H,F)$ in terms of the number of edges of H and the maximum degree of H.*

One such result was proven by Perarnau and Reed [29]: for any graph G with maximum degree at most Δ,

$$\mathrm{ex}(G,K_{a,b}) = \Omega\left(\frac{\mathrm{ex}(\Delta,K_{a,b})}{\Delta^2}\right)\cdot e(G), \tag{1}$$

where $K_{a,b}$ is the complete bipartite graph with parts of sizes a and b, and $e(G)$ denotes the number of edges of G. We note that this result is essentially best possible because $G = K_\Delta$ has maximum degree at most Δ and satisfies

$$\mathrm{ex}(G,K_{a,b}) = \mathrm{ex}(\Delta,K_{a,b}) \approx \mathrm{ex}(\Delta,K_{a,b})\cdot\frac{e(G)}{\Delta^2}.$$

Surprisingly, the bound (1) holds despite the fact that the order of magnitude of $\mathrm{ex}(\Delta,K_{a,b})$ is unknown for most values of a and b. In order to generalize (1), the following was essentially conjectured by Foucaud, Krivelevich, and Perarnau [14]:

Conjecture 3. *If F and G are graphs such that G has maximum degree at most Δ, then*

$$\mathrm{ex}(G,F) = \Omega\Big(\frac{\mathrm{ex}(\Delta,F)}{\Delta^2}\Big)\cdot e(G).$$

One might naively conjecture that an analogous statement holds for r-uniform hypergraphs, namely that

$$\mathrm{ex}(H,F) = \Omega\Big(\frac{\mathrm{ex}(\Delta^{1/(r-1)},F)}{\Delta^{r/(r-1)}}\Big)\cdot e(H),$$

since a clique $H = K_n^r$ with $n \approx \Delta^{1/(r-1)}$ once again shows that such a bound would be best possible. With Verstraëte [42], we proved that this naive conjecture for hypergrpahs is very false, even for hypergraph analogs of $K_{a,b}$.

Theorem 4 (S., Verstraëte [42]). *Let $K^3_{a,b,c}$ be the complete 3-partite 3-uniform hypergraph with parts of sizes $a \leq b \leq c$. There exists a 3-uniform hypergraph H with maximum degree Δ such that*

$$\mathrm{ex}(H, K_{a,b,c}) = O(\Delta^{\frac{-1}{ab+a}}) \cdot e(H).$$

Moreover, if b is sufficiently large in terms of a, and if c is sufficiently large in terms of b, then for all 3-uniform hypergraphs H of maximum degree at most Δ, we have

$$\mathrm{ex}(H, K_{a,b,c}) \geq \Delta^{\frac{-1}{ab+a} - o(1)} \cdot e(H).$$

We emphasize that these bounds are in general *not* what one gets by considering $H = K^3_n$, since in this case it is conjectured that we have $\mathrm{ex}(K^3_n, K_{a,b,c}) = \Theta(n^{3-\frac{1}{ab}}) = \Theta(\Delta^{\frac{-1}{2ab}}) \cdot e(K^3_n)$. Thus the natural analog of (1) fails for 3-uniform hypergraphs. A generalization of Theorem 4 for complete r-partite r-graphs is also proven in [42].

We have proven analogs of Theorem 4 for other hypergraphs F. For example, with Verstraëte [41, 42] and Nie and Verstraëte [26], we have proven such bounds for various families of hypergraph cycles, with these including loose cycles, Berge cycles, and tight cycles.

Maximal Independent Sets in Clique-free Graphs. Given a graph G, a subset of vertices I is called a *maximal independent set* (or MIS for short), if I is an independent set and if every vertex $v \notin I$ has a neighbor in I (that is, if $I \cup \{v\}$ is not an independent set for any $v \notin I$). A large body of literature is dedicated to MIS's, in part due to applications to areas such as bioinformatics [31] and computer vision [18].

Miller and Muller [23] and Moon and Moser [24] independently proved that if n is a multiple of 3, then every n-vertex graph has at most $3^{n/3}$ MIS's, and this bound can be seen to be best possible by considering a disjoint union of triangles. Given this, it is natural to ask how many MIS's a graph G can have if it is "far" from a disjoint union of triangles. For example, Hujter and Tuza [19] showed that if n is even, then every n-vertex triangle-free graph has at most $2^{n/2}$ MIS's, and this bound is best possible by considering a disjoint union of edges. This latter result has found numerous applications: it was used by Balogh and Petříčková [2] to determine the number of maximal triangle-free graphs on n vertices, and by Balogh, Liu, Sharifzadeh, and Treglown [1] to count the number of maximal sum-free subsets of the first n integers.

Nielsen [27] determined that the maximum number of MIS's of a given size k that an n-vertex graph can have is asymptotic to $(n/k)^k$, with the extremal construction being k disjoint cliques with sizes as close to n/k as possible. Again it is natural to consider what happens for graphs which are "far" from a disjoint union of cliques. In this spirit, we proved the following with He and Nie:

Theorem 5 (He, Nie, S. [17]). *Let $m_t(n,k)$ denote the maximum number of MIS's of size k that an n-vertex K_t-free graph can have. For any fixed $k \geq 2(t-1)$, we have*

$$m_t(n,k) \geq n^{\frac{(t-2)k}{t-1} - o(1)}.$$

Our constructions involved certain blowups of hypergraphs related to a famous construction to Ruzsa and Szemerédi [30], as well as to a recent generalization of [30] due to Gowers and Janzer [15]. We believe that these constructions are essentially best possible:

Conjecture 6 (He, Nie, S. [17]). *For $k \geq 2(t-1)$, there exists a constant C depending on k such that*

$$m_t(n,k) \leq Cn^{\frac{(t-2)k}{t-1}}.$$

In [17] we showed that this conjecture is true for $t = 3$ and $k = 4$, but beyond this the problem is wide open. In [17] we showed that many different constructions achieve the lower bound of Theorem 5, which suggests that proving Conjecture 6 may be difficult. In addition to Conjecture 6, we are interested in studying how $m_t(n,k)$ behaves when k grows with n, as well as exploring analogous problems for hypergraphs.

References

[1] J. Balogh, H. Liu, M. Sharifzadeh, and A. Treglown. The number of maximal sum-free subsets of integers. *Proceedings of the American Mathematical Society*, 143(11):4713–4721, 2015.

[2] J. Balogh and Š. Petříčková. The number of the maximal triangle-free graphs. *Bulletin of the London Mathematical society*, 46(5):1003–1006, 2014.

[3] D. Blackwell and J. Hodges Jr. Design for the control of selection bias. *The Annals of Mathematical Statistics*, 28(2):449–460, 1957.

[4] B. Bukh. Extremal graphs without exponentially-small bicliques. *arXiv preprint arXiv:2107.04167*, 2021.

[5] F. Chung, R. Graham, and S. Spiro. Slow Fibonacci walks. *Journal of Number Theory*, 210:142–170, 2020.

[6] M. Ciucu. No-feedback card guessing for dovetail shuffles. *Annals of Applied Probability*, 8(4):1251–1269, 1998.

[7] D. Conlon and A. Ferber. Lower bounds for multicolor Ramsey numbers. *Advances in Mathematics*, 378:107528, 2021.

[8] D. Conlon and W. T. Gowers. Combinatorial theorems in sparse random sets. *Annals of Mathematics*, pages 367–454, 2016.

[9] P. Diaconis and R. Graham. The analysis of sequential experiments with feedback to subjects. *The Annals of Statistics*, 9(1):3–23, 1981.

[10] P. Diaconis, R. Graham, X. He, and S. Spiro. Card guessing with partial feedback. *Combinatorics, Probability and Computing*, pages 1–20, 2021.

[11] P. Diaconis, R. Graham, and S. Spiro. Guessing about guessing: Practical strategies for card guessing with feedback. *The American Mathematical Monthly (To Appear)*, 2020.

[12] P. Erdös. Some remarks on the theory of graphs. *Bulletin of the American Mathematical Society*, 53(4):292–294, 1947.

[13] S. Ethier and D. A. Levin. On the fundamental theorem of card counting, with application to the game of trente et quarante. *Advances in applied probability*, 37(1):90–107, 2005.

[14] F. Foucaud, M. Krivelevich, and G. Perarnau. Large subgraphs without short cycles. *SIAM Journal on Discrete Mathematics*, 29(1):65–78, 2015.

[15] W. Gowers and B. Janzer. Generalizations of the Ruzsa–Szemerédi and rainbow Turán problems for cliques. *Combinatorics, Probability and Computing*, pages 1–18, 2020.

[16] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions. *Annals of Mathematics*, pages 481–547, 2008.

[17] X. He, J. Nie, and S. Spiro. Maximal indpendent sets of clique-free graphs. *arXiv preprint arXiv:2107.09233*, 2021.

[18] R. Horaud and T. Skordas. Stereo correspondence through feature grouping and maximal cliques. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(11):1168–1180, 1989.

[19] M. Hujter and Z. Tuza. The number of maximal independent sets in triangle-free graphs. *SIAM Journal on Discrete Mathematics*, 6(2):284–288, 1993.

[20] P. Keevash. The existence of designs. *arXiv preprint arXiv:1401.3665*, 2014.

[21] P. Liu. On card guessing game with one time riffle shuffle and complete feedback. *Discrete Applied Mathematics*, 288:270–278, 2021.

[22] W. Mantel. Problem 28. *Wiskundige Opgaven*, 10(60-61), 1907.

[23] R. E. Miller and D. E. Muller. A problem of maximum consistent subsets. Technical report, IBM Research Report RC-240, JT Watson Research Center, Yorktown Heights, NY, 1960.

[24] J. W. Moon and L. Moser. On cliques in graphs. *Israel journal of Mathematics*, 3(1):23–28, 1965.

[25] R. Morris and D. Saxton. The number of $C_{2\ell}$-free graphs. *Advances in Mathematics*, 298:534–580, 2016.

[26] J. Nie, S. Spiro, and J. Verstraëte. Triangle-free subgraphs of hypergraphs. *Graphs and Combinatorics*, pages 1–16, 2021.

[27] J. M. Nielsen. On the number of maximal independent sets in a graph. *BRICS Report Series*, 9(15), 2002.

[28] L. Pehlivan. No feedback card guessing for top to random shuffles. *arXiv preprint arXiv:1006.1321*, 2010.

[29] G. Perarnau and B. Reed. Existence of spanning F-free subgraphs with large minimum degree. *arXiv preprint arXiv:1404.7764*, 2014.

[30] I. Z. Ruzsa and E. Szemerédi. Triple systems with no six points carrying three triangles. *Combinatorics (Keszthely, 1976), Coll. Math. Soc. J. Bolyai*, 18:939–945, 1978.

[31] R. Samudrala and J. Moult. A graph-theoretic algorithm for comparative modeling of protein structure11edited by f. cohen. *Journal of Molecular Biology*, 279(1):287–302, 1998.

[32] M. Schacht. Extremal results for random discrete structures. *Annals of Mathematics*, pages 333–365, 2016.

[33] S. Spiro. Polynomial relations between matrices of graphs. *Journal of Graph Theory*, 90(3):288–303, 2019.

[34] S. Spiro. Subset parking functions. *Journal of integer sequences*, 22(7), 2019.

[35] S. Spiro. An averaging processes on hypergraphs. *arXiv preprint arXiv:2004.13935*, 2020.

[36] S. Spiro. Ballot permutations and odd order permutations. *Discrete Mathematics*, 343(6):111869, 2020.

[37] S. Spiro. Counting labeled threshold graphs with eulerian numbers. *Australasian Journal of Combinatorics*, 77(2):249–255, 2020.

[38] S. Spiro. Online card games. *Electronic Journal of Probability (Accepted)*, 2021.

[39] S. Spiro. Slow recurrences. *Journal of Number Theory*, 218:370–401, 2021.

[40] S. Spiro and J. Verstraëte. Counting hypergraphs with large girth. *Journal of Graph Theory (Accepted)*, 2020.

[41] S. Spiro and J. Verstraëte. Relative Turán numbers for hypergraph cycles. *Discrete Mathematics*, 344(10):112542, 2021.

[42] S. Spiro and J. Verstraëte. Relative Turán problems for uniform hypergraphs. *SIAM Journal on Discrete Mathematics*, 35(3):2170–2191, 2021.

[43] E. Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith*, 27(299-345):5, 1975.