

Introduction and Background

Broadly speaking, questions in extremal combinatorics ask how large or small a combinatorial object can be. For example, a classical theorem of Mantel's [20] states that every n -vertex triangle-free graph has at most $\frac{1}{4}n^2$ edges. Many of the tools used in extremal combinatorics come from other areas of math, with such tools including the probabilistic method, the combinatorial Nullstellensatz, finite geometries, and topological methods. In turn, results in extremal combinatorics are often used to solve problems in other areas of math such as number theory, discrete geometry, computer science, and coding theory.

A large portion of our work lies at the intersection of extremal combinatorics and probability. The first connection between these two areas was established by Erdős who utilized random graphs to solve a problem in Ramsey theory, and since then many open problems in extremal combinatorics have been solved by using tools from probability. In addition to using probability to solve extremal problems, there has been a growing interest in using random objects as a source of new extremal problems to study, and this proposal focuses on problems of this form.

Past Accomplishments and Future Research Objectives

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 continuing 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 [2, 10], casino games [13], and many other real-life problems.

Diaconis and Graham [7] 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 much harder. This is because the strategies which achieve the maximum and minimums in this game are unknown. Moreover, 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 problems remained open for nearly 40 years, but recently the maximum problem was essentially solved by Diaconis, Graham, He, and the PI:

Theorem 1 (Diaconis, Graham, He, S. [8]). *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 get 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 than m correct guesses, and we believe a similar phenomenon occurs when the player tries to minimize their 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 the PI [9] who proved a lower bound of $\frac{1}{2}$, which is far from the conjectured value given in Problem 1. We have talked about Theorem 1 at several seminars, and in doing so we have developed new ideas on how we could modify the proof of Theorem 1 to solve 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 [4, 19] and top to random shuffles [24]. Recently, the PI [27] 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. This problem was essentially solved by the PI [27], 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 [32] famously proved that any dense subset of the integers contains arbitrarily long arithmetic progressions. Building on this, Green and Tao [17] 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 $\{1, 2, \dots, n\}$ 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 [5] and Schacht [25]. The methods used in [5, 25] extend to many other probabilistic versions of classical problems, and of particular interest to us 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 are a natural generalization of graphs. We define the *random r -graph* $G_{n,p}^r$ to be the r -graph on n vertices obtained by including each possible hyperedge independently and with probability p . For example, $G_{n,1}^2$ 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 $\text{ex}(G_{n,p}^r, F)$ denote the maximum number of edges of an F -free subgraph of $G_{n,p}^r$. For example, when $p = 1$, the (deterministic) function $\text{ex}(G_{n,1}^r, F)$ is the maximum number of hyperedges that an F -free r -graph on n vertices can have. This special case when $p = 1$ is known as Turán’s problem, which is a fundamental problem in extremal combinatorics.

For general p , determining $\mathbb{E}[\text{ex}(G_{n,p}^r, F)]$ when F is not an r -partite r -graph was essentially solved independently by Conlon and Gowers [5] and by Schacht [25], 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 the following was proven by the PI and Verstraëte:

Theorem 2 (S., Verstraëte [31]). *Let K_{s_1, \dots, s_r}^r denote the complete r -partite r -graph with parts of sizes s_1, \dots, s_r . There exist constants $\beta_1, \beta_2, \beta_3, \gamma$ depending on s_1, \dots, s_r such that, for s_r sufficiently large in terms of s_1, \dots, s_{r-1} , we have*

$$\mathbb{E}[\text{ex}(G_{n,p}^r, K_{s_1, \dots, s_r}^r)] = \begin{cases} \Theta(pn^r) & 0 \leq p \leq n^{-\beta_1}, \\ n^{r-\beta_1+o(1)} & n^{-\beta_1} \leq p \leq n^{-\beta_2}(\log n)^\gamma, \\ \Theta(p^{1-\beta_3}n^{r-\beta_3}) & n^{-\beta_2}(\log n)^\gamma \leq p \leq 1. \end{cases}$$

Theorem 2 generalizes results of Morris and Saxton [21] when $r = 2$. More broadly, we are interested in the following problem:

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

One concrete case of Problem 3 that we are interested in exploring is when F is a theta graph. We define the *theta graph* $\theta_{a,b}$ to be the graph consisting of a internally disjoint paths which are of length b and which all have the same endpoints x, y .

Problem 4. *Determine $\mathbb{E}[\text{ex}(G_{n,p}^2, \theta_{a,b})]$ for theta graphs $\theta_{a,b}$.*

The main motivation for this specific problem comes from work of Morris and Saxton [21]. There they proved bounds on $\mathbb{E}[\text{ex}(G_{n,p}^2, F)]$ when F is an even cycle $C_{2\ell}$ or a complete bipartite graph $K_{s,t}$, and these bounds are known to be best possible assuming some well known conjectures in extremal graph theory. Observe that $C_{2\ell} = \theta_{2,\ell}$ and $K_{2,t} = \theta_{t,2}$, so theta graphs can be viewed as a natural generalization of the F considered in [21].

Non-trivial bounds for $\mathbb{E}[\text{ex}(G_{n,p}^2, \theta_{a,b})]$ are implicitly given by Corsten and Tran [6], but these bounds are far from optimal. In particular, when $\theta_{a,b} = C_{2\ell}$ or $\theta_{a,b} = K_{2,t}$, the bounds of [6] are weaker than the bounds given by [21]. For these particular cases, we believe it should be possible to unify the ideas of [6, 21] to recover the bounds of [21], with the hope being that these methods can be used to give optimal bounds for all $\theta_{a,b}$, and possibly to other r -graphs F as well. We have looked through the papers [6, 21] many times, with the PI making use of their results in [30, 31]. As such, we believe that we are in a strong position for tackling Problem 4.

Counting F -free Hypergraphs. For an r -graph F , define $N_m(n, F)$ to be the number of F -free r -graphs which have n vertices and m hyperedges. A simple first moment argument shows that upper bounds on $N_m(n, F)$ imply upper bounds on $\mathbb{E}[\text{ex}(G_{n,p}^r, F)]$, and almost every known upper bound for $\mathbb{E}[\text{ex}(G_{n,p}^r, F)]$ comes from this relationship with $N_m(n, F)$. This motivates the following:

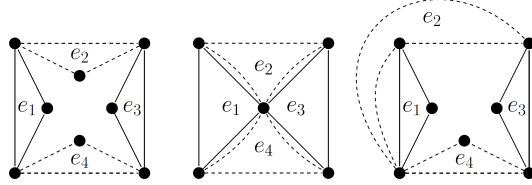
Problem 5. *Determine $N_m(n, F)$ for r -graphs F .*

We note that determining $N_m(n, F)$ is a refinement of determining $N(n, F)$, the number of F -free r -graphs on n vertices (with no restriction on the number of hyperedges). The problem of determining $N(n, F)$ has in large part been solved by Erdős, Frankl, and Rödl [12] and by Ferber, McKinley, and Samotij [14] by utilizing the method of hypergraph contains. However, only sporadic results for the refined quantity $N_m(n, F)$ are known.

Problem 5 was essentially solved when F is a graph cycle by Morris and Saxton [21]. A natural extension of this work is to consider hypergraph cycles. There are several notions of what it means

to be a hypergraph cycle, with the two most popular definitions being loose cycles and Berge cycles.

The r -uniform *loose ℓ -cycle* C_ℓ^r is the hypergraph consisting of ℓ hyperedges e_1, \dots, e_ℓ of size r such that there exist distinct vertices v_1, \dots, v_ℓ with $\{v_i\} = e_i \cap e_{i+1}$ for all $1 \leq i \leq r$, where the indices of these hyperedges are written cyclically. Note that when $r = 2$ this defines the graph cycle C_ℓ . An r -graph C is said to be a *Berge ℓ -cycle* if it consists of ℓ hyperedges e_1, \dots, e_ℓ such that there exist vertices v_1, \dots, v_r with $v_i \in e_i \cap e_{i+1}$ for all $1 \leq i \leq r$. For example, the loose cycle C_ℓ^r is a Berge ℓ -cycle, but there are many other examples of Berge cycles when $r > 2$. We let \mathcal{B}_ℓ^r denote the set of all r -uniform Berge ℓ -cycles. In the picture below, we show three 3-uniform Berge 4-cycles, with the leftmost hypergraph being the loose 4-cycle C_4^3 .



For loose cycles, Nie, the PI, and Verstraëte [23] obtained bounds for $N_m(n, C_3^r)$, with our bounds being tight for $r = 3$. Results for $N_m(n, C_{2\ell}^r)$ were obtained by Mubayi and Yepremyan [22], but these bounds are not tight in general. In particular, the following problem remains open:

Problem 6. Determine $N_m(n, C_4^3)$.

We next consider $N_m(n, \mathcal{B}_\ell^r)$, the number of n -vertex r -graphs with m hyperedges that contain no Berge ℓ -cycle. The best known bounds for $N_m(n, \mathcal{B}_\ell^r)$ are due to the PI and Verstraëte [30]. In [30] we used a novel reduction argument to prove the following result, which allows one to lift the bounds of [21] for the graph setting to the hypergraph setting:

Theorem 3 (S., Verstraëte [30]). For all $\ell, r \geq 2$, there exists $c = c(\ell, r) > 0$ such that

$$N_m(n, \mathcal{B}_{2\ell}^r) \leq 2^{cm} \cdot N_{\leq m}(n, C_{2\ell})^{r^{1/2}},$$

where $N_{\leq m}(n, C_{2\ell})$ denotes the number of n -vertex $C_{2\ell}$ -free graphs on at most m edges.

We do not know in general if the bounds of Theorem 3 are tight, and in particular we propose the following problem:

Problem 7. Determine $N_m(n, \mathcal{B}_4^3)$.

In [30], we focused on bounding $N_m(n, \mathcal{B}_{2\ell}^r)$ for general ℓ and r , and we suspect that we can improve upon our bounds by focusing on the concrete case of \mathcal{B}_4^3 . One particular direction to attack both this problem and Problem 6 would be to review the known proofs which give tight bounds for $N_m(n, C_4)$. There are three different proofs which achieves this result: the original proof by Füredi [16], as well as two proofs given by Morris and Saxton [21]. It is plausible that an appropriate generalization of one of these proofs could be used to resolve either Problem 6 or Problem 7.

Choice of Sponsoring Scientist and Host Institution

Our sponsoring scientist is Bhargav Narayanan from Rutgers University. Narayanan is an expert in probabilistic and extremal combinatorics, and is a leading authority on material related to the

proposed work. We are familiar with the work of Narayanan, and we know that his research interests closely align with our own. For example, our work with Nie and Verstraëte [23] crucially used a supersaturation result proven by Balogh, Narayanan, and Skokan [1]. Further, in [29] the PI gave a common generalization of breakthrough work by Frankston, Kahn, Narayanan, and Park [15] and by Kahn, Narayanan, and Park [18] related to thresholds in random structures.

In addition to Narayanan, Rutgers University is home to many leaders in extremal and probabilistic combinatorics, and the school is located near many other excellent universities such as Princeton and the Institute for Advanced Study. Being nearby so many great researchers will allow for many collaborative projects, making Rutgers University an ideal host institution.

Career Development

The most direct way that the MSPRF will help develop our career is by allowing the PI to more easily work together with many outstanding collaborators, such as Narayanan. These new collaborations will give the PI the opportunity to develop new skills, and to form strong ties with a number of new collaborators.

In addition to this, the MSPRF will give the PI more time to dedicate to activities such as research projects with graduate and undergraduate students. We have already led several research projects, most of which involved graduate students. We hope to use the knowledge and experience we have gained from these projects towards future projects at our host institution. In particular, our past work in combinatorial games [8, 9, 11, 26, 27] and elementary number theory [3, 28] contain many open problems that would be well suited for students.

Broader Impacts

The card guessing games considered in this proposal were motivated by real world problems, and it is likely that further work on these problems could lead to additional applications. We plan to disseminate our work by giving talks about our research at seminars and conferences. For example, we recently spoke at an invited minisymposium on extremal graph theory at SIAM Conference on Discrete Mathematics 2021. Many of our talks have been recorded and uploaded to YouTube, allowing for others to more easily access our talks. Another way we disseminate our work is through expository writing. On our website we maintain a set of notes related to methods in extremal combinatorics, which were written to help demystify the general techniques used throughout our work. In the coming years, we plan to further develop these notes, as well as to continue to speak at conferences and seminars.

We have organized a number of programs at our home institution UC San Diego (UCSD), such as the graduate student seminar, as well as the graduate student combinatorics seminar at UCSD. For the latter seminar, we organized two reading courses: one on spectral graph theory, and another on hypergraph containers. We are an organizer for the Graduate Student Combinatorics Conference 2022, which is a national conference run by and for graduate students in combinatorics.

The PI served as a mentor for a number of graduate and undergraduate students at UCSD. This included mentoring students from our school's Association for Women in Mathematics chapter, which has recently received the 2021 AWM Student Chapter Award for Professional Development. Through our school's Math Graduate Student Council, we helped develop a bootcamp for incoming graduate students at UCSD. We plan to take part in similar activities at Rutgers.

References

- [1] J. Balogh, B. Narayanan, and J. Skokan. The number of hypergraphs without linear cycles. *Journal of Combinatorial Theory, Series B*, 134:309–321, 2019.
- [2] D. Blackwell and J. Hodges Jr. Design for the control of selection bias. *The Annals of Mathematical Statistics*, 28(2):449–460, 1957.
- [3] F. Chung, R. Graham, and S. Spiro. Slow fibonacci walks. *Journal of Number Theory*, 210:142–170, 2020.
- [4] M. Ciucu. No-feedback card guessing for dovetail shuffles. *Annals of Applied Probability*, 8(4):1251–1269, 1998.
- [5] D. Conlon and W. T. Gowers. Combinatorial theorems in sparse random sets. *Annals of Mathematics*, pages 367–454, 2016.
- [6] J. Corsten and T. Tran. Balanced supersaturation for some degenerate hypergraphs. *Journal of Graph Theory*, 2021.
- [7] P. Diaconis and R. Graham. The analysis of sequential experiments with feedback to subjects. *The Annals of Statistics*, 9(1):3–23, 1981.
- [8] P. Diaconis, R. Graham, X. He, and S. Spiro. Card guessing with partial feedback. *Combinatorics, Probability and Computing*, pages 1–20, 2021.
- [9] P. Diaconis, R. Graham, and S. Spiro. Guessing about guessing: Practical strategies for card guessing with feedback. *The American Mathematical Monthly (To Appear)*, 2020.
- [10] B. Efron. Forcing a sequential experiment to be balanced. *Biometrika*, 58(3):403–417, 1971.
- [11] S. English, T. Masařík, G. McCourt, E. Meger, M. S. Ross, and S. Spiro. Linear bounds for cycle-free saturation games. *arXiv preprint arXiv:2108.05295*, 2021.
- [12] P. Erdős, P. Frankl, and V. Rödl. The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent. *Graphs and Combinatorics*, 2(1):113–121, 1986.
- [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] A. Ferber, G. McKinley, and W. Samotij. Supersaturated sparse graphs and hypergraphs. *International Mathematics Research Notices*, 2020(2):378–402, 2020.
- [15] K. Frankston, J. Kahn, B. Narayanan, and J. Park. Thresholds versus fractional expectation-thresholds. *Annals of Mathematics*, 194(2):475–495, 2021.
- [16] Z. Füredi. Random ramsey graphs for the four-cycle. *Discrete Mathematics*, 126(1-3):407–410, 1994.
- [17] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions. *Annals of Mathematics*, pages 481–547, 2008.

- [18] J. Kahn, B. Narayanan, and J. Park. The threshold for the square of a hamilton cycle. *Proceedings of the American Mathematical Society*, 149(08):3201–3208, 2021.
- [19] P. Liu. On card guessing game with one time riffle shuffle and complete feedback. *Discrete Applied Mathematics*, 288:270–278, 2021.
- [20] W. Mantel. Problem 28. *Wiskundige Opgaven*, 10(60-61), 1907.
- [21] R. Morris and D. Saxton. The number of $C_{2\ell}$ -free graphs. *Advances in Mathematics*, 298:534–580, 2016.
- [22] D. Mubayi and L. Yepremyan. Random Turán theorem for hypergraph cycles. *arXiv preprint arXiv:2007.10320*, 2020.
- [23] J. Nie, S. Spiro, and J. Verstraëte. Triangle-free subgraphs of hypergraphs. *Graphs and Combinatorics*, pages 1–16, 2021.
- [24] L. Pehlivan. No feedback card guessing for top to random shuffles. *arXiv preprint arXiv:1006.1321*, 2010.
- [25] M. Schacht. Extremal results for random discrete structures. *Annals of Mathematics*, pages 333–365, 2016.
- [26] S. Spiro. Saturation games for odd cycles. *The Electronic Journal of Combinatorics*, pages P4–11, 2019.
- [27] S. Spiro. Online card games. *Electronic Journal of Probability (Accepted)*, 2021.
- [28] S. Spiro. Slow recurrences. *Journal of Number Theory*, 218:370–401, 2021.
- [29] S. Spiro. A smoother notion of spread hypergraphs. *arXiv preprint arXiv:2106.11882*, 2021.
- [30] S. Spiro and J. Verstraëte. Counting hypergraphs with large girth. *Journal of Graph Theory (Accepted)*, 2020.
- [31] S. Spiro and J. Verstraëte. Relative turán problems for uniform hypergraphs. *SIAM Journal on Discrete Mathematics*, 35(3):2170–2191, 2021.
- [32] E. Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith*, 27(299-345):5, 1975.