Title
Extremal Problems for Random Objects

Author
Spiro, Sam

Publication Date
2022

UNIVERSITY OF CALIFORNIA SAN DIEGO

Extremal Problems for Random Objects

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Mathematics

by

Sam Spiro

Committee in charge:

Professor Jacques Verstraëte, Chair
Professor Fan Chung Graham
Professor Shachar Lovett
Professor Frederick Manners
Professor Andrew Suk

2022

The dissertation of Sam Spiro is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I have been incredibly fortunate to have so many people help in my mathematical journey. I'm very thankful for my family who supported me at every stage of my life. I'm eternally grateful for my high school teachers Mr. Will and Mr. Power for really getting me interested in math, as well as my college professors Drew Armstrong, Richard Stanley, Ervin Győri, András Gyárfás, and Attila Sali for getting me hooked on combinatorics. I thank my many friends and colleagues who supported me mathematically and otherwise throughout my life, and in particular I want to thank Eric Silva, Geoffrey Sangston, Michael Fishman, Andre Kundgen, Persi Diaconis, and my (soon to be) academic brother Jason O'Neill. I also want to give some particular praise to each of the three advisors I had while at UC San Diego.

First of all, I need to thank my PhD adviser Jacques Verstraëte. He has been by my side since year one, and his perpetually high standards have always pushed me to do my best work possible. I particularly want to thank Jacques for always being incredibly generous with his time. He has gone above and beyond with helping us run the Graduate Student Combinatorics Conference, and his incredibly thorough advice increased the quality of all of my application materials tenfold.

Next I want to thank Fan Chung. She jointly advised me with Jacques for my first three years of graduate school, and Fan was the very first person I started doing research with at UC San Diego. I always appreciated her honest feedback, her infectious love for working on problems, as well as her overwhelming hospitality. Even though Fan is no

longer my official adviser, I will always think of myself as one of Fan's students.

Last but certainly not least, I want to thank Ron Graham. While he was never officially my adviser, Ron was a constant source of mentorship and inspiration. I was always happy to chat with Ron, whether it be about a math problem, a juggling problem, or just to hear one of his many stories. While I only knew Ron for a few short years, I know that his positive influence will continue to permeate throughout my life, and because of this I would like to dedicate this dissertation to Ron Graham.

Sam Spiro

May 18th, 2022

Chapter 2 contains material from: P. Diaconis, R. Graham, X. He, and S. Spiro, *"Card Guessing with Partial Feedback"*, Combinatorics, Probability, and Computing (2021) 1–20. The dissertation author was one of the primary investigators and authors of this paper.

Chapter 3 contains material from: S. Spiro and J. Verstraëte, *"Counting Hypergraphs with Large Girth"*, Journal of Graph Theory, accepted (2021). The dissertation author was one of the primary investigators and authors of this paper.

Chapter 4 contains material from: S. Spiro, *"A Smoother Notion of Spread Hypergraphs"*, submitted (2021). The dissertation author was the primary investigator and author of this paper.

VITA

2017	Bachelor of Science, University of Miami
2022	Doctor of Philosophy, University of California San Diego

ABSTRACT OF THE DISSERTATION

Extremal Problems for Random Objects

by

Sam Spiro

Doctor of Philosophy in Mathematics

University of California San Diego, 2022

Professor Jacques Verstraëte, Chair

This dissertation lies at the intersection of extremal combinatorics and probabilistic combinatorics. Roughly speaking, *extremal combinatorics* studies how large a combinatorial object can be. For example, a classical result of Mantel's says that every n-vertex triangle-free graph has at most $\frac{1}{4}n^2$ edges. The area of *probabilistic combinatorics* encompasses both the application of probability to combinatorial problems, as well as the study of random combinatorial objects such as random graphs and random permutations.

In this dissertation we study three problems related to extremal properties of random objects.

First, we study the maximum score of a certain guessing game which uses a randomly shuffled deck of cards, and in doing so we solve a 40 year conjecture of Diaconis and Graham.

Next, we study the Turán problem in random hypergraphs. Specifically, we examine the function $\mathrm{ex}(H^r_{n,p}, \mathcal{F})$, which is the maximum number of hyperedges that an $\mathcal{F}$-free subgraph of the random hypergraph $H^r_{n,p}$ can have. By using a novel counting technique, we obtain effective bounds when $\mathcal{F}$ consists of a collection of Berge cycles.

Finally, we study thresholds of random graphs and hypergraphs, which essentially asks how large p must be in order for $H^r_{n,p}$ to contain a given structure. We give a common generalization of recent breakthrough work done by Alweiss, Lovett, Wu, and Zhang related to the Erdős sunflower problem; and of work by Kahn, Narayanan, and Park related to the threshold for squares of Hamiltonian cycles in $G_{n,p}$.

Chapter 1

Introduction

This thesis focuses on problems at the intersection of extremal combinatorics and probabilistic combinatorics. Roughly speaking, *extremal combinatorics* studies how large a parameter associated to a combinatorial object can be. For example, Mantel's Theorem [Man07] says that every n-vertex triangle-free graph has at most $\frac{1}{4}n^2$ edges, and this bound is best possible due to the complete bipartite graph $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$.

Roughly speaking, *probabilistic combinatorics* studies both the application of 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 [Erd47] used random graphs to give the first exponential lower bound for diagonal Ramsey numbers. Since then, the application of probability to problems in extremal combinatorics has been well developed into what is commonly known as *the probabilistic method*.

The focus of this thesis involves extremal problems for random objects. One classical problem of this form is the problem of determining the (expected) length of a longest increasing subsequence in a random permutation. This innocent sounding problem turns out to have deep connections to many areas of math such as representation theory, and it is the subject of the book by Romik [Rom15]. In a similar spirit, we investigate three problems which ask how large a given parameter related to a random object can be.

1.1 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 [BHJ57], casino games [EL05], and many other real-life problems; see [DGS20] for more information about applications.

We study how large or small of a score the player in this game can guarantee in expectation. Our main result answers a question of Diaconis and Graham from 1981 who asked if the expected score in the partial feedback model could be bounded uniformly in n. We answer this question in a strong form.

Theorem 1.1.1. *There exists a sufficiently large constant C such that regardless of their strategy, the player can guarantee at most $m+Cm^{3/4}\log^{1/4} m$ correct guesses in expectation in the partial feedback model whenever n is sufficiently large in terms of m.*

Note that the player can always guarantee a score of m by guessing the same card type every round, so this result is asymptotically best possible.

1.2 The Turán Problem in Random Hypergraphs

Szemerédi [Sze75] famously proved that any dense subset of the integers contains arbitrarily long arithmetic progressions. Building on this, Green and Tao [GT08] 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. Given this result for pseudorandom sets, it is natural to ask 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 [CG16] and Schacht [Sch16]. The methods used in [CG16, Sch16] extend to many other probabilistic variants of classical problems, and one such problem that we are interested in is finding large $\mathcal{F}$-free subgraphs of random graphs and hypergraphs.

A *hypergraph* H is a set of vertices $V(H)$ together with a set $E(H)$ of subsets of $V(H)$ which are called *edges* or *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. Given a set of r-graphs $\mathcal{F}$, we say that an r-graph H is *$\mathcal{F}$-free* if H does not contain a subgraph isomorphic to any $F \in \mathcal{F}$. We let $\mathrm{ex}(n, \mathcal{F})$ denote the maximum number of edges that an $\mathcal{F}$-free r-graph on n vertices can have. Determining $\mathrm{ex}(n, \mathcal{F})$ is known as *Turán's problem* and is one of the central problems in extremal combinatorics.

We are interested in a random variant of Turán's problem. We define the *random r-graph* $H^r_{n,p}$ to be the r-graph on n vertices obtained by including each possible hyperedge independently and with probability p, and when $r = 2$ we will denote this as $G_{n,p}$. For example, $G_{n,1}$ is the complete graph K_n since each possible edge is included with probability 1.

Let $\mathrm{ex}(H^r_{n,p}, \mathcal{F})$ denote the maximum number of edges of an $\mathcal{F}$-free subgraph of $H^r_{n,p}$. For example, when $p = 1$, the (deterministic) function $\mathrm{ex}(H^r_{n,1}, \mathcal{F})$ is the maximum number of hyperedges that an $\mathcal{F}$-free r-graph on n vertices can have, which is exactly Turán's problem. This problem has been essentially solved if $\mathcal{F}$ contains no r-partite r-graphs due to the work of Conlon and Gowers [CG16] and Schacht [Sch16], but the problem is wide open when $\mathcal{F}$ contains an r-partite r-graph.

In the graph setting, Morris and Saxton [MS16] essentially determined $\mathrm{ex}(n, F)$

when F is either a complete bipartite graph or an even cycle, assuming some well known conjectures in extremal graph theory. As is common in the area, their approach used the heavy machinery of hypergraph containers which was developed independently by Balogh, Morris, and Samotij [BMS15] and Saxton and Thomason [ST15]. We extend the results of Morris and Saxton to the hypergraph setting. In doing so, we develop a novel approach to counting $\mathcal{F}$-free hypergraphs that avoids the (explicit) use of containers. Our new approach relies heavily on turning counting problems for hypergraphs into counting problems for graphs. One result of this form is the following, where the girth of a hypergraph is defined formally in Section 3.

Theorem 1.2.1. *For $\ell, r \geq 3$ let $\mathrm{N}_m^r(n,\ell)$ denote the number of n-vertex r-graphs with m hyperedges and girth larger than ℓ. For $\lambda = \lceil (r-2)/(\ell-2) \rceil$ and all $m, n \geq 1$,*

$$\mathrm{N}_m^r(n,\ell) \leq \mathrm{N}_m^2(n,\ell)^{r-1+\lambda}. \tag{1.1}$$

1.3 Spread Hypergraphs

Given a property $\mathcal{P}$ that a graph can have, we say that a function $p = p(n)$ is a *threshold* for $\mathcal{P}$ if asymptotically almost surely (or a.a.s. for short), $G_{n,p'}$ satisfies $\mathcal{P}$ whenever $p'(n)/p(n) \to \infty$, and it fails to satisfy $\mathcal{P}$ a.a.s. whenever $p'(n)/p(n) \to 0$. For example, it is well known that $G_{n,p'}$ contains a Hamiltonian cycle a.a.s. if $p' \gg \log n/n$ and that it fails to do so a.a.s. if $p' \ll \log n/n$, so we say that $p(n) = \log n/n$ is a threshold for the property of containing a Hamiltonian cycle.

The study of thresholds is one of, if not the central problem, in the study of random graphs. It was proven by Bollobás and Thomason [BT87] that thresholds always exist for monotone properties $\mathcal{P}$, and in particular they always exist when considering the property of containing a given subgraph F. An elegant technique has recently been developed which can be used to find upper bounds on thresholds for $G_{n,p}$ containing a given subgraph F, provided the copies of F are sufficiently "spread out" in K_n.

To motivate these ideas, we observe that for any graph F, we can define an auxiliary hypergraph $\mathcal{H}_F$ whose vertex set consists of the edges of K_n and whose hyperedges consist of sets of edges which form a copy of F in K_n. Let V_p be a random subset of the vertices of $\mathcal{H}_F$ obtained by including each vertex independently and with probability p. If one unwinds the definitions, one sees the probability that $G_{n,p}$ contains a copy of F is exactly the probability that V_p contains a hyperedge of $\mathcal{H}_F$. Thus we have reduced the problem of studying thresholds in $G_{n,p}$ to studying when random subsets of hypergraphs contain a hyperedge.

A general technique for solving this latter problem was developed by Frankston, Kahn, Narayanan, and Park [FKNP21] based off of breakthrough work of Alweiss, Lovett, Wu, and Zhang [ALWZ20] regarding the sunflower conjecture. To state their result, if A is a set of vertices of a hypergraph $\mathcal{H}$, we define the *degree of* A to be the number of edges of $\mathcal{H}$ containing A, and we denote this quantity by $d_{\mathcal{H}}(A)$ or simply as $d(A)$ if $\mathcal{H}$ is understood. We say that a hypergraph $\mathcal{H}$ is *q-spread* if it is non-empty and if $d(A) \le q^{|A|}|\mathcal{H}|$ for all sets of vertices A. It was proven by Frankston, Kahn, Narayanan, and Park [FKNP21] that if $\mathcal{H}$ is a q-spread r-graph, then with high probability a random subset of $V(\mathcal{H})$ of size $Kq \log r$ contains a hyperedge of $\mathcal{H}$, where K is some absolute constant. From this result, one can obtain very short proofs of problems which were previously thought to be incredibly difficult. For example, in Theorem 4.2.1 we show how this result gives a less than one page proof of Shamir's problem, which asks for the threshold of $H^r_{n,p}$ containing a perfect matching. This problem was originally solved by Johansson, Kahn, and Vu [JKV08] using a significantly harder argument.

Using a variant of this technique, Narayanan, Kahn, and Park [KNP21] determined the threshold of the square of a Hamiltonian cycle, which was a longstanding open problem of Kühn and Osthus [KO12]. During a talk, Naryanan asked the following question:

Question 1.3.1. *Is it possible to give a common generalization of the proofs of [ALWZ20, FKNP21, KNP21]?*

We answer this question in the positive, giving a smooth interpolation between the conditions needed for variants of q-spread techniques to apply.

1.4 Notation

We say a sequence of events A_n occurs *asymptotically almost surely* or *a.a.s.* if $\lim_{n\to\infty} \Pr[A_n] = 1$. Throughout 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. We write $f \sim g$ if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$. Occasionally in our exposition we will write $f \gg g$ to informally mean that f is significantly larger than g.

Chapter 2

Card Guessing with Feedback

2.1 Introduction

Let $\mathfrak{S}_{m,n}$ be the set of words π over the alphabet $[n] := \{1, 2, \ldots, n\}$ where each character in $[n]$ appears exactly m times in π. We think of π as some way to shuffle a deck of cards which has m suits and n card types. For example, a standard deck of 52 cards has $n = 13$ values (Ace, Two, ..., King), each appearing $m = 4$ times. We find it helpful to think that m stands for <u>m</u>ultiplicity and n is for <u>n</u>umber of values. We refer to the elements of $\mathfrak{S}_{m,n}$ as permutations, even though for $m > 1$ this is technically not the case. If X is a finite set, we write $\boldsymbol{x} \sim X$ to indicate that $\boldsymbol{x}$ is chosen uniformly at random from X.

Consider the following experiment: a deck with m copies of n different card types is randomly shuffled according to some $\boldsymbol{\pi} \sim \mathfrak{S}_{m,n}$, and a guesser attempts to guess each card as it is drawn, and the drawn card is discarded after the guess is made (i.e. this is sampling without replacement). Each time a guess is made, some amount of "feedback" is given. For example, one could tell the guesser the true identity of the card they just guessed (this is called the *complete feedback model*) or they could be told nothing at all (the *no feedback model*). This can also be viewed as a one player game where the guesser tries to either maximize or minimize the number of times their guesses are correct, and we will often refer to these models as games.

These sorts of models were considered by Blackwell and Hodges [BHJ57] and Efron [Efr71] in relation to clinical trials. For example, in a medical trial comprising 4 treatments and 100 subjects, a deck of 100 cards with 25 cards labeled by each treatment is prepared. Subjects are assigned to treatments as they come into the clinic, sequentially, using the next card (which is then discarded). Hospital staff have the option of ruling subjects ineligible as they come in. If the staff has strong opinions about the efficacy of treatments and observes which treatments have already been given out, they may guess what the next treatment is and bias the experiment by ruling a sickly subject ineligible. It is thus of interest to be able to evaluate the expected potential bias.

Card guessing is also a mainstay of classical experiments to test "Extra Sensory Perception" (ESP). The most common experiment utilizes a deck of 25 cards where there are five copies of five different types of cards (so $m = n = 5$ in our language) where the subjects iteratively try and guess the identity of the next card, and experimenters routinely give various kinds of feedback to enhance "learning". Diaconis [Dia78] and

Diaconis and Graham [DG81] give a review of these problems.

In the no feedback model every strategy guesses m cards correctly in expectation. The distribution of correct guesses depends on the guessing strategy: if the guesser always guesses the same card type then the variance is 0, and it can be shown that the variance is largest if the guesser uses a permutation of the mn values, see [DG81].

The complete feedback model is more complicated, but optimal strategies were determined in [DG81]. Given a strategy $\mathcal{G}$ for the guesser, let $C(\mathcal{G}, \pi)$ denote the number of correct guesses the guesser gets in the complete feedback model if they use strategy $\mathcal{G}$ and the deck is shuffled according to π. Let $\mathcal{C}^+_{m,n} = \max_{\mathcal{G}} \mathbb{E}[C(\mathcal{G}, \boldsymbol{\pi})]$, where $\boldsymbol{\pi} \sim \mathfrak{S}_{m,n}$ and the maximum ranges over all possible strategies $\mathcal{G}$. Similarly define $\mathcal{C}^-_{m,n} = \min_{\mathcal{G}} \mathbb{E}[C(\mathcal{G}, \boldsymbol{\pi})]$. The following is proven in [DG81].

Theorem 2.1.1 ([DG81])**.** *If $\mathcal{G}^+$ (respectively $\mathcal{G}^-$) is the strategy where one guesses a most likely (respectively least likely) card at each step, then $\mathcal{C}^\pm_{m,n} = \mathbb{E}[C(\mathcal{G}^\pm, \boldsymbol{\pi})]$. Moreover,*

$$\mathcal{C}^\pm_{m,n} = m \pm M_n\sqrt{m} + o_n(\sqrt{m}),$$

where $M_n = \Theta(\sqrt{\log n})$ is the expected maximum value of n independent standard normal variables.

The main focus in this chapter is on a feedback model called the partial feedback model, which returns an intermediate amount of information to the guesser. After each guess, the guesser is only told whether their guess was correct or not (and thus not the identity of the card if they were incorrect). This feedback protocol was recommended when conducting ESP trials and is a natural notion of bias if card guessing experiments are performed with experimenter and subject in the same room. Given a strategy $\mathcal{G}$ for this game, let $P(\mathcal{G},\pi)$ denote the number of cards the guesser guesses correctly using strategy $\mathcal{G}$ if the deck is shuffled according to π, and define $\mathcal{P}^+_{m,n} = \max_{\mathcal{G}} \mathbb{E}[P(\mathcal{G},\boldsymbol{\pi})]$ and $\mathcal{P}^-_{m,n} = \min_{\mathcal{G}} \mathbb{E}[P(\mathcal{G},\boldsymbol{\pi})]$ for the maximum and minimum expected number of correct guesses possible, respectively.

The partial feedback model is significantly more difficult to analyze than the other two models, and relatively little is known about it. This is in large part due to the fact that we do not understand the optimal strategy in this game, and in particular it is not the case that the strategy $\mathcal{G}^+$ of guessing a maximum likelihood card satisfies $\mathbb{E}[P(\mathcal{G}^+,\boldsymbol{\pi})] = \mathcal{P}^+_{m,n}$ for $m \geq 2$, see [DG81].

Define $\mathcal{N}^\pm_{m,n}$ for the no feedback model analogous to how $\mathcal{C}^\pm_{m,n}$ and $\mathcal{P}^\pm_{m,n}$ were defined, and note that $\mathcal{N}^\pm_{m,n} = m$. One can easily show that $\mathcal{N}^+_{m,n} \leq \mathcal{P}^+_{m,n} \leq \mathcal{C}^+_{m,n}$ for all m and n, with the reverse inequalities holding for $-$ instead of $+$. In particular, by Theorem 2.1.1 and the fact that $\mathcal{N}^\pm_{m,n} = m$, we obtain $\mathcal{P}^\pm_{m,n} = (1+o(1))m$ as m goes to infinity, for any fixed n. Given this, our focus will be in bounding $\mathcal{P}^\pm_{m,n}$ when m is fixed and n is large. As a point of comparison, we first establish the value of $\mathcal{C}^\pm_{m,n}$ in this regime. Here and throughout this chapter we let log denote the natural logarithm.

Theorem 2.1.2 ([DG81, DGHS21, HO21])**.** *For m fixed and $n \to \infty$, we have*

$$\mathcal{C}^+_{m,n} \sim H_m \log n,$$

where $H_m = \sum_{i=1}^m j^{-1}$ *is the* m*-th harmonic number, and*

$$\mathcal{C}^-_{m,n} \sim \Gamma\left(1 + \frac{1}{m}\right) n^{-1/m},$$

where $\Gamma(x)$ *denotes the gamma function.*

Theorem 2.1.2 shows that for any fixed m, in expectation the guesser can achieve arbitrarily many or arbitrarily few correct guesses as n grows in the complete feedback model. In sharp contrast, we show that the guesser cannot obtain arbitrarily many correct guesses in the partial feedback model.

Theorem 2.1.3. *If* n *is sufficiently large in terms of* m*, we have*

$$\mathcal{P}^+_{m,n} = m + O(m^{3/4} \log^{1/4} m).$$

This resolves a 40 year old problem of Diaconis and Graham [DG81], which was open even for $m = 2$ (i.e. a deck with composition $\{1, 1, 2, 2, \ldots, n, n\}$). In particular, this shows that the information from the partial feedback model is not enough for the guesser to correctly guess asymptotically more cards compared to when they are given no feedback at all. We suspect that the error term in Theorem 2.1.3 can be improved to $m^{1/2+o(1)}$, which would be best possible; see the discussion in Section 2.3.

We conclude this introduction with some brief remarks about the related literature. In the partial feedback model, the enumeration of the number of permutations consistent with a given sequence of guesses can be reduced to the evaluation of certain permanents, see Chung, Diaconis, Graham, and Mallows [CDGM81] and Diaconis, Graham, and Holmes [DGH01]. These papers contain applications to the partial feedback model, as well as a fascinating "persistence conjecture": whenever the guesser guesses a card type i incorrectly, it is optimal for them to continue to guess i in the next step.

Throughout this chapter, we focus on evaluating the expected number of correct guesses. The distribution of the number of correct guesses is treated in [DG81], see also Proschan [Pro91]. A variety of other feedback mechanisms have also been explored, such as less feedback if the guesser is doing well, and telling the guesser that their guess is "high" or "low", see Samaniego and Utts [SU83].

Our evaluation for these models gives one point for each card guessed correctly. It is also natural to consider weighted scores: a correct guess early on might be weighted more heavily than a correct guess towards the end since more information is available to the guesser later on. This is known as skill scoring and is discussed in [DG81].

2.2 Proof of Theorem 2.1.3

2.2.1 Definitions and Outline

Throughout this section we fix a guessing strategy $\mathcal{G}$ and a suitable $\varepsilon = \varepsilon(m) > 0$ which will be on the order of $m^{-1/4} \log^{1/4} m$. Our goal is to prove for large enough n that $\mathbb{E}[P(\mathcal{G}, \boldsymbol{\pi})] \leq (1 + O(\varepsilon))m$. In this section, we simply refer to the partial feedback model as "the game."

A *history* $h = (g, y)$ of a completed game is a pair of vectors: the $[n]$-valued vector g of all mn guesses made throughout the game, and the boolean vector y of feedback received, so that $y_t = 1$ if and only if the t-th card in the deck has value g_t. A *history at time* t, denoted h_t, is a truncation of some complete history h to the first t values in each vector, representing all the information available to the guesser after they make the t-th guess.

We let H denote a sample of the history of the game given the fixed strategy $\mathcal{G}$ and that the deck is shuffled according to a uniform random $\boldsymbol{\pi} \sim \mathfrak{S}_{m,n}$. Similarly H_t denotes a sample of the history of the game at time t.

Given a history $h = (g, y)$, we write $Y(h) := \|y\|$ for the total number of correct guesses, where here and throughout this chapter $\|v\| := \sum |v_i|$ denotes the ℓ^1 norm. Define $a_i(h) := |\{t : g_t = i\}|$ to be the number of times card type i has been guessed, and $m_i(h) := m - |\{t : g_t = i \text{ and } y_t = 1\}|$ to be the number of copies of card i left to be found in the deck. For a partial history h_t, the values $Y(h_t)$, $a_i(h_t)$, and $m_i(h_t)$ are defined in the same way.

We are ready to outline the proof. The first and most important step is to prove the following "pointwise" lemma, which roughly shows that for all typical histories h_{t-1}, the probability that the t-th guess is correct is at most $(1 + o(1))n^{-1}$.

Lemma 2.2.1. *For any history h_{t-1} of the game up to time $t - 1$ and any $i \in [n]$,*

$$\Pr[\boldsymbol{\pi}_t = i | H_{t-1} = h_{t-1}] \leq \frac{m_i(h_{t-1})}{mn - a_i(h_{t-1}) - Y(h_{t-1})}.$$

Note that the fraction on the right hand side is a natural estimate for $\Pr[\boldsymbol{\pi}_t = i | H_{t-1} = h_{t-1}]$: the numerator is exactly the number of copies of i in the deck that have yet to be found, and the denominator is approximately the total number of positions among $[mn]$ at which such a copy could lie (this may not be exact because $a_i(h_{t-1})$ and $Y(h_{t-1})$ can count the same position twice). We use a simple bijective argument to prove Lemma 2.2.1 in Section 2.2.2.

The second step of the proof is to show that the term $Y(H_{t-1})$ in Lemma 2.2.1 is negligible with high probability, which is done by the following lemma.

Lemma 2.2.2. *For any* $0 < \lambda \leq 1/6$, $n^{1/2} \geq 12\lambda^{-1}$, *and any fixed strategy* $\mathcal{G}$,

$$\Pr[P(\mathcal{G}, \boldsymbol{\pi}) > \lambda mn] \leq 2e^{-mn^{1/2}}.$$

This bound is proved in Section 2.2.3 using Lemma 2.2.1 and Chernoff bounds. Combining Lemmas 2.2.1 and 2.2.2, and since $Y(h_{t-1}) \leq Y(h)$, we see that with high probability for any $\varepsilon > 0$,

$$\Pr[\boldsymbol{\pi}_t = i] \leq \frac{m_i(H_{t-1})}{(1-\varepsilon)mn - a_i(H_{t-1})}.$$

We now break guesses into three types, based on how many times a given card i has already been guessed. A guess at time t, say with $g_t = i$, is called *subcritical* if $a_i(H_{t-1}) < \varepsilon mn$, *critical* if $\varepsilon mn \leq a_i(H_{t-1}) < (1-\varepsilon)mn$, and *supercritical* if $a_i(H_{t-1}) \geq (1-\varepsilon)mn$. Note that if even a single supercritical guess is made, then almost all guesses must have been of that same card type, which makes the situation easy to analyze.

By adaptively re-numbering the cards during the game if necessary, we may assume without loss of generality that if there are k card types for which critical guesses are made, then they are exactly the first k cards $1, \ldots, k$. For any given history h, let $b_0(h)$ be the number of subcritical guesses made, let $b_i(h)$, $1 \le i \le k$ be the number of critical guesses made with $g_t = i$, and let $b_\infty(h)$ be the number of supercritical guesses made. Define $Y_0(h)$, $Y_i(h)$, and $Y_\infty(h)$ to be the number of correct guesses made in each regime.

We finish the proof by showing with high probability that each of the $Y_i(H)$ values are not much larger than their means. The subcritical guesses $Y_0(H)$ are handled in Section 2.2.4, the critical guesses $Y_i(H)$ in Section 2.2.5, and the supercritical regime is simple enough to not merit its own subsection. The proof is then completed in Section 2.2.6.

Throughout the proof we will often omit floors and ceilings for ease of presentation. For an event E we let $\overline{E}$ denote its complement. For real valued random variables X and Y, we write $X \succeq Y$ if X stochastically dominates Y, i.e. if for all $x \in \mathbb{R}$, $\Pr[X \ge x] \ge \Pr[Y \ge x]$. We also recall a standard variant of the Chernoff bound, see for instance [KQ21].

Lemma 2.2.3. *Let $B(N,p)$ be a binomial random variable with N trials and probability of success p. Then for all $\lambda > 0$,*

$$\Pr[B(N,p) > (1+\lambda)pN] \le e^{-\frac{\lambda^2 pN}{2+\lambda}}.$$

2.2.2 The Pointwise Lemma

In this section we show Lemma 2.2.1, which is equivalent to an upper bound on the number of $\pi \in \mathfrak{S}_{m,n}$ for which at each position up through t, either π_t is specified or a single value is disallowed for π_t. We reduce to the following setup.

Definition 2.2.4. *Let* $\mathbf{m} = (m_1, \ldots, m_n)$ *and* $\mathbf{a} = (a_1, \ldots, a_n)$ *be vectors of nonnegative integers satisfying* $\|\mathbf{a}\| < \|\mathbf{m}\|$. *An* $\mathbf{m}$-permutation *is a word of length* $\|\mathbf{m}\|$ *over alphabet* $[n]$ *where* i *appears exactly* m_i *times. An* $(\mathbf{m}, \mathbf{a})$*-permutation* π *is an* $\mathbf{m}$*-permutation where the first* a_1 *terms are not* 1*, the next* a_2 *terms are not* 2*, and so on, so that exactly* a_i *terms in* π *are forbidden from taking value* i*.*

It is significant that $\|a\| < \|m\|$ strictly in the definition of $(\mathbf{m}, \mathbf{a})$-permutations, guaranteeing that no restrictions are made on the value of the last term. Given a history h_{t-1} up to time $t-1$, we let $\mathbf{m}$ be the vector $(m_1(h_{t-1}), \ldots, m_n(h_{t-1}))$, and $\mathbf{a}$ be the vector $(a_1(h_{t-1}), \ldots, a_n(h_{t-1}))$. We claim that the following bound on $(\mathbf{m}, \mathbf{a})$-permutations implies Lemma 2.2.1.

Lemma 2.2.5. *If* $f_i(\mathbf{m}, \mathbf{a})$ *is the fraction of all* $(\mathbf{m}, \mathbf{a})$*-permutations for which the last term is* i*, then*

$$f_i(\mathbf{m}, \mathbf{a}) \leq \frac{m_i}{\|\mathbf{m}\| - a_i}.$$

Indeed, by definition $f_i(\mathbf{m}, \mathbf{a})$ is the probability that the last card in $\boldsymbol{\pi}$ is exactly i given the current history h_{t-1}. But all positions past the first $t-1$ are indistinguishable, so $f_i(\mathbf{m}, \mathbf{a})$ is also the probability that the next card (at index t) is i. Thus it suffices to prove Lemma 2.2.5.

Proof of Lemma 2.2.5. It suffices to show the lemma for $i = 1$. First we make a technical reduction to the case $a_1 = 0$ for convenience. Let $\tilde{\pi}$ be any sequence of a_1 cards in which

1 does not appear and i appears at most m_i times for all $i > 1$. Define an $(\mathbf{m}, \mathbf{a}, \tilde{\pi})$-*permutation* to be an $(\mathbf{m}, \mathbf{a})$-permutation where the first a_1 terms agree with $\tilde{\pi}$.

Define $f_i(\mathbf{m}, \mathbf{a}, \tilde{\pi})$ to be the fraction of $(\mathbf{m}, \mathbf{a}, \tilde{\pi})$-permutations which have last term i. Since $f_1(\mathbf{m}, \mathbf{a})$ is some convex combination of the values $f_1(\mathbf{m}, \mathbf{a}, \tilde{\pi})$, it suffices to show that for every specific choice of $\tilde{\pi}$,

$$f_1(\mathbf{m}, \mathbf{a}, \tilde{\pi}) \leq \frac{m_1}{\|\mathbf{m}\| - a_1}. \tag{2.1}$$

Let $\mathbf{m}'$ be the vector of card counts remaining when the cards in $\tilde{\pi}$ are taken out, and let $\mathbf{a}' = (0, a_2, a_3, \ldots, a_n)$, so that an $(\mathbf{m}, \mathbf{a}, \tilde{\pi})$-permutation is just $\tilde{\pi}$ concatenated with an $(\mathbf{m}', \mathbf{a}')$-permutation π'. Since $m_1' = m_1$ and $\|\mathbf{m}'\| = \|\mathbf{m}\| - a_1$, it suffices to show

$$f_1(\mathbf{m}', \mathbf{a}') \leq \frac{m_1'}{\|\mathbf{m}'\|},$$

which is just the case $a_1 = 0$ in the original lemma statement. Thus, it remains to show that if $a_1 = 0$, we have

$$f_1(\mathbf{m}, \mathbf{a}) \leq \frac{m_1}{\|\mathbf{m}\|}. \tag{2.2}$$

In fact, we will prove that for any i,

$$\frac{f_1(\mathbf{m}, \mathbf{a})}{f_i(\mathbf{m}, \mathbf{a})} \leq \frac{m_1}{m_i}. \tag{2.3}$$

The case $i = 1$ is trivial, so we just need to prove this for $i > 1$, and without loss of generality we can assume $i = 2$. We divide the $(\mathbf{m}, \mathbf{a})$-permutations π which end in either 1 or 2 into classes as follows. For each π which ends in either 1 or 2, consider all positions past the first a_2 which contain either a 1 or a 2. Let $S(\pi)$ denote the set of π'

obtained by cyclically shifting the 1's and 2's in these positions within π, fixing all other values. Note that with this we never move a 1 into a forbidden position (as $a_1 = 0$) nor a 2 into a forbidden position (as we only shift past the first a_2 positions). It follows that every $\pi' \in S(\pi)$ is a $(\mathbf{m}, \mathbf{a})$-permutation ending in 1 or 2.

Note that the total number of 2's past the first a_2 positions is exactly m_2, since every 2 appears past the first a_2, while the total number of 1's past the first a_2 positions is at most m_1, since there are exactly m_1 1's in total. Thus, we see that the fraction of $\pi' \in S(\pi)$ which end in 1 is at most $\frac{m_1}{m_1+m_2}$ for every π. As the $S(\pi)$ partition all possible $(\mathbf{m}, \mathbf{a})$-permutations π which end in either 1 or 2, (2.3) follows for $i = 2$.

Finally, to derive (2.2) it suffices to write (2.3) as

$$\frac{m_i}{m_1} f_1(\mathbf{m}, \mathbf{a}) \leq f_i(\mathbf{m}, \mathbf{a})$$

and sum over i, noting that $\sum_i f_i(\mathbf{m}, \mathbf{a}) = 1$ since every $(\mathbf{m}, \mathbf{a})$-permutation must end in some i. □

2.2.3 Weak Bound on $\mathcal{P}^+_{m,n}$

The next step is to show that the $Y(h_{t-1})$ term in Lemma 2.2.1 is negligible with high probability. Since $Y(h_{t-1})$ is bounded by just $Y(h)$, the total number of cards guessed correctly, it suffices to show a weak upper bound on the total number of correct guesses in the form of Lemma 2.2.2. To do this we first show the following.

Lemma 2.2.6. *Let $B_1, \ldots, B_k$ be (not necessarily independent) Bernoulli random vari-*

ables with $\Pr[B_t = 1 | \sum_{s<t} B_s = x] \le p$ *for all* t *and* x*. Then* $\sum_{t=1}^{k} B_t$ *is stochastically dominated by a binomial random variable* $B(k, p)$*.*

This lemma will be proved by induction. The induction step is the following simple observation.

Lemma 2.2.7. *Let* X, X', Y, Y' *be integer-valued random variables such that* X' *and* Y' *are* $\{0, 1\}$*-valued,* $X \succeq Y$*, and for all* $x \in \mathbb{Z}$*,* $(X'|X = x) \succeq (Y'|Y = x)$*. Then,*

$$X + X' \succeq Y + Y'.$$

Proof. Our goal is to show that for any $y \in \mathbb{Z}$, $\Pr[X + X' \ge y] \succeq \Pr[Y + Y' \ge y]$. But clearly

$$\begin{aligned}
\Pr[X + X' \ge y] &= \Pr[X \ge y] + \Pr[(X = y - 1) \wedge (X' = 1)] \\
&= \Pr[X \ge y] + \Pr[X = y - 1] \Pr[X' = 1 | X = y - 1] \\
&\ge \Pr[X \ge y] + \Pr[X = y - 1] \Pr[Y' = 1 | Y = y - 1] \\
&\ge \Pr[Y \ge y] + \Pr[Y = y - 1] \Pr[Y' = 1 | Y = y - 1] \qquad (2.4) \\
&= \Pr[Y + Y' \ge y].
\end{aligned}$$

Here only (2.4) is worth explaining. Since $X \succeq Y$ we have $\Pr[X \ge y] \ge \Pr[Y \ge y]$ and $\Pr[X \ge y] + \Pr[X = y - 1] \ge \Pr[Y \ge y] + \Pr[Y = y - 1]$, so by taking convex combinations of these two inequalities, we have for any $t \in [0, 1]$, $\Pr[X \ge y] + t \Pr[X = y - 1] \ge \Pr[Y \ge y] + t \Pr[Y = y - 1]$ as well. Taking $t = \Pr[Y' = 1 | Y = y - 1]$ completes the proof. $\square$

Lemma 2.2.6 follows by iterating Lemma 2.2.7 with $X = \sum_{s<t} B_s$, $X' = B_t$, Y a binomial random variable $B(t-1,p)$, and Y' a Bernoulli random variable with probability p. We omit the details.

We next prove the following, which immediately implies Lemma 2.2.2.

Lemma 2.2.8. *For any* $0 < \lambda \le 1/6$, $n \ge 200\lambda^{-1}$, *and any fixed strategy* $\mathcal{G}$,

$$\Pr[P(\mathcal{G}, \boldsymbol{\pi}) > \lambda mn] \le 2e^{-\lambda mn/12}.$$

Proof. We first show that few correct guesses are made in the first third of the game, i.e. when $t \le mn/3$. In this case we apply Lemma 2.2.1 to find that for any $i \in [n]$,

$$\Pr[\boldsymbol{\pi}_t = i | H_{t-1} = h_{t-1}] \le \frac{m_i(h_{t-1})}{mn - a_i(h_{t-1}) - Y(h_{t-1})} \le \frac{m}{mn - mn/3 - mn/3} = \frac{3}{n},$$

since up to this point there have been at most $mn/3$ correct guesses and each i has been guessed at most $mn/3$ times. It follows that for $t \le mn/3$, conditional on any h_{t-1}, the probability that the t-th guess is correct is at most $3/n$. In particular the t-th guess is correct with probability at most $3/n$ regardless of the value of $Y(H_{t-1})$, so by Lemma 2.2.6 the number of correct guesses in the first third of the game $Y(H_{mn/3})$ is stochastically dominated by a binomial random variable $B(mn/3, 3/n)$. Applying Lemma 2.2.3 gives for all $\delta \ge 2$,

$$\Pr[Y(H_{mn/3}) > (1+\delta)m] \le \Pr[B(mn/3, 3/n) > (1+\delta)m] \le e^{-\delta m/2}.$$

Taking $\delta = \lambda n/4 - 1 \ge \lambda n/6 \ge 2$ since $n \ge 12\lambda^{-1}$, we find

$$\Pr[Y(H_{mn/3}) > \lambda mn/4] \le e^{-\lambda mn/12}. \tag{2.5}$$

Let T be the set of i such that $a_i(h_t) < mn/4$ for all t, and note that there are at most four card types not in T (since only mn total guesses are made). Let E be the event that $Y(H_{mn/3}) \leq \lambda mn/4$, and observe that conditional on E we have $Y(H_t) \leq (2/3 + \lambda/4)mn$ for all t since at most $2mn/3$ correct guesses can be made in the last $2/3$ of the game. Thus by Lemma 2.2.5 and the above observations, we have for $i \in T$, all $t > mn/3$, and any possible history h_{t-1} for which E occurs,

$$\Pr[\boldsymbol{\pi}_t = i | H_{t-1} = h_{t-1}] \leq \frac{m}{mn - mn/4 - (2/3 + \lambda/4)mn} \leq \frac{24}{n}, \tag{2.6}$$

where we used $\lambda \leq 1/6$.

Let $Y'(H)$ denote the total number of correct guesses of card types $i \in T$ and let $Y''(H)$ denote the total number of correct guesses involving $i \notin T$. Observe that

$$Y(H) = Y'(H) + Y''(H) \leq Y'(H) + 4m \leq Y'(H) + \lambda mn/2,$$

where this last step used $n \geq 8\lambda^{-1}$ (which is implicit in our hypothesis of the lemma). By (2.6) we see that conditional on E, $Y'(H) - Y(H_{mn/3})$ is stochastically dominated by a binomial random variable $B(2mn/3, 24/n)$. Thus

$$\begin{aligned}\Pr[Y(H) > \lambda mn] \leq \Pr[Y'(H) > \lambda mn/2] &\leq \Pr[Y'(H) - Y(H_{mn/3}) > \lambda mn/4 | E] + \Pr[\overline{E}] \\ &\leq \Pr[B(2mn/3, 24/n) > \lambda mn/4] + \Pr[\overline{E}] \leq e^{-\lambda mn/12} + e^{-\lambda mn/12},\end{aligned}$$

where the last inequality used the Chernoff bound with $\delta = \lambda n/64 - 1 \geq \lambda n/96 \geq 2$ and (2.5). $\square$

2.2.4 Concentration of Subcritical Guesses

In this section we handle the subcritical guesses. If X_t denotes the indicator variable that the t-th subcritical guess is correct, then intuitively the X_t variables are dominated by Bernoulli random variables with parameter $p = \frac{1}{(1-2\varepsilon)n}$, so the total number of correct subcritical guesses is dominated by a binomial distribution $B(b_0(H), p)$, where we recall that $b_0(H)$ is the number of subcritical guesses in history H.

We would like to say that this binomial distribution is close to its expectation with high probability. It is not enough, however, to prove this for a fixed binomial distribution. The main technical issue is that the number of trials $b_0(H)$ can be chosen adaptively by the guesser. For example, they can use a strategy where they repeatedly make subcritical guesses until they have guessed an above average number of cards correctly. This is essentially equivalent to the guesser simulating a summation of Bernoulli random variables $\sum_{t=1}^{mn} B_t$, and then choosing some number of trials $b \leq mn$ such that the number of correct subcritical guesses is $\sum_{t=1}^{b} B_t$. We thus wish to show that for B_t a sequence of independent Bernoulli variables, $\sum_{t=1}^{b} B_t$ is not much larger than its expectation for all large b. With this, no matter how the guesser chooses b, they can never do much better than pb.

A weak upper bound for this probability comes from applying the Chernoff bound to all $b \leq mn$ and then using a union bound. Unfortunately when p is very small this upper bound is not effective. A more careful application of the union bound gives the following technical result, where we think of the Z_k's in its statement as centered binomial random variables with k trials.

Lemma 2.2.9. *Let $0 \leq p \leq 1$, $c, c' > 0$, and let $0 \equiv Z_0, Z_1, Z_2, \ldots$ be random variables such that $Z_k - Z_{k-1} \geq -p$ for all k, and such that for all integers $0 \leq k' < k$ and all $0 < \lambda \leq 1$,*

$$\Pr[Z_k - Z_{k'} > \lambda p(k - k')] \leq c' e^{-c\lambda^2 p(k-k')}.$$

Then for all $0 < \lambda \leq 1$ and integers $k_1 \geq k_0 \geq 2\lambda^{-1}$, we have

$$\Pr[\exists k \in [k_0, k_1], Z_k > \lambda pk] \leq \frac{8c' k_1}{\lambda k_0} e^{-\frac{1}{256} c\lambda^3 pk_0}.$$

Proof. Define $\ell = \frac{1}{2}\lambda k_0 \geq 1$. The idea of the proof is to take a union bound over the events $Z_{\ell a} - Z_{\ell(a-1)} > \lambda p\ell$ for all integers $a \leq \frac{k_1}{\ell}$, which will turn out to be strong enough to conclude the stated result. To be precise, let $0 = x_0 < x_1 < \cdots < x_r = k_1$ be any sequence of integers such that $\frac{1}{2}\ell \leq x_a - x_{a-1} \leq \ell$ for all $a > 0$, and note that the number of terms in this sequence satisfies

$$r \leq \lceil 2k_1/\ell \rceil \leq \frac{8k_1}{\lambda k_0}. \tag{2.7}$$

Let E be the event that $Z_{x_b} > \frac{1}{8}\lambda pb\ell$ for some b. Observe that $Z_{x_b} = \sum_{a=1}^{b} Z_{x_a} - Z_{x_{a-1}}$, so $Z_{x_b} > \frac{1}{8}\lambda pb\ell$ implies that some $a \leq b$ has

$$Z_{x_a} - Z_{x_{a-1}} > \frac{1}{8}\lambda p\ell \geq \frac{1}{8}\lambda p(x_a - x_{a-1}).$$

Thus by the union bound, the hypothesis of the lemma, the fact that $x_a - x_{a-1} \geq \frac{1}{2}\ell$, and inequality (2.7), we have

$$\Pr[E] \leq \sum_{a=1}^{r} \Pr[Z_{x_a} - Z_{x_{a-1}} > \frac{1}{8}\lambda p(x_a - x_{a-1})] \leq r \cdot c' e^{-\frac{1}{128} c\lambda^2 p\ell} \leq \frac{8c' k_1}{\lambda k_0} e^{-\frac{1}{256} c\lambda^3 pk_0}.$$

We claim that if $Z_k > \lambda pk$ for some $k \in [k_0, k_1]$, then E occurs. Indeed, suppose such a k exists and let b be the smallest integer such that $k \le x_b$, which in particular implies $x_b - k \le \ell$. We also have $b \ge 2$ because $k_0 \le k \le b\ell$ and $k_0/\ell = 2\lambda^{-1} \ge 2$, and thus

$$k \ge \frac{1}{2}(b-1)\ell \ge \frac{1}{4}b\ell. \tag{2.8}$$

Note that $Z_{x_b} - Z_k \ge -\ell p$ because $Z_k - Z_{k-1} \ge -p$ for all k. Using this, $\ell = \frac{1}{2}\lambda k_0 \le \frac{1}{2}\lambda k$, and inequality (2.8), we have

$$Z_{x_b} > \lambda pk - \ell p \ge \frac{1}{2}\lambda pk \ge \frac{1}{8}pb\ell,$$

so E occurs. Thus,

$$\Pr[\exists k \in [k_0, k_1], Z_k > \lambda pk] \le \Pr[E] \le \frac{8c'k_1}{\lambda k_0} e^{-\frac{1}{256}c\lambda^3 pk_0}$$

as desired. □

Using Lemma 2.2.9, we can show that subcritical guesses are well behaved.

Lemma 2.2.10. *If $\varepsilon \le \frac{1}{8}$ and n is sufficiently large in terms of ε, m, then*

$$\Pr\left[Y_0(H) > (1+4\varepsilon)\frac{b_0(H)}{n}\right] \le c'\varepsilon^{-2}e^{-c\varepsilon^4 m},$$

for some absolute constants $c, c' > 0$.

Proof. Given $t \le b_0(H)$, let t' be the smallest positive integer for which $b_0(H_{t'}) = t$, so that t' is the time of the t-th subcritical guess (note that t' is itself a random variable), and let $X_t \coloneqq Y_0(H_{t'}) - Y_0(H_{t'-1})$. In other words, X_t is the indicator of the t-th subcritical guess.

Let E be the event that $Y(H) > \varepsilon mn$, and define E_t to be the event that $Y(H_{t'}) > \varepsilon mn$. Observe that $\overline{E}$ implies that no E_t occurs.

Note that $Y_0(H) = \sum_{t=1}^{b_0(H)} X_t$. We modify $Y_0(H)$ to ignore the events E_t as follows. Define $X'_t = X_t$ if E_{t-1} does not occur and $X'_t = 0$ otherwise, and let $Y'_0 = \sum_{t=1}^{b_0(H)} X'_t$. With $L := (1+4\varepsilon)\frac{b_0(H)}{n}$, we find

$$\begin{aligned}\Pr[Y_0(H) > L] &\le \Pr[(Y_0(H) > L) \wedge \overline{E}] + \Pr[E] = \Pr[(Y'_0 > L) \wedge \overline{E}] + \Pr[E] \\ &\le \Pr[Y'_0 > L] + \Pr[E],\end{aligned}$$

By Lemma 2.2.2 we know $\Pr[E] \le 2e^{-mn^{1/2}}$, so for n sufficiently large the contribution of $\Pr[E]$ is negligible. It remains to upper bound the probability that Y'_0 is large. Note that $X'_t = 1$ if and only if the next term $\boldsymbol{\pi}_{t'}$ is exactly the next guess i, the total number $a_i(H_{t'-1})$ of times i is guessed is at most εmn, and the total number $Y(H_{t'-1})$ of correct guesses up to this point is also at most εmn. We now have by Lemma 2.2.1 that

$$\Pr[X'_t = 1 | X'_1, \ldots, X'_{t-1}] \le \frac{m_i(H_{t'-1})}{mn - a_i(H_{t'-1}) - Y(H_{t'-1})} \le \frac{m}{(1-2\varepsilon)mn} = \frac{1}{(1-2\varepsilon)n} =: p.$$

Define $B_1, B_2, \ldots, B_{mn}$ to be independent Bernoulli random variables with $\Pr[B_t = 1] = p$ and define $Z_k = \sum_{t=1}^{k} B_t - pk$. By the above inequality, we see that given any history $h_{t'-1}$ up to the t'-th guess, X'_t is stochastically dominated by B_t, and hence Z_k stochastically dominates $\sum_{t=1}^{k} X'_t - pk$. Observe that

$$\sum_{t=1}^{b_0(H)} X'_t - pb_0(H) > \frac{\varepsilon b_0(H)}{(1-2\varepsilon)n} \iff Y'_0 > \frac{(1+\varepsilon)b_0(H)}{(1-2\varepsilon)n} \Longleftarrow Y'_0 > \frac{(1+4\varepsilon)b_0(H)}{n} = L,$$

where the last step used $\varepsilon \le \frac{1}{8}$. Because $Z_{b_0(H)}$ stochastically dominates the above sum, we have

$$\Pr[Y_0' > L] \le \Pr\left[\sum_{t=1}^{b_0(H)} X_t' - pb_0(H) > \frac{\varepsilon b_0(H)}{(1-2\varepsilon)n}\right] \le \Pr\left[Z_{b_0(H)} > \frac{\varepsilon b_0(H)}{(1-2\varepsilon)n}\right]$$
$$\le \Pr\left[\exists k \in [\varepsilon mn, mn],\ Z_k > \frac{\varepsilon k}{(1-2\varepsilon)n}\right],$$

Where this last step used that the number of subcritical guesses $b_0(H)$ must always be at least εmn and at most mn.

Because Z_k is a centered binomial distribution, $\Pr[Z_k - Z_{k'} > \lambda p(k-k')] \le e^{-\frac{1}{3}\lambda^2 p(k-k')}$ for $k' < k$ by Lemma 2.2.3, and also $Z_k - Z_{k-1} \ge -p$ for all k by construction. If n is sufficiently large we have $\varepsilon mn \ge 2\varepsilon^{-1}$, so we can apply Lemma 2.2.9 to the above inequality with $c' = \frac{1}{3}$ and $c = 1$ to conclude

$$\Pr[Y_0' > L] \le \frac{8}{(1-2\varepsilon)\varepsilon^2} e^{-\frac{1}{768(1-2\varepsilon)}\varepsilon^4 m} \le 16\varepsilon^{-2} e^{-\frac{1}{768}\varepsilon^4 m},$$

with this last step using $\varepsilon \le \frac{1}{4}$. $\square$

2.2.5 Concentration of Critical Guesses

In the subcritical region we were able to bound the number of correct guesses by a binomial random variable. For the critical region, we compare the number of correct guesses with a hypergeometric random variable. We recall that a random variable $S \sim \mathrm{Hyp}(N, m, b)$ has a hypergeometric distribution (with parameters N, m, b) if for all integers

$1 \leq k \leq m$ we have

$$\Pr[S = k] = \binom{b}{k}\binom{N-b}{m-k}\binom{N}{m}^{-1}. \tag{2.9}$$

Equivalently one can define this by uniformly shuffling a deck of N cards with m of these cards being "good", and then letting S be the number of good cards one sees in the first b draws from the deck. From this viewpoint, if we let R_t denote the indicator variable which is 1 if the tth draw is a good card, we see that $S = \sum_{t=1}^{b} R_t$ and that the R_t can be defined by

$$\Pr[R_t = 1] = \frac{m - (R_1 + \cdots + R_{t-1})}{N - t + 1}. \tag{2.10}$$

We can use the following lemma to bound random variables by hypergeometric random variables.

Lemma 2.2.11. *Suppose $P_1, \ldots, P_k$ and $R_1, \ldots, R_k$ are $\{0,1\}$-random variables satisfying*

$$\begin{aligned}\Pr[P_t = 1] &\leq \frac{m - (P_1 + \cdots + P_{t-1})}{N - t + 1}\\ \Pr[R_t = 1] &= \frac{m - (R_1 + \cdots + R_{t-1})}{N - t + 1}.\end{aligned}$$

Then $R_1 + \cdots + R_k \succeq P_1 + \cdots + P_k$.

The proof of Lemma 2.2.11 follows from induction and applying Lemma 2.2.7 with $X = R_1 + \cdots + R_{t-1}$, $X' = R_t$, $Y = P_1 + \cdots + P_{t-1}$, and $Y' = P_t$. The last thing we need is to use Lemma 2.2.9 in this hypergeometric setting.

Lemma 2.2.12. *Let $N \geq m^2 + m$, define the indicator random variables $R_1, R_2, \ldots, R_N$*

as in (2.10)*, and let* $S_b := \sum_{t=1}^{b} R_t$ *for all* b*. Then for all* b *and* $0 < \lambda \leq 1$*,*

$$\Pr\left[S_b > \frac{(1+\lambda)bm}{N}\right] \leq 3e^{-\frac{\lambda^2 bm}{3N}}.$$

Further, for all $0 < \lambda \leq 1$ *and integers* b_0, b_1 *satisfying* $2\lambda^{-1} \leq b_0 \leq b_1 \leq N$*, we have*

$$\Pr\left[\exists b \in [b_0, b_1], S_b > \frac{(1+\lambda)bm}{N}\right] \leq \frac{24b_1}{\lambda b_0} e^{-\frac{\lambda^3 b_0 m}{768N}}.$$

Proof. Observe that $S_b \sim \mathrm{Hyp}(N, m, b)$. Thus if $q := b/N$, we have by (2.9) that

$$\begin{aligned}
\Pr[S_b = k] &= \binom{qN}{k}\binom{(1-q)N}{m-k}\binom{N}{m}^{-1} \leq \frac{(qN)^k}{k!}\frac{((1-q)N)^{m-k}}{(m-k)!}\frac{m!}{(N-m)^m} \\
&= \binom{m}{k} q^k (1-q)^{m-k}\left(1 + \frac{m}{N-m}\right)^m \leq \binom{m}{k} q^k (1-q)^{m-k} e^{m^2/(N-m)} \\
&\leq \binom{m}{k} q^k (1-q)^{m-k} \cdot 3,
\end{aligned}$$

where this last step used $N - m \geq m^2$.

We thus see that $\Pr[S_b > (1+\lambda)qm] \cdot 3^{-1}$ is at most the probability that a binomial distribution with m trials and probability q of success has at least $(1+\lambda)qm$ successes, which is at most $e^{-\lambda^2 qm/3}$ by Lemma 2.2.3. This gives the first result.

For the second result, define $p = m/N$ and let $Z_b := S_b - pb$. Note that $S_b - S_{b'} \sim \mathrm{Hyp}(N, m, b - b')$ for $b > b'$ (since this is just a sum of $b - b'$ of the R_t variables), so the first result implies

$$\Pr[Z_b - Z_{b'} > \lambda p(b - b')] = \Pr[S_b - S_{b'} > (1+\lambda)p(b-b')] \leq 3e^{-\frac{1}{3}\lambda^2 pb}.$$

We can thus apply Lemma 2.2.9 to the Z_b variables with $c' = 3$ and $c = -\frac{1}{3}$ to conclude the result. $\square$

Using this we can prove the following.

Lemma 2.2.13. *For $i \geq 1$ finite, $\varepsilon \leq \frac{1}{4}$, and n sufficiently large in terms of ε, m, we have*

$$\Pr\left[Y_i(H) > (1+4\varepsilon)\frac{b_i(H)}{n} + \varepsilon^2 m\right] \leq c'\varepsilon^{-2}e^{-c\varepsilon^4 m}$$

for some absolute constants $c, c' > 0$.

Proof. Fix i positive and finite, and let $X_t := Y_i(H_{t'}) - Y_i(H_{t'-1})$ where t' is the smallest positive integer for which $b_i(H_{t'}) = t$ (note that t' is itself a random variable). In other words, X_t is the indicator of the t-th critical guess of i. Define $X'_t = X_t$ if $Y(H) \leq \varepsilon mn$ and define $X'_t = 0$ otherwise.

Let R_t be random variables as in Lemma 2.2.11 with $N = (1-2\varepsilon)mn$, and define $S_b = \sum_{i=1}^{b} R_i$ for all $1 \leq b \leq (1-2\varepsilon)mn$. By applying Lemma 2.2.5 (and noting that i was guessed εmn times before its critical guesses started), we see

$$\Pr[X'_t = 1] \leq \frac{m - (X'_1 + \cdots + X'_{t-1})}{(1-2\varepsilon)mn - t + 1}.$$

Thus we can apply Lemma 2.2.11 with X'_t taking the role of P_t, and letting

$$L(b) = (1+4\varepsilon)\frac{b}{n} + \varepsilon^2 m$$

gives

$$
\begin{aligned}
\Pr[Y_i(H) > L(b_i(H))] &= \Pr\left[\sum_{t=1}^{b_i(H)} X_t > L(b_i(H))\right] \\
&\leq \Pr[Y(H) \leq \varepsilon mn] \cdot \Pr\left[\sum_{t=1}^{b_i(H)} X'_t > L(b_i(H))\right] + \Pr[Y(H) > \varepsilon mn] \\
&\leq 1 \cdot \Pr[S_{b_i(H)} > L(b_i(H))] + 2e^{-mn^{1/2}} \\
&\leq \Pr\left[\exists b \in [1, (1-2\varepsilon)mn],\ S_b > L(b)\right] + 2e^{-mn^{1/2}},
\end{aligned}
$$

where the second to last step used Lemma 2.2.2 and the last step used that the value of $b_i(H)$ must lie in 1 and $(1-2\varepsilon)mn$.

To bound $\Pr\left[\exists b \in [1, (1-2\varepsilon)mn],\ S_b > L(b)\right]$, we partition $[1, (1-2\varepsilon)mn]$ into intervals $[b_{j-1}, b_j]$ (which we define below) and show that $\Pr\left[\exists b \in [b_{j-1}, b_j],\ S_b > L_j(b)\right]$ is small for all j, where $L_j(b)$ is some quantity upper bounded by $L(j)$. Taking a union bound will then give the desired result.

Let $b_0 := \frac{1}{2}\varepsilon^2(1-2\varepsilon)mn \geq 2$ for n sufficiently large. By taking $\lambda = 1$ in Lemma 2.2.12, we have

$$
\Pr[\exists b \in [1, b_0],\ S_b > \varepsilon^2 m] = \Pr\left[S_{b_0} > \frac{2b_0 m}{(1-2\varepsilon)mn}\right] \leq 3e^{-\frac{b_0 m}{(1-2\varepsilon)mn}} = 3e^{-\frac{1}{4}\varepsilon^2 m}. \tag{2.11}
$$

Define $b_j = 2^j b_0$. Observe that for all $b \leq b_j$ we have $\varepsilon^2 \geq 2^{1-j}\frac{b}{(1-2\varepsilon)mn}$. Thus for

j such that $2^{1-j} \geq 4\varepsilon$ we find for n sufficiently large in terms of j,

$$\begin{aligned}
Pr[\exists b \in [b_{j-1}, b_j],\ S_b > \frac{b}{n} + \varepsilon^2 m] &\leq \Pr\left[\exists b \in [b_{j-1}, b_j],\ S_b > (1 - 2\varepsilon + 2^{1-j})\frac{bm}{(1-2\varepsilon)mn}\right] \\
&\leq \Pr\left[\exists b \in [b_{j-1}, b_j],\ S_b > (1 + 2^{-j})\frac{bm}{(1-2\varepsilon)mn}\right] \\
&\leq 48 \cdot 2^j e^{-\frac{2^j b_0}{2^{3j+9}(1-2\varepsilon)n}} = 48 \cdot 2^j e^{-2^{-2j-10}\varepsilon^2 m} \\
&\leq 48\varepsilon^{-1} e^{-2^{-10}\varepsilon^4 m}, \qquad (2.12)
\end{aligned}$$

where this third inequality used Lemma 2.2.12.

Let $J = \lfloor \log_2(\varepsilon^{-1}) \rfloor - 1$, noting that we can apply the above bound up to

$$b_J = 2^{J-1}\varepsilon^2(1-2\varepsilon)mn \geq \frac{1}{8}\varepsilon(1-2\varepsilon)mn.$$

Observe that for $\varepsilon \leq \frac{1}{8}$,

$$\frac{(1+4\varepsilon)b}{n} \geq \frac{(1+\varepsilon)b}{(1-2\varepsilon)n}.$$

Thus by Lemma 2.2.12 applied with $\lambda = \varepsilon$, we see that

$$\Pr\left[\exists b \in [b_J, (1-2\varepsilon)mn],\ S_b > \frac{(1+4\varepsilon)b}{n}\right] \leq \frac{24(1-2\varepsilon)mn}{\varepsilon b_J} e^{-\frac{\varepsilon^3 b_J}{768(1-2\varepsilon)n}} \leq 96\varepsilon^{-2} e^{\frac{-1+\varepsilon}{3072(1-2\varepsilon)}\varepsilon^4 m}.$$

Taking the union bound over this, (2.11), and (2.12) for the at most $-\log_2(\varepsilon) \leq \varepsilon^{-1}$ values of $j \leq J$ gives the result. □

2.2.6 Completing the Proof

We need the following simple consequence of Lemma 2.2.8.

Lemma 2.2.14. *If n is sufficiently large and A is an event with $\Pr[A] = p > 0$, then*

$$\mathbb{E}[Y(H)|A] < 200m + 20p^{-1}.$$

Proof. The statement of Lemma 2.2.8 implies the following: For any $0 < \lambda \leq 1/6$ and $n \geq 200\lambda^{-1}$,

$$\Pr[Y(H) > \lambda mn] \leq 2e^{-\lambda mn/12}.$$

By taking $x = \lambda n$, this is equivalent to saying that for $200 \leq x \leq n/6$ we have

$$\Pr[Y(H) > xm] \leq 2e^{-xm/12}.$$

In particular, even after conditioning on the event A,

$$\Pr[Y(H) > xm|A] \leq 2p^{-1}e^{-xm/12}.$$

With this we have

$$\begin{aligned}
\mathbb{E}[Y(H)|A] &= m\int_0^n \Pr[Y(H) > xm|A]dx \\
&\leq 200m + m\int_{200}^{n/6} \Pr[Y(H) > xm|A]dx + mn \cdot \Pr[Y(H) > mn/6] \\
&\leq 200m + m\int_0^{\infty} 2p^{-1}e^{-xm/12}dx + 2mne^{-mn^{1/2}} \\
&= 200m + 12p^{-1} + 2mne^{-mn^{1/2}},
\end{aligned}$$

where the second inequality used the observation made above and Lemma 2.2.2. This gives the result by taking n to be sufficiently large in terms of m. □

Finally we have all the tools to prove the main theorem.

Proof of Theorem 2.1.3. We will pick $\varepsilon = O((\log m/m)^{1/4})$, and show that for an appropriate such ε and n sufficiently large in terms of m and ε, $\mathbb{E}[Y(H)] \leq (1+\varepsilon)m$. To this end, we define the following three "atypical" events: E_0, E_1 and E_∞.

- The event E_0 is the event that $Y_0(H) > (1 + 4\varepsilon)b_0(H)/n$, in other words that significantly more than the average number of subcritical guesses are correct.

- The event E_1 is the event that $Y_i(H) > (1 + 4\varepsilon)b_i(H)/n + \varepsilon^2 m$ for some $i \geq 1$, in other words that for some critical card i, significantly more than the average number of critical guesses of card i are correct.

- The event E_∞ is the event that there is at least one supercritical card. In this case, this single card is guessed at least $(1 - \varepsilon)mn$ times.

Our goal will be to calculate the conditional expectation of $Y(H)$ depending on whether or not the exceptional events above occur. It will be convenient to group E_0 and E_1 together and define their union $A = E_0 \vee E_1$. Then,

$$\mathbb{E}[Y(H)] = \Pr[A]\mathbb{E}[Y(H)|A]+\Pr[\overline{A}\wedge\overline{E}_\infty]\cdot\mathbb{E}[Y(H)|\overline{A}\wedge\overline{E}_\infty]+\Pr[\overline{A}\wedge E_\infty]\mathbb{E}[Y(H)|\overline{A}\wedge E_\infty]. \quad (2.13)$$

We first observe that if none of the events E_0, E_1 and E_∞ occur, then the conditional expectation of $Y(H)$ is small. Indeed, we have

$$\mathbb{E}[Y(H)|\overline{A} \wedge \overline{E}_\infty] = \mathbb{E}[Y(H)|\overline{E}_0 \wedge \overline{E}_1 \wedge \overline{E}_\infty] \leq (1 + 5\varepsilon)m, \quad (2.14)$$

since all guesses must be subcritical or critical, and there are at most ε^{-1} distinct critical card types i.

Define $p_j = \Pr[E_j]$ for $j \in \{0, 1\}$. We have by Lemma 2.2.10 and Lemma 2.2.13 that for some absolute constants $c, c' > 0$,

$$p_0 \leq c'\varepsilon^{-2}e^{-c\varepsilon^4 m}$$

and

$$p_1 \le c' \varepsilon^{-3} e^{-c\varepsilon^4 m},$$

where there is an extra multiplicative factor of ε^{-1} in the second inequality because there may be up to ε^{-1} critical cards. In particular, we have

$$\Pr[A] = \Pr[E_0 \vee E_1] \le p := 2c'\varepsilon^{-3} e^{-c\varepsilon^4 m}. \tag{2.15}$$

By Lemma 2.2.14 we find

$$\mathbb{E}[Y(H)|A] \le 200m + 20 \Pr[A]^{-1},$$

and so

$$\Pr[A]\mathbb{E}[Y(H)|A] \le 200m \Pr[A] + 20 \le 200mp + 20,$$

for p defined above. By picking an appropriate $\varepsilon = O((\log m/m)^{1/4})$, we find

$$\Pr[A]\mathbb{E}[Y(H)|A] \le m^{-\Omega(1)} + 20 < \varepsilon m \tag{2.16}$$

for m sufficiently large.

Finally, to control the third term of (2.13), note that if there is a supercritical card, at most m guesses are correct for that card (since there are a total of m copies of that card in the deck), and at most εmn guesses are made of any other card, so all other guesses are subcritical. In particular, including guesses of the unique supercritical card, there at most $b_0(H) \le 2\varepsilon mn$ subcritical guesses. Thus, by the definition of E_0, we get

$$\mathbb{E}[Y(H)|\overline{A} \wedge E_\infty] \le m + (1 + 4\varepsilon)(2\varepsilon m) \le (1 + 3\varepsilon)m.$$

In total, using (2.13), (2.14), (2.16), and the inequality above, we find that for m, n sufficiently large,

$$\mathbb{E}[Y(H)] \leq \varepsilon m + \Pr[\overline{A}] \cdot (1+5\varepsilon)m \leq (1+6\varepsilon)m = m + O(m^{3/4}\log^{1/4} m),$$

completing the proof. $\square$

2.3 Concluding Remarks

2.3.1 Sharper Bounds

Our main result was a tight asymptotic upper bound on $\mathcal{P}^+_{m,n}$, which is the most number of cards one can guess correctly in expectation in the partial feedback model. Specifically, we proved $\mathcal{P}^+_{m,n} = m + O(m^{3/4}\log^{1/4} m)$ provided n is sufficiently large. We have a trivial lower bound of $\mathcal{P}^+_{m,n} \geq m$ coming from the strategy of guessing the same card type every round, so we know $\mathcal{P}^+_{m,n} \sim m$. At this point it is natural to ask to improve upon the bounds for the error term of $\mathcal{P}^+_{m,n}$, and especially to improve upon the trivial lower bound of m. For this latter task, we need to exhibit effective strategies for the partial feedback model. It will also be of independent interest to construct practical strategies, i.e. strategies for which a human player could reasonably implement.

Perhaps the simplest non-trivial strategy one could consider is the strategy of guessing 1 until m correct guesses are made, then 2 until m correct guesses are made, and so on. While this does better than the trivial strategy of guessing the same card type every round, one can show that this strategy only achieves about $m+1-\frac{1}{m+1}$ correct guesses in expectation, see [DGS20].

Another natural strategy to consider is the *shifting strategy* $\mathcal{G}$ defined by guessing 1 until you get a correct guess, then 2 until you get a correct guess, and so on; and if you guess n correctly, play arbitrarily for the remaining trials. The score in this game turns out to be somewhat complex, and it is closely related to certain increasing subsequences of random multiset permutations.

For $\pi \in \mathfrak{S}_{m,n}$, let $L(\pi)$ denote the largest integer p such that there exist $i_1 < \cdots < i_p$ with $\pi_{i_j} = j$ for all $1 \le j \le p$; i.e. $L(\pi)$ is the longest subsequence of the form $12\cdots p$. It is not difficult to see that if one plays the partial feedback model with a deck shuffled according to $\pi \in \mathfrak{S}_{m,n}$, then $L(\pi)$ is at least the score obtained when using the shifting strategy (and it is equal to the score provided the player does not get n correct guesses, which happens with overwhelming probability). Together with Clifton, Deb, Huang, and Yoo [CDH+21], we obtained very precise estimates for $\mathbb{E}[L(\pi)]$, and hence for the expected score under the shifting strategy.

Theorem 2.3.1 ([CDH+21])**.** *Let $\mathcal{L}_{m,n} = \mathbb{E}[L(\boldsymbol{\pi})]$ where $\boldsymbol{\pi} \sim \mathfrak{S}_{m,n}$. For any integer $m \ge 1$, let $\alpha_1, \ldots, \alpha_m$ be the zeroes of $E_m(x) := \sum_{k=0}^{m} \frac{x^k}{k!}$. We have*

$$\mathcal{L}_m := \lim_{n\to\infty} \mathcal{L}_{m,n} = -1 - \sum \alpha_i^{-1} e^{-\alpha_i}. \tag{2.17}$$

Moreover, there exists an absolute constant $\beta > 0$ such that

$$\left| \mathcal{L}_m - \left(m + 1 - \frac{1}{m+2} \right) \right| \le O(e^{-\beta m}).$$

For example, this says that the limiting score of the shifting strategy when $m = 1$ is $\mathcal{L}_1 = e - 1$, which is easy to prove by hand. However, even at $m = 2$ we get the non-trivial conclusion that $\mathcal{L}_2 = e(\cos(1) + \sin(1)) - 1$.

Theorem 2.3.1 shows that the shifting strategy also only gives about $m + 1$ correct guesses in expectation, so we see that the simplest possible strategies do not do much better than the trivial one. The following result due to Diaconis, Graham, and ourselves [DGS20] gives the best known strategy for the partial feedback model.

Theorem 2.3.2 ([DGS20])**.** *If n is sufficiently large in terms of m, then*

$$\mathcal{P}^{+}_{m,n} = m + \Omega(m^{1/2}).$$

Sketch of Proof. Consider the following strategy. Guess 1 a total of $mn/2$ times. If you guessed at least $\frac{1}{2}m + \sqrt{m}$ cards correctly, guess 2 for the rest of the game, otherwise keep guessing 1. In the latter scenario we always get exactly m correct guesses. One can show that the first scenario happens with some constant probability, and given this, the expected number of 2's left in the second half of the deck is at least $m/2$. In total this gives a lower bound of $m + \Omega(m^{1/2})$. □

We suspect that this lower bound is close to the truth.

Conjecture 2.3.3. *For all $\varepsilon > 0$ and n sufficiently large,*

$$\mathcal{P}^{+}_{m,n} = m + O(m^{1/2+\varepsilon}).$$

The current proof overshoots this bound at two points. The first is in Lemma 2.2.9 where we try and bound the probability that an "adversarial" binomial distribution deviates significantly from its mean. Our proof of this lemma essentially only used a union bound, and it's plausible that more sophisticated techniques could decrease this error term.

The second point is in the bounds of Lemmas 2.2.10 and 2.2.13 where we bound the probability that the subcritical or critical guesses are much larger than average. We note that by adding in an error term of εm to the lower bound of Y_0 in Lemma 2.2.10, one can decrease the probability from roughly $e^{-\varepsilon^4 m}$ to $e^{-\varepsilon^3 m}$ (which could go down to $e^{-\varepsilon^2 m}$ if Lemma 2.2.9 is improved), so the central issue is the critical case, and it seems like new ideas are needed here.

2.3.2 Minimizing Scores

Another problem of interest is bounding $\mathcal{P}^-_{m,n}$, the fewest number of cards one can guess correctly in expectation in the partial feedback model. In [DGS20] we proved $\mathcal{P}^-_{m,n} \leq m - \Omega(m^{1/2})$ using an analog of the strategy for $\mathcal{P}^+_{m,n}$ sketched in Theorem 2.3.2. We also proved an asymptotic lower bound for $\mathcal{P}^-_{m,n}$ of $1 - e^{-m}$ by showing that one always has probability at least roughly $1 - e^{-m}$ of guessing at least one card correctly. Thus $\mathcal{P}^-_{m,n} = \Omega(1)$, which is again in sharp contrast to the complete feedback model where one can get arbitrarily few correct guesses in expectation. There is still a large gap between these bounds, and as in Theorem 2.1.3 we suspect that the partial feedback model does not allow one to guess significantly fewer guesses than in the no feedback model.

Conjecture 2.3.4. *If n is sufficiently large in terms of m, then*

$$\mathcal{P}^-_{m,n} \sim m.$$

The central difficulty in this setting is that there does not exist an analog of Lemma 2.2.1 which lower bounds $\Pr[\boldsymbol{\pi}_t = i]$ given that we have not guessed i many times and that we have guessed few cards correctly. For example, say we incorrectly guessed 1 a total of $(m-1)n$, so the remaining cards are m copies of 1. Then the probability that the next card is 2 is 0 despite the fact that we have not guessed 2 at all nor guessed any cards correctly.

2.3.3 Other Models

One can consider variants of these models where $\boldsymbol{\pi}$ is chosen according to some non-uniform distribution. For $m = 1$, the case when π is obtained from a single riffle shuffle is studied by Ciucu [Ciu98] (under no feedback), and Liu [Liu21] (under complete feedback). Analysis under repeated "top to random shuffles" is done by Pehlivan [Peh10].

In [Spi21], we considered a variant of the complete feedback model where an adversary is allowed to shuffle the deck according to an arbitrary distribution. Our main result can be stated as follows.

Theorem 2.3.5 ([Spi21])**.** *There exists a distribution for $\boldsymbol{\pi} \in \mathfrak{S}_{m,n}$ such that for any strategy $\mathcal{G}$ in the complete feedback model,*

$$\Gamma\left(1+\frac{1}{m}\right)(m!)^{1/m}\cdot n^{-1/m}+o(n^{-1/m}) \leq \mathbb{E}[C(\mathcal{G},\boldsymbol{\pi})] \leq \log n + o(\log n).$$

Moreover, both of these bounds are asymptotically best possible.

By Theorem 2.1.2 we see that these bounds differ from those of $\mathcal{C}^{\pm}_{m,n}$ (i.e. the maximum expected score when the deck is shuffled according to a uniform distribution) by multiplicative factors of $(m!)^{1/m}$ and H_m, respectively. Thus an adversary can significantly alter the extremal scores that the player can obtain in expectation.

This chapter contains material from: P. Diaconis, R. Graham, X. He, and S. Spiro, *"Card Guessing with Partial Feedback"*, Combinatorics, Probability, and Computing (2021) 1–20. The dissertation author was one of the primary investigators and authors of this paper.

Chapter 3

Turán's Problem in Random Hypergraphs

3.1 Introduction

Recall that $H^r_{n,p}$ denotes the random r-graph obtained by keeping each possible hyperedge on K_n independently and with probability p, and that $\mathrm{ex}(H^r_{n,p}, \mathcal{F})$ denotes the maximum size of an $\mathcal{F}$-free subgraph of $H^r_{n,p}$. The problem of understanding the behavior of $\mathrm{ex}(H^r_{n,p}, \mathcal{F})$ when $\mathcal{F}$ contains no r-partite r-graphs was essentially solved due to independent work of Conlon and Gowers [CG16] and Schacht [Sch16], but only a few sporadic results are known when $\mathcal{F}$ contains an r-partite r-graph. The first result in this direction was obtained by Füredi [Für94] who essentially solved the problem for $\mathcal{F} = \{C_4\}$. Kohayakawa, Kreuter, and Steger [KKS98] considered the more general problem of avoiding $C_{2\ell}$ and proved essentially tight bounds whenever p is not too large. In breakthrough work, Morris and Saxton [MS16] essentially solved the problem for $\mathcal{F} = \{C_{2\ell}\}$ and $\mathcal{F} = \{K_{s,t}\}$ for all values of p provided certain well known conjectures in extremal graph theory are true.

In this chapter we consider analogs of these results for hypergraphs. For example, we proved the following result with Verstraëte [SV21], where here $K_{s_1,\ldots,s_r}$ denotes the complete r-partite r-graph with parts of sizes $s_1, \ldots, s_r$.

Theorem 3.1.1 ([SV21])**.** *For $r \geq 2$, let $2 \leq s_1 \leq \cdots \leq s_r$ be integers, $a_i = \prod_{j=1}^{i-1} s_j$ for $i = r, r+1$, and*

$$\beta_1 = \frac{\sum_{i=1}^{r} s_i - r}{a_{r+1} - 1} \qquad \textit{and} \qquad \beta_2 = \frac{a_r(\sum_{i=1}^{r-1} s_i - r) + 1}{(a_r - 1)(a_{r+1} - 1)}.$$

If $\mathrm{ex}(n, K_{s_1,\ldots,s_r}) = \Omega(n^{r-1/a_r})$, then a.a.s.

$$\mathrm{ex}(H^r_{n,p}, K_{s_1,\ldots,s_r}) = \begin{cases} \Theta(pn^r) & n^{-r}\log n \leq p \leq n^{-\beta_1}, \\ n^{r-\beta_1+o(1)} & n^{-\beta_1} \leq p \leq n^{-\beta_2}(\log n)^{2a_r/(a_r-1)}, \\ \Theta(p^{1-1/a_r} n^{r-1/a_r}) & n^{-\beta_2}(\log n)^{2a_r/(a_r-1)} \leq p \leq 1. \end{cases}$$

For example, in Figure 3.1 we have plotted $f(n,p) = \mathbb{E}[\mathrm{ex}(G_{n,p}, K_{2,2})]$. Note that this plot goes through three distinct phases: when p is very small, one can find a $K_{2,2}$-free subgraph using almost every edge of $G_{n,p}$. The function is essentially constant for medium values of p, and for larger p the function grows in some non-trivial way with p. This kind of behavior is typical for these sorts of problems.

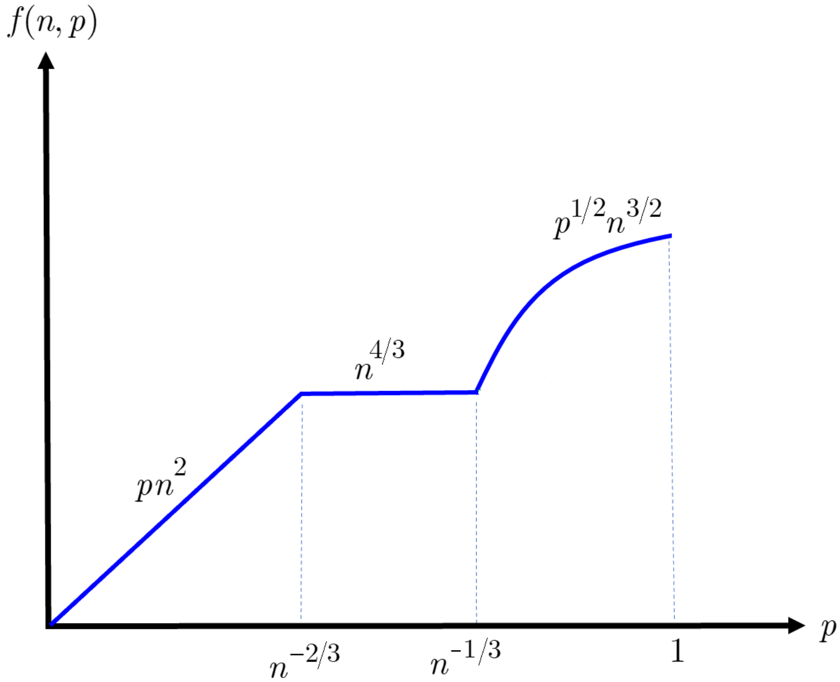

Figure 3.1: Plot of $f(n,p) = \mathbb{E}[\mathrm{ex}(G_{n,p}, K_{2,2})]$.

We next turn to hypergraph cycles. There are several different notions of hypergraph cycles, and we start by considering loose cycles. The r-uniform loose ℓ-cycle C_ℓ^r is the r-graph which is obtained from the graph cycle C_ℓ by inserting $r-2$ distinct vertices into each edge of $C_{2\ell}$. More precisely, the hypergraph consits of hyperedges $e_1, \ldots, e_\ell$ such that there exist distinct vertices $v_1, \ldots, v_\ell$ with $e_{i-1} \cap e_i = \{v_i\}$ (where the indices are written cyclically) and $e_i \cap e_j = \emptyset$ otherwise. See Figure 3.2 for a picture of C_4^3.

With Nie and Verstraëte [NSV20], we obtained the following bound for the loose triangle C_3^3.

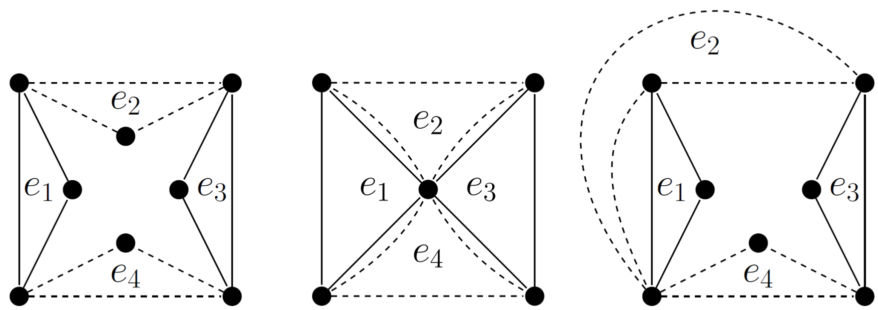

Figure 3.2: The leftmost hypergraph depicts C_4^3. All of the hypergraphs are examples of 3-uniform Berge 4-cycles.

Theorem 3.1.2 ([NSV20])**.** *If $p \geq n^{-\frac{3}{2}}(\log n)^3$, then a.a.s.*

$$p^{\frac{1}{3}} n^{2-o(1)} \leq \mathrm{ex}(H_{n,p}^3, C_3^3) \leq p^{\frac{1}{3}} n^{2+o(1)}.$$

We note that it is easy to show $\mathrm{ex}(H_{n,p}^3, C_3^3) = \Theta(pn^3)$ provided $n^{-3} \ll p \ll n^{-3/2}$ by a standard deletion argument. Bounds for $r > 3$ were also obtained in [NSV20], but these bounds are not tight and somewhat cumbersome to state. The best known bounds involving loose even cycles $C_{2\ell}^r$ are due to Mubayi and Yepremyan [MY20], and once again there are significant gaps between the best known upper and lower bounds.

We next turn to Berge cycles, which is the main focus of this chapter. For $\ell \geq 2$, an r-graph F is said to be a *Berge ℓ-cycle* if there exist distinct vertices $v_1, \ldots, v_\ell$ and distinct hyperedges $e_1, \ldots, e_\ell$ with $v_i, v_{i+1} \in e_i$ for all $1 \leq i \leq \ell$. See Figure 3.2 for some examples of 3-uniform Berge 4-cycles.

Observe that the loose cycle C_ℓ^r is a Berge ℓ-cycle, and that a hypergraph H is *linear* (i.e. no two hyperedges of H intersect in at least two vertices) if and only if it contains no Berge 2-cycle. We denote by $\mathcal{C}_\ell^r$ the family of all r-uniform Berge ℓ-cycles. If

H is an r-graph containing a Berge cycle, then the *girth* of H is the smallest $\ell \geq 2$ such that H contains a Berge ℓ-cycle. Let $\mathcal{C}^r_{[\ell]} = \mathcal{C}^r_2 \cup \mathcal{C}^r_3 \cup \cdots \cup \mathcal{C}^r_\ell$ denote the family of all r-uniform Berge cycles of length at most ℓ, and when $r = 2$ we simply write $\mathcal{C}_{[\ell]}$. With this an r-graph has girth larger than ℓ if and only if it is $\mathcal{C}^r_{[\ell]}$-free.

Note that $\mathrm{ex}(H^r_{n,p}, \mathcal{C}^r_{[2]})$ denotes the largest subgraph of $H^r_{n,p}$ which has girth 3, i.e. the largest subgraph which is linear. It is not hard to show by a simple first moment calculation that if $p \geq n^{-r} \log n$, then a.a.s

$$\mathrm{ex}(H^r_{n,p}, \mathcal{C}^r_{[2]}) = \Theta(\min\{pn^r, n^2\}).$$

By using a non-trivial argument, we can determines the a.a.s. behavior of the number of edges in an extremal subgraph of $H^r_{n,p}$ of girth four. In this theorem we omit the case $p < n^{-r+\frac{3}{2}}$, as it is straightforward to show that a.a.s $\mathrm{ex}(H^r_{n,p}, \mathcal{C}^r_{[3]}) = \Theta(pn^r)$ in this range when $p \geq n^{-r} \log n$.

Theorem 3.1.3. *Let $r \geq 3$. If $p \geq n^{-r+\frac{3}{2}} (\log n)^{2r-3}$, then a.a.s.*

$$p^{\frac{1}{2r-3}} n^{2-o(1)} \leq \mathrm{ex}(H^r_{n,p}, \mathcal{C}^r_{[3]}) \leq p^{\frac{1}{2r-3}} n^{2+o(1)}.$$

We are not able to obtain tight bounds for larger girths. Part of the difficulty here lies with the fact that the Turán numbers $\mathrm{ex}(n, \mathcal{C}^r_{[\ell]})$ are unknown in general. A reasonable guess is the following, which is essentially a strengthening of a conjecture of Erdős and Simonovits [ES83] for graphs.

Conjecture 3.1.4. *For all $\ell \geq 3$ and $r \geq 2$ and $k = \lfloor \ell/2 \rfloor$,*

$$\mathrm{ex}(n, \mathcal{C}^r_{[\ell]}) = n^{1+1/k-o(1)}.$$

Conjecture 3.1.4 is known to hold for $r = 2$ without the $o(1)$ term for $\ell \in \{3,4,5,6,7,10,11\}$. It is also known to hold for $\ell = 3,4$ and $r \geq 3$ – see [EFR86, LV03, RS78, TV15] – but is open and evidently difficult for $\ell \geq 5$ and $r \geq 3$. Györi and Lemons [GL12] proved $\mathrm{ex}(n, \mathcal{C}_\ell^r) = O(n^{1+1/k})$ with $k = \lfloor \ell/2 \rfloor$, so the conjecture concerns constructions of dense r-graphs of girth more than ℓ. We emphasize that the $o(1)$ term in Conjecture 3.1.4 is necessary for $\ell = 3$, due to the Ruzsa-Szemerédi Theorem [EFR86, RS78], and for $\ell = 5$, due to work of Conlon, Fox, Sudakov and Zhao [CFSZ20].

With this conjecture in mind, we can prove the following.

Theorem 3.1.5. *Let $\ell \geq 4$ and $r \geq 2$, and let $k = \lfloor \ell/2 \rfloor$ and $\lambda = \lceil (r-2)/(\ell-2) \rceil$. Then a.a.s.*

$$\mathrm{ex}(H_{n,p}^r, \mathcal{C}_{[\ell]}^r) \leq \begin{cases} n^{1+\frac{1}{\ell-1}+o(1)} & n^{-r+1+\frac{1}{\ell-1}} \leq p < n^{\frac{-(r-1+\lambda)(k-1)}{2k-1}} (\log n)^{(r-1+\lambda)k}, \\ p^{\frac{1}{(r-1+\lambda)k}} n^{1+\frac{1}{k}+o(1)} & n^{\frac{-(r-1+\lambda)(\ell-1-k)}{\ell-1}} (\log n)^{(r-1+\lambda)k} \leq p \leq 1. \end{cases}$$

If Conjecture 3.1.4 is true, then

$$\mathrm{ex}(H_{n,p}^r, \mathcal{C}_{[\ell]}^r) \geq \begin{cases} n^{1+\frac{1}{\ell-1}+o(1)} & n^{-r+1+\frac{1}{\ell-1}} \leq p < n^{\frac{-(r-1)(\ell-1-k)}{\ell-1}}, \\ p^{\frac{1}{(r-1)k}} n^{1+\frac{1}{k}-o(1)} & n^{\frac{-(r-1)(\ell-1-k)}{\ell-1}} \leq p \leq 1. \end{cases}$$

We emphasize that there is a significant gap in the bounds of Theorem 3.1.5 due to the presence of λ in the exponent of p in the upper bound and its absence in the lower bound, and we close this gap when $\ell = 3$ through Theorem 3.1.3 by an improvement to the lower bound. Thus the simplest case where a gap remains is when $\ell = 4$ and $r = 3$.

Letting $g(n,p) = \mathbb{E}[\mathrm{ex}(H^3_{n,p}, \mathcal{C}^3_{[4]})]$, we plot the bounds of Theorem 3.1.5 in Figure 3.3, where the upper bound is in blue and the lower bound is in green.

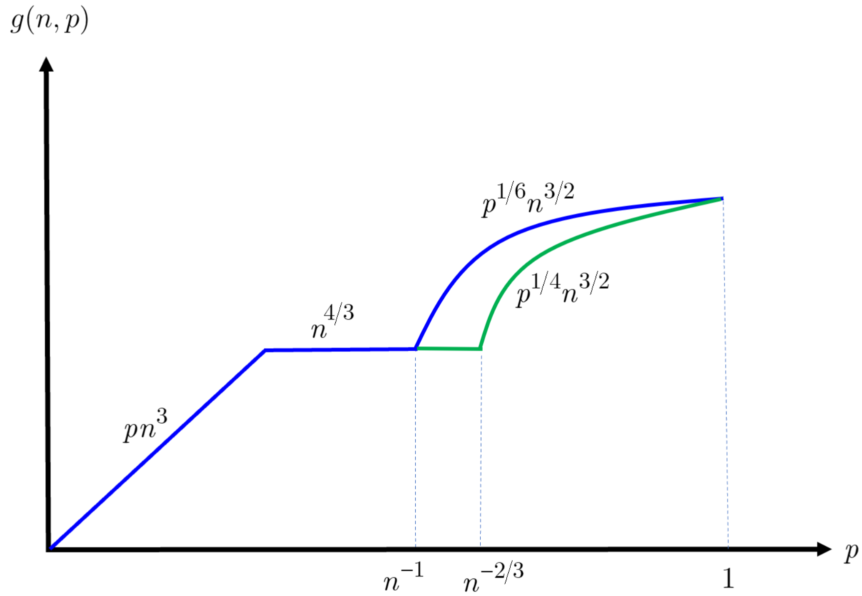

Figure 3.3: Plot of $g(n,p) = \mathbb{E}[\mathrm{ex}(H^3_{n,p}, \mathcal{C}^3_{[4]})]$, the expected maximum size of a subgraph of $H^3_{n,p}$ of girth five.

3.1.1 Counting $\mathcal{F}$-free Hypergraphs

There are two somewhat standard techniques for proving bounds on $\mathrm{ex}(H^r_{n.p}, \mathcal{F})$. For lower bounds, one (roughly speaking) considers a random map $\phi : V(H^r_{n,p}) \to V(J)$ where J is some $\mathcal{F}$-free r-graph with many edges, and then one takes $H \subseteq H^r_{n,p}$ to consist of all the hyperedges which get mapped to hyperedges of T via ϕ. This exact approach as stated does not work in general, but some standard adjustments can be used to get an effective bound in many cases. We will see examples of this in Section 3.4.

The problem of upper bounding $\mathrm{ex}(H^r_{n,p}, \mathcal{F})$ is typically harder and is intimately related to the problem of counting $\mathcal{F}$-free r-graphs with a given number of hyperedges, which is a problem of independent interest. To this end, define $\mathrm{N}^r(n, \mathcal{F})$ to be the number of $\mathcal{F}$-free r-graphs on $[n] := \{1, \ldots, n\}$, and define $\mathrm{N}^r_m(n, \mathcal{F})$ to be the number of $\mathcal{F}$-free r-graphs on $[n]$ with exactly m hyperedges. A simple first moment argument can be used to show that upper bounds on $\mathrm{N}^r_m(n, \mathcal{F})$ directly translate to upper bounds for $\mathbb{E}[\mathrm{ex}(H^r_{n,p}, \mathcal{F})]$, and this is essentially what we will use to prove all of the upper bounds of our main results.

Recalling that $\mathrm{ex}(n, \mathcal{F})$ denotes the maximum number of hyperedges in an $\mathcal{F}$-free r-graph on $[n]$; it is not difficult to see that for $1 \le m \le \mathrm{ex}(n, \mathcal{F})$,

$$\left(\frac{\mathrm{ex}(n, \mathcal{F})}{m}\right)^m \le \binom{\mathrm{ex}(n, \mathcal{F})}{m} \le \mathrm{N}^r_m(n, \mathcal{F}) \le \binom{\binom{n}{r}}{m} \le \left(\frac{en^r}{m}\right)^m,$$

and summing over m one obtains $2^{\Omega(\mathrm{ex}(n,\mathcal{F}))} = \mathrm{N}^r(n, \mathcal{F}) = 2^{O(\mathrm{ex}(n,\mathcal{F}) \log n)}$. The state of the art for bounding $\mathrm{N}^r(n, \mathcal{F})$ is the work of Ferber, McKinley, and Samotij [FMS20] which shows that if F is an r-uniform hypergraph with $\mathrm{ex}(n, F) = O(n^\alpha)$ and α not too small, then

$$\mathrm{N}^r(n, F) = 2^{O(n^\alpha)},$$

and this result encompasses many of the earlier results in the area [BNS19, BS11, CT17, MS16].

There are relatively few families for which effective bounds for $\mathrm{N}^r_m(n, \mathcal{F})$ are known. One family where results are known is $\mathcal{C}_{[\ell]} = \{C_3, C_4, \ldots, C_\ell\}$, the family of all graph cycles of length at most ℓ. Morris and Saxton implicitly proved the following in this setting:

Theorem 3.1.6 ([MS16]). *For $\ell \geq 3$ and $k = \lfloor \ell/2 \rfloor$, there exists a constant $c = c(\ell) > 0$ such that if n is sufficiently large and $m \geq n^{1+1/(2k-1)}(\log n)^2$, then*

$$\mathrm{N}^2_m(n, \mathcal{C}_{[\ell]}) \leq e^{cm}(\log n)^{(k-1)m}\left(\frac{n^{1+1/k}}{m}\right)^{km}.$$

We note that if $\mathrm{ex}(n, \mathcal{C}_{[2\ell]}) = \Theta(n^{1+1/\ell})$, then this result would be best possible up to the exponent of $(\log n)^m$

Essentially all of the previously known upper bounds for $\mathrm{N}^2_m(n, \mathcal{F})$ required proving a technical result known as a "balanced supersaturation" lemma together with a routine (but tedious) calculation using the powerful machinery of hypergraph containers. Indeed, this exact approach is what we used to prove Theorems 3.1.1 and 3.1.2, and is what was used to prove Theorem 3.1.6. The main goal of this chapter is to present a fairly simple argument which, assuming Theorem 3.1.6, can be used to count hypergraphs of a given girth.

To this end, we write $\mathrm{N}^r_m(n, \ell) := \mathrm{N}^r_m(n, \mathcal{C}^r_{[\ell]})$ for the number of n-vertex r-graphs with m edges and girth larger than ℓ. By refining an argument of Balogh and Li [BL20], we prove effective and almost tight bounds on $\mathrm{N}^r_m(n, \ell)$ relative to $\mathrm{N}^2_m(n, \ell)$.

Theorem 3.1.7. *Let $\ell, r \geq 3$ and $\lambda = \lceil (r-2)/(\ell-2) \rceil$. Then for all $m, n \geq 1$,*

$$\mathrm{N}^r_m(n, \ell) \leq \mathrm{N}^2_m(n, \ell)^{r-1+\lambda}. \tag{3.1}$$

This inequality is essentially tight when $\ell - 2$ divides $r - 2$ due to standard probabilistic arguments (see for instance Janson, Łuczak and Rucinski [JŁR00]): it is possible to show that when $m \leq n^{1+1/(\ell-1)}$, the uniform model of random n-vertex r-graphs with

m edges has girth larger than ℓ with probability at least a^{-m} for some constant $a > 1$ depending only on ℓ and r. In particular, there exists some constants $b, c > 1$ such that for $m \leq n^{1+1/(\ell-1)}$ we have

$$\mathrm{N}^r_m(n,\ell) \geq a^{-m}\binom{\binom{n}{r}}{m} \geq b^{-m}(n^r/m)^m \geq b^{-m}(n^2/m)^{(r-1+\frac{r-2}{\ell-2})m} \geq c^{-m}\cdot \mathrm{N}^2_m(n,\ell)^{r-1+\frac{r-2}{\ell-2}}, \tag{3.2}$$

where the third inequality used $m \leq n^{1+1/(\ell-1)}$ and the last inequality used the trivial bound $\mathrm{N}^2_m(n,\ell) \leq (en^2/m)^m$. This shows that the bound of Theorem 3.1.7 is best possible when $\ell - 2$ divides $r - 2$ up to a multiplicative error of c^{-m} for some constant $c > 1$. We believe that (3.2) should define the optimal exponent, and propose the following conjecture:

Conjecture 3.1.8. *For all $r \geq 2$, $\ell \geq 3$ and $m, n \geq 1$,*

$$\mathrm{N}^r_m(n,\ell) \leq \mathrm{N}^2_m(n,\ell)^{r-1+\frac{r-2}{\ell-2}}.$$

Theorem 3.1.7 shows that this conjecture is true when $\ell - 2$ divides $r - 2$, so the first open case of Conjecture 3.1.8 is when $\ell = 4$ and $r = 3$. We note that a proof of Conjecture 3.1.8 would give improved bounds to Theorem 3.1.5, but even with this there would still be a gap between the upper and lower bounds.

In the case that Berge ℓ-cycles are forbidden instead of all Berge cycles of length at most ℓ, we can prove an analog of Theorem 3.1.7 with weaker quantitative bounds. To this end, let $\mathrm{N}^r_{[m]}(n,\mathcal{F})$ denote the number of n-vertex $\mathcal{F}$-free r-graphs on at most m hyperedges.

Theorem 3.1.9. *For each $\ell, r \geq 3$, there exists $c = c(\ell, r)$ such that*

$$\mathrm{N}^r_m(n, \mathcal{C}^r_\ell) \leq 2^{cm} \cdot \mathrm{N}^2_{[m]}(n, C_\ell)^{r!/2}.$$

We suspect that this result continues to hold with $\mathrm{N}^2_{[m]}(n, C_\ell)$ replaced by $\mathrm{N}^2_m(n, C_\ell)$. In any case, one can use Theorem 3.1.9 to obtain non-trivial upper bounds on $\mathrm{ex}(H^r_{n,p}, \mathcal{C}^r_\ell)$, but these bounds are very far from the lower bounds when r is large.

Organization. Theorem 3.1.7 and the upper bounds of Theorems 3.1.5 and 3.1.3 are proven in Section 3.2. Theorem 3.1.9 is proven in Section 3.3. The lower bounds for Theorems 3.1.5 and 3.1.3 are proven in Section 3.4.

Notation. A set of size k will be called a *k-set*. Given a hypergraph H on $[n]$, we define the *k-shadow* $\partial^k H$ to be the k-graph on $[n]$ consisting of all k-sets e which lie in a hyperedge of $E(H)$. As much as possible, when working with a k-graph G and an r-graph H with $k < r$, we will refer to elements of $E(G)$ as edges and elements of $E(H)$ as hyperedges. If $G_1, \ldots, G_q$ are k-graphs on $[n]$, then $\bigcup G_i$ denotes the k-graph G on $[n]$ which has edge set $\bigcup E(G_i)$.

3.2 Counting Hypergraphs with Large Girth

As Balogh and Li [BL20] observed, if $\ell \geq 3$ and H has girth larger than ℓ, then H is uniquely determined by $\partial^2 H$, which we can view as the graph obtained by replacing each hyperedge of H by a clique. A key insight in proving Theorem 3.1.7 is that we can replace each hyperedge of H with a sparser graph B and still uniquely recover H from this graph. To this end, we say that a graph B is a *book* if there exist cycles $F_1, \ldots, F_k$ and an edge xy such that $B = \bigcup F_i$ and $E(F_i) \cap E(F_j) = \{xy\}$ for all $i \neq j$. In this case we call the cycles F_i the *pages* of B and we call the common edge xy the *spine* of B. The following lemma shows that if we replace each hyperedge in H by a book on r vertices which has small pages, then the vertex sets of books in the resulting graph are exactly the hyperedges of H.

Lemma 3.2.1. *Let H be an r-graph of girth larger than ℓ. If $\partial^2 H$ contains a book B on r vertices such that every page has length at most ℓ, then there exists a hyperedge $e \in E(H)$ such that $V(B) = e$.*

Proof. Let F be a cycle in $\partial^2 H$ with $V(F) = \{v_1, \ldots, v_p\}$ such that $v_i v_{i+1} \in E(\partial^2 H)$ for $i < p$ and $v_1 v_p \in E(\partial^2 H)$. If $p \leq \ell$ we claim that there exists an $e \in E(H)$ such that $V(F) \subseteq e$. Indeed, by definition of $\partial^2 H$ there exists some hyperedge $e_i \in E(H)$ with $v_i, v_{i+1} \in e_i$ for all $i < p$ and some hyperedge e_p with $v_1, v_p \in e_p$. If all of these e_i hyperedges are equal then we are done, so we may assume $e_1 \neq e_p$. Define i_1 to be the largest index such that $e_i = e_1$ for all $i \leq i_1$, define i_2 to be the largest index so that $e_i = e_{i_1+1}$ for all $i_1 < i \leq i_2$, and so on up to $i_q = p$, and note that $2 \leq q \leq p$ since $e_1 \neq e_p$. If all the e_{i_j} hyperedges are distinct, then they form a Berge q-cycle in H since $v_{1+i_j} \in e_{i_j} \cap e_{1+i_j} = e_{i_j} \cap e_{i_{j+1}}$ for all j, a contradiction. Thus we can assume $e_{i_j} = e_{i_{j'}}$ for some $j < j'$. We can further assume that $e_{i_s} \neq e_{i_{s'}}$ for any $j \leq s < s' < j'$, as otherwise we could replace j, j' with s, s'. Finally note that $j < j' - 1$, as otherwise we would have $e_{i_j} = e_{i_{j'}} = e_{i_j+1}$, contradicting the maximality of i_j. We conclude that the distinct hyperedges $e_{i_j}, e_{i_{j+1}}, \dots, e_{i_{j'-1}}$ form a Berge $(j'-j)$-cycle with $2 \le j'-j \le \ell$ in H, a contradiction. This proves the claim.

Now let B be a book with spine xy and pages $F_1, \dots, F_k$ of length at most ℓ. By the claim there exist hyperedges $e_1, \dots, e_k \in E(H)$ such that $V(F_i) \subseteq e_i$ for all i, and in particular $x, y \in e_i$ for all i. Because H is linear, this implies that all of these hyperedges are equal and we have $V(B) \subseteq e_1$. If B has r vertices, then we further have $V(B) = e_1$. □

Proof of Theorem 3.1.7. With $\lambda := \lceil (r-2)/(\ell-2) \rceil$, we observe for all $\ell, r \ge 3$ that there exists a book graph B on r vertices $\{x_1, \dots, x_r\}$ with $r-1+\lambda$ edges $f_1, \dots, f_{r-1+\lambda}$. Indeed if $\ell-2$ divides $r-2$ one can take λ copies of C_ℓ which share a common edge, and otherwise one can take $\lambda-1$ copies of C_ℓ and a copy of C_p with $p = r-(\lambda-1)(\ell-2) \ge 3$. From now on we let B denote this book graph. If $f_i = \{x_j, x_{j'}\} \in E(B)$ and $e = \{v_1, \dots, v_r\} \subseteq [n]$ is any r-set with $v_1 < \dots < v_r$, define $\phi_i(e) = \{v_j, v_{j'}\}$. If H is an r-graph on $[n]$, define $\phi_i(H)$ to be the graph on $[n]$ which has all edges of the form $\phi_i(e)$ for $e \in E(H)$; so in particular $\bigcup \phi_i(H)$ is the graph obtained by replacing each hyperedge of H with a copy of B. See Figure 3.4 for an example.

Let $\mathcal{H}_{m,n}$ denote the set of r-graphs on $[n]$ with m hyperedges and girth more than ℓ, and let $\mathcal{G}_{m,n}$ be the set of graphs on $[n]$ with m edges and girth more than ℓ. We claim that ϕ_i maps $\mathcal{H}_{m,n}$ to $\mathcal{G}_{m,n}$. Indeed, if $H \in \mathcal{H}_{m,n}$, then each hyperedge of H contributes a distinct edge to $\phi_i(H)$ since H is linear, so $e(\phi_i(H)) = e(H) = m$. One can show that if $\phi_i(e_1), \dots, \phi_i(e_p)$ form a p-cycle in $\phi_i(H)$, then $e_1, \dots, e_p$ form a Berge p-cycle in H; so

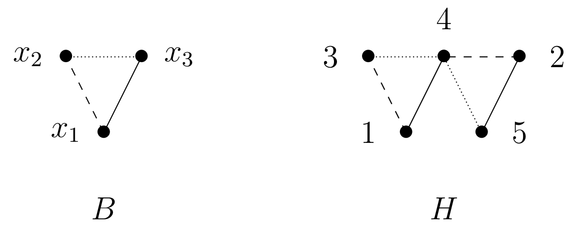

Figure 3.4: A book B on 3 vertices and a 3-graph H on [5] which consists of two hyperedges sharing a vertex. If $f_1 = \{x_1, x_2\}$, then $\phi_1(H)$ is the graph on [5] using the two dashed edges 13 and 24 (since these two pairs consist of the two smallest elements of each hyperedge of H).

$H \in \mathcal{H}_{m,n}$ implies $\phi_i(H)$ does not contain a cycle of length at most ℓ.

Let $\mathcal{G}^t_{m,n} = \{(G_1, G_2, \ldots, G_t) : G_i \in \mathcal{G}_{m,n}\}$ and define the map $\phi : \mathcal{H}_{m,n} \to \mathcal{G}^{r-1+\lambda}_{m,n}$ by

$$\phi(H) = (\phi_1(H), \ldots, \phi_{r-1+\lambda}(H)).$$

We claim that this map is injective. Indeed, fix some $H \in \mathcal{H}_{m,n}$ and let $\mathcal{B}(G)$ denote the set of books B in the graph $G := \bigcup \phi_i(H) \subseteq \partial^2 H$. By definition of ϕ we have $E(H) \subseteq \mathcal{B}(G)$ for all H. Moreover, if $H \in \mathcal{H}_{m,n}$, then Lemma 3.2.1 implies $\mathcal{B}(G) \subseteq E(H)$. Thus $E(H)$ (and hence H) is uniquely determined by G, which is itself determined by $\phi(H)$, so the map is injective. In total we conclude

$$\mathrm{N}^r_m(n, \ell) = |\mathcal{H}_{m,n}| \le |\mathcal{G}^{r-1+\lambda}_{m,n}| = \mathrm{N}^2_m(n, \ell)^{r-1+\lambda},$$

proving Theorem 3.1.7. $\square$

Before going onward, let us show how Theorem 3.1.7 implies our desired upper bounds for the random Turán problem.

Proof of Upper Bound of Theorem 3.1.5. Let

$$p_0 = n^{-\frac{(r-1+\lambda)(k-1)}{2k-1}} (\log n)^{(r-1+\lambda)k}.$$

For $p \geq p_0$, define

$$m = p^{\frac{1}{(r-1+\lambda)k}} n^{1+\frac{1}{k}} \log n,$$

and note that this is large enough to apply Theorem 3.1.6 for $p \geq p_0$. Let Y_m denote the number of subgraphs of $H^r_{n,p}$ which are $\mathcal{C}^r_{[\ell]}$-free and have exactly m edges, and note that $\mathrm{ex}(H^r_{n,p}, \mathcal{C}^r_{[\ell]}) \geq m$ if and only if $Y_m \geq 1$. By Markov's inequality, Theorem 3.1.7, and Theorem 3.1.6:

$$\begin{aligned}
\Pr[Y_m \geq 1] &\leq \mathbb{E}[Y_m] = p^m \cdot \mathrm{N}^r_m(n, \ell) \\
&\leq p^m \cdot \mathrm{N}^2_m(n, \ell)^{r-1+\lambda} \\
&\leq \left(p^{\frac{1}{r-1+\lambda}} e^c (\log n)^{k-1} \left(\frac{n^{1+\frac{1}{k}}}{m} \right)^k \right)^{m(r-1+\lambda)} \\
&= \left(\frac{e^c}{\log n} \right)^{m(r-1+\lambda)}.
\end{aligned}$$

The right hand side converges to zero, so for $p \geq p_0$, a.a.s:

$$\mathrm{ex}(H^r_{n,p}, \mathcal{C}^r_{[\ell]}) < m.$$

As $\mathbb{E}[\mathrm{ex}(H^r_{n,p}, \mathcal{C}^r_{[\ell]})]$ is non-decreasing in p, the bound

$$\mathrm{ex}(H^r_{n,p}, \mathcal{C}^r_{[\ell]}) < n^{1+\frac{1}{\ell-1}} (\log n)^2$$

continues to hold a.a.s. for all $p < p_0$. $\square$

Proof of Upper Bound of Theorem 3.1.5. This proof is almost identical to the previous, so we omit some of the redundant details. Let $m = p^{\frac{1}{2r-3}} n^2 \log n$ and let Y_m denote the number of subgraphs of $H^r_{n,p}$ which are $\mathcal{C}^r_{[\ell]}$-free and have exactly m edges. By Markov's inequality, Theorem 3.1.7, and the trivial bound $\mathrm{N}^2_m(n,3) \le \binom{n^2}{m}$ which is valid for all m, we find for all p

$$\Pr[Y_m \ge 1] \le p^m (en^2/m)^{(2r-3)m} = (e/\log n)^m.$$

This quantity converges to zero, so we conclude the result by the same reasoning as in the previous proof. $\square$

3.3 Counting Hypergraphs avoiding a Single Berge Cycle

For arbitrary hypergraphs H, the map $\phi(H) = \partial^{r-1} H$ (let alone the map to $\partial^2 H$) is not injective. However, we will show that this map is "almost" injective when considering H which are $\mathcal{C}^r_\ell$-free. To this end, we say that a set of vertices $\{v_1, \ldots, v_r\}$ is a *core set* of an r-graph H if there exist distinct hyperedges $e_1, \ldots, e_r$ with $\{v_1, \ldots, v_r\} \setminus \{v_i\} \subseteq e_i$ for all i. The following observation shows that core sets are the only obstruction to $\phi(H) = \partial^{r-1} H$ being injective.

Lemma 3.3.1. *Let H be an r-graph. If $\{v_1, \ldots, v_r\}$ induces a K^{r-1}_r in $\partial^{r-1} H$, then either $\{v_1, \ldots, v_r\} \in E(H)$ or $\{v_1, \ldots, v_r\}$ is a core set of H.*

Proof. By assumption of $\{v_1, \ldots, v_r\}$ inducing a K_r^{r-1} in $\partial^{r-1}H$, for all i there exist $e_i' \in E(\partial^{r-1}H)$ with $e_i' = \{v_1, \ldots, v_r\} \setminus \{v_i\}$. By definition of $\partial^{r-1}H$, this means there exist (not necessarily distinct) $e_i \in E(H)$ with $e_i \supseteq e_i' = \{v_1, \ldots, v_r\} \setminus \{v_i\}$. Given this, either $e_i = \{v_1, \ldots, v_r\}$ for some i, or all of the e_i distinct, in which case $\{v_1, \ldots, v_r\}$ is a core set of H. In either case we conclude the result.

□

We next show that $\mathcal{C}_\ell^r$-free r-graphs have few core sets.

Lemma 3.3.2. *Let $\ell, r \geq 3$ and let H be a $\mathcal{C}_\ell^r$-free r-graph with m hyperedges. The number of core sets in H is at most $\ell^2 r^2 m$.*

Proof. We claim that H contains no core sets if $\ell \leq r$. Indeed, assume for contradiction that H contained a core set $\{v_1, \ldots, v_r\}$ with distinct hyperedges $e_i \supseteq \{v_1, \ldots, v_r\} \setminus \{v_i\}$. It is not difficult to see that the hyperedges $e_1, \ldots, e_\ell$ form a Berge ℓ-cycle, a contradiction to H being $\mathcal{C}_\ell^r$-free. Thus from now on we may assume $\ell > r$.

Let $\mathcal{A}_1$ denote the set of core sets in H, and for any $\mathcal{A}' \subseteq \mathcal{A}_1$ and $(r-1)$-set S, define $d_{\mathcal{A}'}(S)$ to be the number of core sets $A \in \mathcal{A}'$ with $S \subseteq A$. Observe that $d_{\mathcal{A}_1}(S) > 0$ for at most $\binom{r}{r-1}m = rm$ distinct $(r-1)$-sets S, since in particular S must be contained in a hyperedge of H.

Given $\mathcal{A}_i$, define $\mathcal{A}_i' \subseteq \mathcal{A}_i$ to be the core sets $A \in \mathcal{A}_i$ which contain an $(r-1)$-set S with $d_{\mathcal{A}_i}(S) \leq \ell r$, and let $\mathcal{A}_{i+1} = \mathcal{A}_i \setminus \mathcal{A}_i'$. Observe that $|\mathcal{A}_i'| \leq \ell r \cdot rm$ since each

$(r-1)$-set S with $d_{\mathcal{A}_i}(S) > 0$ is contained in at most ℓr elements of $\mathcal{A}_i'$. In particular,

$$|\mathcal{A}_1| \leq (\ell - r) \cdot \ell r^2 m + |\mathcal{A}_{\ell-r+1}| \leq \ell^2 r^2 m + |\mathcal{A}_{\ell-r+1}|. \tag{3.3}$$

Assume for the sake of contradiction that $\mathcal{A}_{\ell-r+1} \neq \emptyset$. We prove by induction on $r \leq i \leq \ell$ that one can find distinct vertices $v_1, \ldots, v_i$ and distinct hyperedges $e_1, \ldots, e_{i-1}, \tilde{e}_i$ such that $v_j, v_{j+1} \in e_j$ for $1 \leq j < i$ and $v_1, v_i \in \tilde{e}_i$, and such that $\{v_i, v_{i-1}, \ldots, v_{i-r+2}, v_1\} \in \mathcal{A}_{\ell-i+1}$. For the base case, consider any $\{v_r, v_{r-1}, \ldots, v_1\} \in \mathcal{A}_{\ell-r+1}$. As this is a core set, there exist distinct hyperedges $e_j \supseteq \{v_1, \ldots, v_r\} \setminus \{v_{j+2}\}$ and $\tilde{e}_r \supseteq \{v_1, \ldots, v_r\} \setminus \{v_2\}$, proving the base case of the induction.

Assume that we have proven the result for $i < \ell$, which in particular implies that $\{v_i, v_{i-1}, \ldots, v_{i-r+2}, v_1\} \in \mathcal{A}_{\ell-i+1}$. Thus we have $\{v_i, v_{i-1}, \ldots, v_{i-r+2}, v_1\} \notin \mathcal{A}_{\ell-i}'$, so there exists a set of vertices $\{u_1, \ldots, u_{\ell r+1}\}$ such that $\{v_i, v_{i-1}, \ldots, v_{i-r+3}, v_1, u_j\} \in \mathcal{A}_{\ell-i}$ for all j. Because $|\bigcup_{k=1}^{i-1} e_k| \leq \ell r$, there exists some j such that $u_j \notin \bigcup_{k=1}^{i-1} e_k$. For this j, let $v_{i+1} := u_j$ and let $e_i, \tilde{e}_{i+1}$ be distinct hyperedges containing v_i, v_{i+1} and v_1, v_{i+1} respectively, which exist by assumption of this being a core set. Note that v_{i+1} is distinct from every other $v_{i'}$ since $v_{i'} \in \bigcup_{k=1}^{i-1} e_k$ for $i' \leq i$, and similarly the hyperedges $e_i, \tilde{e}_{i+1}$ are distinct from every hyperedge $e_{i'}$ with $i' < i$ since these new hyperedges contain $v_{i+1} \notin \bigcup_{k=1}^{i-1} e_k$. This proves the inductive step and hence the claim. The $i = \ell$ case of this claim implies that H contains a Berge ℓ-cycle, a contradiction. Thus $\mathcal{A}_{\ell-r+1} = \emptyset$, and the result follows by (3.3). □

Combining these two lemmas gives the following result, which allows us to reduce from r-graphs to $(r-1)$-graphs. We recall that $\mathrm{N}^r_{[m]}(n,\mathcal{F})$ denotes the number of n-vertex $\mathcal{F}$-free r-graphs on at most m hyperedges.

Proposition 3.3.3. *For each $\ell, r \geq 3$, there exists $c = c(\ell, r)$ such that*

$$\mathrm{N}^r_{[m]}(n, \mathcal{C}^r_\ell) \leq 2^{cm} \cdot \mathrm{N}^{r-1}_{[m]}(n, \mathcal{C}^{r-1}_\ell)^r.$$

Proof. If $e = \{v_1, v_2, \ldots, v_r\} \subseteq [n]$ is any r-set with $v_1 < v_2 < \cdots < v_r$, let $\phi_i(e) = \{v_1, \ldots, v_r\} \setminus \{v_i\}$. Given an r-graph H on $[n]$, let $\phi_i(H)$ be the $(r-1)$-graph on $[n]$ with edge set $\{\phi_i(e) : e \in E(H)\}$, and define $\phi(H) = (\phi_1(H), \phi_2(H), \ldots, \phi_r(H))$ and $\psi(H) = (\phi(H), E(H))$. Observe that $\bigcup \phi_i(H) = \partial^{r-1} H$. Let $\mathcal{H}_{[m],n}$ denote the set of all r-graphs on $[n]$ with at most m hyperedges which are $\mathcal{C}^r_\ell$-free, and let $\phi(\mathcal{H}_{[m],n}), \psi(\mathcal{H}_{[m],n})$ denote the image sets of $\mathcal{H}_{[m],n}$ under these respective maps. Observe that ψ is injective since it records $E(H)$, so it suffices to bound how large $\psi(\mathcal{H}_{[m],n})$ can be.

Let $\mathcal{G}_{[m],n}$ denote the set of $(r-1)$-graphs on $[n]$ which have at most m edges and which are $\mathcal{C}^{r-1}_\ell$-free. It is not difficult to see that $\phi(\mathcal{H}_{[m],n}) \subseteq \mathcal{G}^r_{[m],n}$. We observe by Lemmas 3.3.1 and 3.3.2 that for any $(G_1, G_2, \ldots, G_r) \in \phi(\mathcal{H}_{[m],n})$, say with $\phi(H) = (G_1, \ldots, G_r)$, there are at most $(1 + \ell^2 r^2)m$ copies of K^{r-1}_r in $\bigcup G_i = \partial^{r-1} H$. We also observe that if $((G_1, G_2, \ldots, G_r), E) \in \psi(\mathcal{H}_{[m],n})$, then E is a set of at most m copies of K^{r-1}_r in $\bigcup G_i$. Thus given any $(G_1, \ldots, G_r) \in \phi(\mathcal{H}_{[m],n}) \subseteq \mathcal{G}^r_{[m],n}$, there are at most $2^{(1+\ell^2 r^2)m}$ choices of E such that $((G_1, \ldots, G_r), E) \in \psi(\mathcal{H}_{[m],n})$. We conclude that

$$\mathrm{N}_{[m]}(n, \mathcal{C}^r_\ell) = |\mathcal{H}_{[m],n}| \leq |\mathcal{G}_{[m],n}|^r \cdot 2^{(1+\ell^2 r^2)m} = \mathrm{N}^r_{[m]}(n, \mathcal{C}^{r-1}_\ell)^r \cdot 2^{(1+\ell^2 r^2)m},$$

proving the result. ☐

Applying this proposition repeatedly gives $\mathrm{N}^r_{[m]}(n, \mathcal{C}^r_\ell) \leq 2^{cm}\mathrm{N}^2_{[m]}(n, C_\ell)^{r!/2}$. Combining this with the trivial inequality $\mathrm{N}^r_m(n, \mathcal{C}^r_\ell) \leq \mathrm{N}^r_{[m]}(n, \mathcal{C}^r_\ell)$ gives Theorem 3.1.9.

3.4 Lower Bounds: Random Homomorphisms

To prove lower bounds for $\mathrm{ex}(H^r_{n,p}, \mathcal{C}^r_{[\ell]})$, we use homomorphisms similar to Foucaud, Krivelevich and Perarnau [FKP15] and Perarnau and Reed [PR17]. If F and F' are hypergraphs and $\chi : V(F) \to V(F')$ is any map, we let $\chi(e) = \{\chi(u) : u \in e\}$ for any $e \in E(F)$. For two r-graphs F and F', a map $\chi : V(F) \to V(F')$ is a *homomorphism* if $\chi(e) \in E(F')$ for all $e \in E(F)$, and χ is a *local isomorphism* if χ is a homomorphism and $\chi(e) \neq \chi(f)$ whenever $e, f \in E(F)$ with $e \cap f \neq \emptyset$. A key lemma is the following:

Lemma 3.4.1. *If $F \in \mathcal{C}^r_{[\ell]}$ and $\chi : F \to F'$ is a local isomorphism, then F' has girth at most ℓ.*

Proof. Let F be a Berge p-cycle with $p \leq \ell$ and $E(F) = \{e_1, e_2, \ldots, e_p\}$. Then there exist distinct vertices $v_1, v_2, \ldots, v_p$ such that $v_i \in e_i \cap e_{i+1}$ for $i < p$ and $v_p \in e_p \cap e_1$. First assume there exists $i \neq j$ such that $\chi(e_i) = \chi(e_j)$. By reindexing, we can assume $\chi(e_1) = \chi(e_k)$ for some $k > 1$, and further that $\chi(e_i) \neq \chi(e_j)$ for any $1 \leq i < j < k$. Note that $k \geq 3$ since $e_1 \cap e_2 \neq \emptyset$ and χ is a local isomorphism. If we also have $\chi(v_i) \neq \chi(v_j)$ for all $1 \leq i < j < k$, then $\chi(v_i) \in \chi(e_i) \cap \chi(e_{i+1})$ for $i < k-1$ and $\chi(v_{k-1}) \in \chi(e_{k-1}) \cap \chi(e_1)$, so $\chi(e_1), \chi(e_2), \ldots, \chi(e_{k-1})$ is the edge set of a Berge $(k-1)$-cycle in F' as required.

Suppose $\chi(v_i) = \chi(v_j)$ for some $1 \leq i < j < k$, and as before we can assume there exists no $i \le i' < j' < j$ with $\chi(v_{i'}) = \chi(v_{j'})$. Then $\chi(v_i), \chi(v_{i+1}), \ldots, \chi(v_{j-1})$ are distinct vertices with $\chi(v_h) \in \chi(e_h) \cap \chi(e_{h+1})$ for $i \le h < j-1$ and $\chi(v_{j-1}) \in \chi(e_{j-1}) \cap \chi(e_1)$. Note that $\chi(v_i) \neq \chi(v_{i+1})$ since this would imply $|\chi(e_i)| < r$, contradicting that χ is a homomorphism, so $j > i+1$. Thus the hyperedges $\chi(e_i), \chi(e_{i+1}), \ldots, \chi(e_{j-1})$ form a Berge $(j-i)$-cycle in F' with $j - i \ge 2$ as desired.

This proves the result if $\chi(e_i) = \chi(e_j)$ for some $i \neq j$. If this does not happen and the $\chi(v_i)$ are all distinct, then F' is a Berge p-cycle, and if $\chi(v_i) = \chi(v_j)$ then the same proof as above gives a Berge $(j-i)$-cycle in F'. □

The following lemma allows us to find a relatively dense subgraph of large girth in any r-graph whose maximum i-degree is not too large, where the *i-degree* of an i-set S is the number of hyperedges containing S.

Lemma 3.4.2. *Let $\ell, r \ge 3$ and let H be an r-graph with maximum i-degree Δ_i for each $i \ge 1$. If $t \ge r^2 4^r \Delta_i^{1/(r-i)}$ for all $i \ge 1$, then H has a subgraph H' of girth larger than ℓ with*

$$e(H') \ge \mathrm{ex}(t, \mathcal{C}^r_{[\ell]}) t^{-r} \cdot e(H).$$

Proof. Let J be an extremal $\mathcal{C}^r_{[\ell]}$-free r-graph on t vertices and $\chi : V(H) \to V(J)$ chosen uniformly at random. Let $H' \subseteq H$ be the random subgraph which keeps the hyperedge $e \in E(H)$ if

(1) $\chi(e) \in E(J)$, and

(2) $\chi(e) \neq \chi(f)$ for any other $f \in E(H)$ with $|e \cap f| \neq 0$.

We claim that H' is $\mathcal{C}^r_{[\ell]}$-free. Indeed, assume H' contained a subgraph F isomorphic to some element of $\mathcal{C}^r_{[\ell]}$. Let F' be the subgraph of J with $V(F') = \{\chi(u) : u \in V(F)\}$ and $E(F') = \{\chi(e) : e \in E(F)\}$, and note that $F \subseteq H'$ implies that each hyperedge of F satisfies (1), so every element of $E(F')$ is a hyperedge in J. By conditions (1) and (2), χ is a local isomorphism from F to F'. By Lemma 3.4.1, $F' \subseteq J$ contains a Berge cycle of length at most ℓ, a contradiction to J being $\mathcal{C}^r_{[\ell]}$-free.

It remains to compute $\mathbb{E}[e(H')]$. Given $e \in E(H)$, let A_1 denote the event that (1) is satisfied, let $E_i = \{f \in E(H) : |e \cap f| = i\}$, and let A_2 denote the event that $\chi(f) \not\subseteq \chi(e)$ for any $f \in \bigcup_i E_i$, which in particular implies (2) for the hyperedge e. It is not too difficult to see that $\Pr[A_1] = r!e(J)t^{-r}$, and that for any $f \in E_i$ we have $\Pr[\chi(f) \subseteq \chi(e)|A_1] = (r/t)^{r-i}$. Note for each $i \geq 1$ that $|E_i| \leq 2^r\Delta_i$, as e has at most 2^r subsets of size i each of i-degree at most Δ_i. Taking a union bound we find

$$\Pr[A_2|A_1] \geq 1 - \sum_{i=1}^{r} |E_i|(r/t)^{r-i} \geq 1 - \sum_{i=1}^{r} 2^r\Delta_i(r/t)^{r-i} \geq 1 - \sum_{i=1}^{r} r^{-1}2^{-r} \geq \frac{1}{2},$$

where the second to last inequality used $(r4^r)^{i-r} \geq r^{-1}4^{-r}$ for $i \leq r$. Consequently

$$\Pr[e \in E(H')] = \Pr[A_1] \cdot \Pr[A_2|A_1] \geq r!e(J)t^{-r} \cdot \frac{1}{2} \geq e(J)t^{-r},$$

and linearity of expectation gives $\mathbb{E}[e(H')] \geq e(J)t^{-r} \cdot e(H) = \mathrm{ex}(t, \mathcal{C}^r_{[\ell]})t^{-r} \cdot e(H)$. Thus there exists some $\mathcal{C}^r_{[\ell]}$-free subgraph $H' \subseteq H$ with at least $\mathrm{ex}(t, \mathcal{C}^r_{[\ell]})t^{-r} \cdot e(H)$ hyperedges.

$\square$

Proof of lower bound of Theorem 3.1.5. By the Chernoff bound one can show for

$$p \geq p_1 := n^{\frac{-(r-1)(\ell-1-k)}{\ell-1}}$$

that a.a.s. $H^r_{n,p}$ has maximum i-degree at most $\Theta(pn^{r-i})$ for all i. If Conjecture 3.1.4 is true, then a.a.s for $p \geq p_1$ Lemma 3.4.2 with $t = \Theta(p^{1/(r-1)}n)$ gives:

$$\mathrm{ex}(H^r_{n,p}, \mathcal{C}^r_{[\ell]}) = \Omega(t^{-r}\mathrm{ex}(t, \mathcal{C}^r_{[\ell]})pn^r) = p^{\frac{1}{(r-1)k}} n^{1+\frac{1}{k}-o(1)}.$$

This gives the desired result. $\square$

For Theorem 3.1.3 we use the following variant of Lemma 3.4.2:

Lemma 3.4.3. *Let H be an r-graph and let $R_{\ell,v}(H)$ be the number of Berge ℓ-cycles in H on v vertices. For all $t \geq 1$, H has a subgraph H' of girth larger than 3 with*

$$e(H') \geq \left(e(H)t^{2-r} - \sum_{\ell=2}^{3} \sum_{v} t^{2-v} R_{\ell,v}(H) \right) e^{-c\sqrt{\log t}},$$

where $c > 0$ is an absolute constant.

Proof. By work of Ruzsa and Szemerédi [RS78] and Erdős, Frankl, Rödl [EFR86], it is known for all t that there exists a $\mathcal{C}^r_{[3]}$-free r-graph J on t vertices with $t^2 e^{-c\sqrt{\log t}}$ hyperedges. Choose a map $\chi : V(H) \to V(J)$ uniformly at random and define $H'' \subseteq H$ to be the subgraph which keeps a hyperedge $e = \{v_1, \ldots, v_r\} \in E(H)$ if and only if $\chi(e) \in E(J)$.

We claim that if e_1, e_2, e_3 form a Berge triangle in H'', then $\chi(e_1) = \chi(e_2) = \chi(e_3)$. Observe that if v_1, v_2, v_3 are vertices with $v_i \in e_i \cap e_{i+1}$, then we must have e.g. $\chi(v_1) \neq \chi(v_2)$, as otherwise $|\chi(e_2)| < r$. Because J is linear we must have $|\chi(e_i) \cap \chi(e_j)| \in \{1, r\}$. These hyperedges can not all intersect in 1 vertex since this together with the distinct vertices $\chi(v_1), \chi(v_2), \chi(v_3)$ defines a Berge triangle in J, so we must have say $\chi(e_1) =$

$\chi(e_2)$. But this means $\chi(v_3), \chi(v_2)$ are distinct vertices in $\chi(e_1) = \chi(e_2)$ and $\chi(e_3)$, so $|\chi(e_1) \cap \chi(e_3)| > 1$ and we must have $\chi(e_1) = \chi(e_3)$ as desired.

The probability that a given Berge triangle C on v vertices in H maps to a given hyperedge in J is at most $(r/t)^v$ (since this is the probability that every vertex of C maps into the edge of J). By linearity of expectation, H'' contains at most $\sum_v R_{3,v}(H)e(J)(r/t)^v$ Berge triangles in expectation. An identical proof shows that H'' contains at most $\sum_v R_{2,v}(H)e(J)(r/t)^v$ Berge 2-cycles in expectation. We can then delete a hyperedge from each of these Berge cycles in H'' to find a subgraph H' with

$$\mathbb{E}[e(H')] \geq e(J)t^{-r} \cdot e(H) - \sum_{\ell=2}^{3} \sum_{v} R_{\ell,v}(H)e(J)(r/t)^v.$$

The result follows since $e(J) = t^2 e^{-c\sqrt{\log t}}$. $\square$

Proof of Lower Bound of Theorem 3.1.3. By Markov's inequality one can show that a.a.s. $R_{3,3r-3}(H^r_{n,p}) = O(p^3 n^{3r-3})$. By the Chernoff bound we have a.a.s. that $e(H^r_{n,p}) = \Omega(pn^r)$, so if we take $t = p^{2/(2r-3)}n(\log n)^{-1}$, then a.a.s. $t^{5-3r}R_{3,3r-3}(H^r_{n,p})$ is significantly smaller than $t^{2-r}e^(H^r_{n,p})$. A similar result holds for each term $t^{2-v}R_{\ell,v}(H^r_{n,p})$ with $\ell = 2, 3$ and $v \leq \ell(r-1)$, so by Lemma 3.4.3 we conclude $\mathrm{ex}(H^r_{n,p}, \mathcal{C}^r_{[3]})] \geq p^{1/(2r-3)}n^{2-o(1)}$ a.a.s., proving the lower bound in Theorem 3.1.3. $\square$

We note that the proof of Lemma 3.4.3 fails for larger ℓ. In particular, a Berge 4-cycle can appear in H'' by mapping onto two edges in J intersecting at a single vertex, and with this the bound becomes ineffective.

3.4.1 Relative Turán Problems

The lower bound techniques discussed above extend to a more general set of problems, which we briefly discuss here. Given an r-graph H and a set of r-graphs $\mathcal{F}$, we define the *relative Turán number* $\mathrm{ex}(H,\mathcal{F})$ to be the largest size of an $\mathcal{F}$-free subgraph of H. The central problem here is to bound $\mathrm{ex}(H,\mathcal{F})$ in terms of parameters of H. One particular problem of this form is the following:

Problem 3.4.4. *Given a family of r-graphs $\mathcal{F}$, determine lower bounds for* $\mathrm{ex}(H,\mathcal{F})$ *in terms of the number of edges of H and the maximum degree of H.*

In some sense this problem is a "worst case analysis" of the function $\mathrm{ex}(H,\mathcal{F})$, which complements the "average case analysis" which comes from studying $\mathrm{ex}(H^r_{n,p},\mathcal{F})$.

One result in the spirit of Problem 3.4.4 was proven by Perarnau and Reed [PR17]: for any graph G with maximum degree at most Δ,

$$\mathrm{ex}(G,K_{a,b}) = \Omega\left(\frac{\mathrm{ex}(\Delta,K_{a,b})}{\Delta^2}\right)\cdot e(G). \tag{3.4}$$

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 (3.4) 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 (3.4), the following was essentially conjectured by Foucaud, Krivelevich, and Perarnau [FKP15]:

Conjecture 3.4.5. *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, we proved that this naive conjecture for hypergrpahs is very false, even for hypergraph analogs of $K_{a,b}$.

Theorem 3.4.6 ([SV21])**.** *Let $K_{a,b,c}$ be the complete 3-partite 3-uniform hypergraph with parts of sizes $a \le b \le c$. 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

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

We emphasize that these bounds are *not* what one gets by considering $H = K_n^3$, since in this case it is conjectured that $\mathrm{ex}(K_n^3, K_{a,b,c}) = \Theta(n^{3-\frac{1}{ab}}) = \Theta(\Delta^{\frac{-1}{2ab}})\cdot e(K_n^3)$. Thus the natural analog of (3.4) fails for 3-uniform hypergraphs. A generalization of

Theorem 3.4.6 for complete r-partite r-graphs is also proven in [SV21]. We briefly outline the ideas of this proof.

Sketch of Proof. Let H be a 3-graph with maximum degree Δ and let $D = \Delta^{b/(1+b)}$. If H has maximum 2-degree at most D, then an analog of Lemma 3.4.2 gives the desired lower bound. If every hyperedge of H contains a pair of vertices with codegree D, then we can form a graph G consisting of these pairs with large codegrees, find a subgraph $G' \subseteq G$ which contains no $K_{a,b}$ (again by using an analog of Lemma 3.4.2), and then lift this to a large $K_{a,b,c}$-free subgraph of H. By making some reductions, one can more or less assume one of these two cases happen, giving the desired lower bound.

For the upper bound, let $n = \Delta^{1/(1+b)}$ and take H to be the complete 3-partite 3-graph K_{n,n,n^b}. We note that this example is motivated by the proof of the lower bound: our parameters are such that H has maximum degree $n^{1+b} = \Delta$ and maximum 2-degree $n^b = D$. One can show by using an alteration of the proof of the Kővári-Sós-Turán theorem that $\mathrm{ex}(H, K_{a,b,c}) = O(\Delta^{\frac{-1}{ab+a}}) \cdot e(H)$, proving the result. $\square$

Given our work on Berge cycles, it is natural to ask for lower bounds on $\mathrm{ex}(H, \mathcal{C}^r_{[\ell]})$. Unfortunately one can only prove trivial bounds on $\mathrm{ex}(H, \mathcal{C}^r_{[\ell]})$ for $r \geq 3$ because $\mathcal{C}^r_{[\ell]}$ contains sunflowers. A hypergraph F is said to be a *sunflower* if there exists a set K called the kernel such that $e \cap f = K$ for any distinct edges e, f in F. If $F \in \mathcal{C}^r_\ell$ is a sunflower with kernel of size k, and if H is a sunflower with kernel of size k and Δ edges, then $\mathrm{ex}(H, F) = \ell - 1 = O(\Delta^{-1}) \cdot e(H)$. Thus the best lower bound one can prove is

$\mathrm{ex}(H, \mathcal{C}^r_{[\ell]}) = \Omega(\Delta^{-1}) \cdot e(H)$, and this trivially holds by considering a largest matching in H. With Verstraëte, we showed that sunflowers are the only obstruction to obtaining non-trivial bounds.

Theorem 3.4.7 ([SV20]). *Let $\widehat{\mathcal{C}}^r_{[\ell]}$ consist of all the elements of $\mathcal{C}^r_{[\ell]}$ which are not sunflowers. If $\ell, r \geq 3$ are such that Conjecture 3.1.4 holds, then for all r-graphs H with maximum degree at most Δ, we have*

$$\mathrm{ex}(H, \widehat{\mathcal{C}}^r_{[\ell]}) \geq \Delta^{-1+\frac{1}{(r-1)\lfloor \ell/2 \rfloor}-o(1)} \cdot e(H), \tag{3.5}$$

and this bound is tight up to the $o(1)$ term in the exponent for $H = K^r_{\Delta^{1/(r-1)}}$.

We note that the $r = 2$ case of Theorem 3.4.7 was originally proven by Perarnau and Reed [PR17].

3.5 Concluding remarks

In this chapter, we extended ideas of Balogh and Li to bound the number of n-vertex r-graphs with m edges and girth more than ℓ in terms of the number of n-vertex graphs with m edges and girth more than ℓ. The reduction is best possible when $m = \Theta(n^{\ell/(\ell-1)})$ and $\ell - 2$ divides $r - 2$. Theorem 3.1.9 shows that similar reductions can be made when forbidding a single family of Berge cycles.

By using variations of our method, we can prove the following generalization. For a graph F, a hypergraph H is a *Berge-F* if there exists a bijection $\phi : E(F) \to E(H)$

such that $e \subseteq \phi(e)$ for all $e \in E(F)$. Let $\mathcal{B}^r(F)$ denote the family of r-uniform Berge-F. We can prove the following extension of Theorem 3.1.9: if there exists a vertex $v \in V(F)$ such that $F - v$ is a forest, then there exists $c = c(F, r)$ such that

$$\mathrm{N}^r_m(n, \mathcal{B}^r(F)) \leq 2^{cm} \cdot \mathrm{N}^2_{[m]}(n, F)^{r!/2}.$$

For example, this result applies when F is a theta graph. We do not believe that the exponent $r!/2$ is optimal in general, and we propose the following problem.

Problem 3.5.1. *Let $\ell, r \geq 3$. Determine the smallest value $\beta = \beta(\ell, r) > 0$ such that there exists a constant $c = c(\ell, r)$ so that, for all $m, n \geq 1$,*

$$\mathrm{N}^r_m(n, \mathcal{C}^r_\ell) \leq 2^{cm} \cdot \mathrm{N}^2_{[m]}(n, C_\ell)^{\beta}.$$

Theorem 3.1.9 shows that $\beta \leq r!/2$ for all ℓ, r, but in principle it could be that $\beta = O_\ell(r)$. We claim without proof that it is possible to use variants of our methods to show $\beta(3, r), \beta(4, r) \leq \binom{r}{2}$, but beyond this we do not know any non-trivial upper bounds on β.

It seems likely that the following conjecture is true:

Conjecture 3.5.2. *Let $\ell, r \geq 3$ and $k = \lfloor \ell/2 \rfloor$. Then there exists $\gamma = \gamma(\ell, r)$ such that a.a.s.*

$$\mathrm{ex}(H^r_{n,p}, \mathcal{C}^r_{[\ell]}) = \begin{cases} n^{1+\frac{1}{\ell-1}+o(1)} & n^{-r+1+\frac{1}{\ell-1}} \leq p < n^{-\frac{\gamma(\ell-1-k)}{\ell-1}}, \\ p^{\frac{1}{\gamma k}} n^{1+\frac{1}{k}+o(1)} & n^{-\frac{\gamma(\ell-1-k)}{\ell-1}} \leq p \leq 1. \end{cases}$$

Conjecture 3.1.8 suggests the possible value $\gamma(\ell, r) = r - 1 + (r-2)/(\ell-2)$, which is the correct value for $\ell = 3$ by Theorem 3.1.3. We are not certain that this is the right value of γ in general, even when $r = 3$ and $\ell = 4$, and more generally Conjecture 3.1.4 is an obstacle for $r \geq 3$ and $\ell \geq 5$. Theorem 3.1.5 shows that if γ exists, then $(r-1)k \leq \gamma \leq (r-1+\lambda)k$ provided Conjecture 3.1.4 holds. It would be interesting as a test case to know if $\gamma(3,4) = 5/2$:

Problem 3.5.3. *Prove or disprove that Conjecture 3.5.2 holds with* $\gamma(3,4) = 5/2$.

This chapter contains material from: S. Spiro and J. Verstraëte, *"Counting Hypergraphs with Large Girth"*, Journal of Graph Theory, accepted (2021). The dissertation author was one of the primary investigators and authors of this paper.

Chapter 4

Spread Hypergraphs

4.1 Introduction

Throughout this chapter we allow our hypergraphs to have repeated edges. If A is a set of vertices of a hypergraph $\mathcal{H}$, we define the *degree of* A to be the number of edges of $\mathcal{H}$ containing A, and we denote this quantity by $d_{\mathcal{H}}(A)$, or simply by $d(A)$ if $\mathcal{H}$ is understood. We say that a hypergraph $\mathcal{H}$ is *q-spread* if it is non-empty and if $d(A) \leq q^{|A|}|\mathcal{H}|$ for all sets of vertices A. A hypergraph is said to be *r-bounded* if each of its edges have size at most r and it is *r-uniform* if all of its edges have size exactly r.

The notion of q-spread hypergraphs was introduced by Alweiss, Lovett, Wu, and Zhang [ALWZ20] where it was a key ingredient in their groundbreaking work which significantly improved upon the bounds on the largest size of a set system which contain no sunflower. Their method was refined by Frankston, Kahn, Narayanan, and Park [FKNP21] who proved the following.

Theorem 4.1.1 ([FKNP21])**.** *There exists an absolute constant K_0 such that the following holds. Let $\mathcal{H}$ be an r-bounded q-spread hypergraph on V. If W is a set of size $K_0(\log r)q|V|$ chosen uniformly at random from V, then W contains an edge of $\mathcal{H}$ with probability tending to 1 as r tends towards infinity.*

This theorem can be used to give short proofs of a number of important results, see Section 4.2 for some examples.

Kahn, Narayanan, and Park [KNP21] used a variant of the method from [FKNP21] to show that for certain q-spread hypergraphs, the conclusion of Theorem 4.1.1 holds for random sets W of size only $K_0q|V|$. They used this to determine the threshold for when a square of a Hamiltonian cycle appears in the random graph $G_{n,p}$, which was a long-standing open problem.

In a talk, Narayanan asked if there was a "smoother" definition of spread hypergraphs, one which interpolates between q-spread hypergraphs and hypergraphs like those in [KNP21] where the $\log r$ term of Theorem 4.1.1 can be dropped. The aim of this chapter is to provide such a definition.

Definition 1. *Let $0 < q \leq 1$ be a real number and $r_1 > \cdots > r_\ell$ positive integers. We say that a hypergraph $\mathcal{H}$ on V is $(q; r_1, \ldots, r_\ell)$-spread if $\mathcal{H}$ is non-empty, r_1-bounded, and if for all $1 \leq i < \ell$, every $A \subseteq V$ with $d(A) > 0$ and every integer j satisfying*

$r_i \geq |A| \geq j \geq r_{i+1}$ has

$$M_j(A) := |\{S \in \mathcal{H} : |A \cap S| \geq j\}| \leq q^j |\mathcal{H}|.$$

Roughly speaking, this condition says that every set A of r_i vertices intersects few edges of $\mathcal{H}$ in more than r_{i+1} vertices. Note that we always have

$$M_j(A) \leq \sum_{B \subseteq A : |B| = j} d(B),$$

and in practice it is often easiest to upper bound this sum rather than trying to upper bound $M_j(A)$ directly. As a warm-up, we show how this definition relates to the definition of being q-spread.

Proposition 4.1.2. *We have the following.*

(a) If $\mathcal{H}$ is $(q; r_1, \ldots, r_\ell, 1)$-spread, then it is q-spread.

(b) If $\mathcal{H}$ is q-spread and r_1-bounded, then it is $(4q; r_1, \ldots, r_\ell)$-spread for any sequence of integers r_i satisfying $r_i > r_{i+1} \geq \frac{1}{2} r_i$.

Proof. For (a), assume $\mathcal{H}$ is $(q; r_1, \ldots, r_\ell, 1)$-spread and let $r_{\ell+1} = 1$. Let A be a set of vertices of $\mathcal{H}$. If $A = \emptyset$, then $d(A) = |\mathcal{H}| = q^{|A|}|\mathcal{H}|$, so we can assume A is non-empty. If $d(A) = 0$, then trivially $d(A) \leq q^{|A|}|\mathcal{H}|$, so we can assume $d(A) > 0$. This means $|A| \leq r_1$ since in particular $\mathcal{H}$ is r_1 bounded. Thus there exists an integer $1 \leq i \leq \ell$ such that $r_i \geq |A| \geq r_{i+1}$, so the hypothesis that $\mathcal{H}$ is $(q; r_1, \ldots, r_\ell, 1)$-spread and $d(A) > 0$ implies

$$d(A) \leq M_{|A|}(A) \leq q^{|A|}|\mathcal{H}|,$$

proving that $\mathcal{H}$ is q-spread.

For (b), assume $\mathcal{H}$ is q-spread and r_1-bounded. If A is any set of vertices of $\mathcal{H}$, then for all $j \geq \frac{1}{2}|A|$ we have

$$M_j(A) \leq \sum_{B \subseteq A : |B| = j} d(B) \leq 2^{|A|} \cdot q^j |\mathcal{H}| \leq (4q)^j |\mathcal{H}|.$$

In particular, if $r_i \geq |A| \geq r_{i+1}$, then this bound holds for any $j \geq r_{i+1}$ since $r_{i+1} \geq \frac{1}{2} r_i \geq \frac{1}{2}|A|$. We conclude that $\mathcal{H}$ is $(4q; r_1, \ldots, r_\ell)$-spread. $\square$

We now state our main result for uniform hypergraphs, which says that a random set of size $C\ell q|V|$ will contain an edge of an r_1-uniform $(q; r_1, \ldots, r_\ell, 1)$-spread hypergraph with high probability as $C\ell$ tends towards infinity. An analogous result can be proven for non-uniform hypergraphs, but for ease of presentation we defer this result to Section 4.4.

Theorem 4.1.3. *There exists an absolute constant K_0 such that the following holds. Let $\mathcal{H}$ be an r_1-uniform $(q; r_1, \ldots, r_\ell, 1)$-spread hypergraph on V. If W is a set of size $C\ell q|V|$ chosen uniformly at random from V with $C \geq K_0$, then*

$$\Pr[W \text{ contains an edge of } \mathcal{H}] \geq 1 - \frac{K_0}{C\ell}.$$

We note that Theorem 4.1.3 with $\ell = \Theta(\log r)$ together with Proposition 4.1.2(b) implies Theorem 4.1.1 for uniform $\mathcal{H}$. In [KNP21], it is implicitly proven that the hypergraph $\mathcal{H}$ encoding squares of Hamiltonian cycles is a $(2n)$-uniform $(Cn^{-1/2}; 2n, C_0 n^{1/2}, 1)$-spread hypergraph for some appropriate constants C, C_0, so the $\ell = 2$ case of Theorem 4.1.3 suffices to prove the main result of [KNP21]. Thus, at least in the uniform case, Theorem 4.1.3 provides an interpolation between the results of [FKNP21, KNP21]. Theorem 4.1.3 can also be used to recover results from very recent work of Espuny Díaz and Person [EP21] who extended the results of [KNP21] to other spanning subgraphs of $G_{n,p}$.

4.2 Applications

In this section we showcase the power of Theorems 4.1.1 and 4.1.3 through a number of examples.

Theorem 4.2.1 ([JKV08]). *Let $H^r_{n,m}$ be the r-graph chosen uniformly at random amongst all r-graphs with n vertices and m edges. Then there exists a constant C such that if $m \geq Cn \log n$ and n is a multiple of r, then $H^r_{n,m}$ contains a perfect matching a.a.s.*

It is not too difficult to show that this bound on m is essentially best possible. We note that $H^r_{n,m}$ behaves very similarly to $H^r_{n,p}$ where $p = m/\binom{n}{r}$. In particular, one can use Theorem 4.2.1 to prove that $H^r_{n,p}$ contains a perfect matching a.a.s. if p is significantly larger than $n^{1-r} \log n$. Proving Theorem 4.2.1 for $r = 2$ is not hard, but the result for general r was thought to be very difficult, with its first proof due to Johansson, Kahn, and Vu [JKV08] using a rather involved argument. We will prove Theorem 4.2.1 in just a few lines with Theorem 4.1.3.

Proof. Let $\mathcal{H}$ be the hypergraph with vertex set $V = E(K^r_n)$ where each hyperedge S is a perfect matching of K_n^r. Observe that for any set $A \subseteq E(K_n^r)$, we have

$$
\begin{aligned}
d(A) \cdot |\mathcal{H}|^{-1} &= \frac{(n - r|A|)!}{(r!)^{n/r-|A|}(n/r - |A|)!} \cdot \frac{(r!)^{n/r}(n/r)!}{n!} \\
&= (r!)^{|A|} \binom{n/r}{|A|} \binom{n}{r|A|}^{-1} \frac{|A|!}{(r|A|)!} \\
&\leq (r!)^{|A|}(en/r|A|)^{|A|} \cdot (n/r|A|)^{-r|A|} \cdot (|A|)^{|A|} \cdot (r|A|/e)^{-r|A|} \\
&= (r!)^{|A|}e^{(r+1)|A|}n^{-(r-1)|A|} \leq (n/re^3)^{-(r-1)|A|}.
\end{aligned}
$$

Thus $\mathcal{H}$ is $(n/re^3)^{-r+1}$-spread. It is also (n/r)-uniform and has a ground set $V = E(K_n^r)$ of size $\binom{n}{r}$. By Theorem 4.1.3, we see that if m is at least as large as in our hypothesis, then with high probability a random m-subset of $\mathcal{H}$ will contain a hyperedge, i.e. $H_{n,m}^r$ will contain a perfect matching with high probability. $\square$

Another basic example is the following.

Proposition 4.2.2. *Let F be an r-graph and define*

$$
t(F) = \max \left\{ \frac{|E(F')|}{|V(F')|} : F' \subseteq F \right\}.
$$

Let $H_{n,m}^r$ be as in Theorem 4.2.1. There exists a constant $C(F)$ such that if $m \geq C(F)n^{r-1/t(F)}$, then $H_{n,m}^r$ contains a copy of F a.a.s.

A simple first moment argument shows that this bound is tight. One can prove Proposition 4.2.2 using a standard but somewhat tedious second moment argument, but using Theorem 4.1.3 gives a shorter proof.

Proof. Let $\mathcal{H}$ be the hypergraph on $E(K_n^r)$ whose hyperedges correspond to copies of F. Observe that $\mathcal{H}$ being q-spread is equivalent to having $(d(A)/|\mathcal{H}|)^{1/|A|} \leq q$ for all

$A \subseteq V = E(K_n^r)$. Any set $A \subseteq E(K_n^r)$ of positive degree in $\mathcal{H}$ forms some subgraph $F' \subseteq F$ with $|E(F')| = |A|$, and in this case

$$\left(\frac{d(A)}{|\mathcal{H}|}\right)^{1/|A|} \leq \left(\frac{n^{|V(F)|-|V(F')|}}{\binom{n}{|V(F)|}}\right)^{1/|A|} \leq |V(F)|^{|V(F)|} \cdot n^{-|V(F')|/|E(F')|}.$$

This implies that $\mathcal{H}$ is q-spread with

$$q = \max\{|V(F)|^{|V(F)|} \cdot n^{-|V(F')|/|E(F')|} : F' \subseteq F\} = |V(F)|^{|V(F)|} \cdot n^{-1/t(F)}.$$

Plugging this into Theorem 4.1.3 gives the result. $\square$

The study of q-spread hypergraphs was initiated by Alweiss, Lovett, Wu, and Zhang [ALWZ20] where they proved a slightly weaker version of Theorem 4.1.3. Their motivation came from the Erdős sunflower conjecture. A *k-sunflower* is a hypergraph with edges $S_1, \ldots, S_k$ such that there exists a set K called the *kernel* which has $S_i \cap S_j = K$ for all $i \neq j$.

Theorem 4.2.3 ([ALWZ20, Rao19, BCW21])**.** *There exists a constant C such that if $\mathcal{H}$ is an r-graph with more than $(Ck \log r)^r$ edges, then $\mathcal{H}$ contains a k-sunflower.*

We note that [ALWZ20] was the first to prove a theorem of this form, with [Rao19, BCW21] later giving better bounds in terms of k. When k is fixed, Theorem 4.2.3 gives a bound of the form $(\log r)^{r+o(1)}$. Prior to [ALWZ20], the best known bounds were of the form $r^{r-o(1)}$. It is a famous conjecture of Erdős that one can prove a bound of the form $c_k^{r+o(1)}$.

Proof. We prove the result by induction on r, the $r = 1$ case being trivial. Let $\mathcal{H}$ be an r-graph with at least $(Ck \log r)^r$ edges. If $\mathcal{H}$ is not q-spread with $q = (Ck \log r)^{-1}$, then there exists some $A \subseteq V(H)$ such that $d(A) \geq (Ck \log r)^{r-|A|}$. This means that the link hypergraph $\mathcal{H}_A = \{S \setminus A : S \in \mathcal{H}, A \subseteq S\}$ has size at least $(Ck \log r)^{r-|A|}$. Since $\mathcal{H}_A$ is an $(r - |A|)$-uniform hypergraph, by induction $\mathcal{H}_A$ contains a k-sunflower, say with edges $S_1 \setminus A, \ldots, S_k \setminus A \in \mathcal{H}_A$. It is not difficult to check that $S_1, \ldots, S_k \in \mathcal{H}$ forms a k-sunflower in $\mathcal{H}$. We conclude that any $\mathcal{H}$ with at least $(Ck \log r)^r$ edges which is not q-spread contains a k-sunflower, so from now on we may assume $\mathcal{H}$ is q-spread.

Possibly by adding isolated vertices to $\mathcal{H}$, we can assume that the size of the vertex set V of $\mathcal{H}$ is a multiple of $2k$. Let $V_1, \ldots, V_{2k}$ be a random partition of V such that each $V_i \subseteq V$ has size $(2k)^{-1}|V|$. This means that each V_i is a uniformly chosen set of V of size $(2k)^{-1}|V| = \frac{1}{2}C(\log r)q|V|$. Let 1_i be the indicator variable for the event that V_i contains an edge of $\mathcal{H}$. By Theorem 4.1.3, we have $\Pr[1_i = 1] \geq \frac{1}{2}$ provided C is sufficiently large. In this case, $\mathbb{E}[\sum 1_i] \geq k$, and hence there exists some partition $V_1, \ldots, V_{2k}$ such that $\sum 1_i \geq k$, which in particular means there exist k disjoint edges of $\mathcal{H}$. This is a k-sunflower in $\mathcal{H}$, proving the result. $\square$

For our last application, we say that a subgraph $G \subseteq K_n$ contains the square of a Hamiltonian cycle if there exist a cyclic ordering of the vertices $v_1, \ldots, v_n$ such that v_i is adjacent to v_{i+1} and v_{i+2} for all i in G. Let $G_{n,m}$ denote the random graph obtained by uniformly choosing a graph on n vertices with m edges. Kühn and Osthus [KO12] asked when $G_{n,m}$ contains the square a of a Hamiltonian cycle with high probability. One can show using Theorem 4.1.1 that $G_{n,m}$ will contain the square of a Hamiltonian cycle with high probability if $m \geq Cn^{3/2} \log n$. It turns out that one can do better by utilizing Theorem 4.1.3.

Theorem 4.2.4 ([KNP21])**.** *There exists an absolute constant C such that if $m \geq Cn^{3/2}$, then $G_{n,m}$ contains the square of a Hamiltonian cycle a.a.s.*

Proof. Let $\mathcal{H}$ be the hypergraph on $E(K_n)$ whose hyperedges are squares of Hamiltonian cycles. For this proof, we will rely on two elementary combinatorial results proven in [KNP21]. First, [KNP21, Proposition 2.1] stated in our language says that if $B \subseteq E(K_n)$ is such that it consists of $j \leq n/4$ edges and c non-trivial connected components, then

$$d(B) \leq (16)^j \left(n - \left\lceil \frac{j+c}{2} \right\rceil - 1\right)! \tag{4.1}$$

Second, [KNP21, Proposition 2.2] says that if $A \subseteq V$ has $d(A) > 0$, then the number of subsets $B \subseteq A$ with $|B| = j$ and which have c non-trivial connected components is at most

$$(8e)^j \binom{2|A|}{c}. \tag{4.2}$$

Using these results, we will show that $\mathcal{H}$ is $(q; 2n, 4\sqrt{n}, 1)$-spread for some $q = \Theta(n^{-1/2})$, and from this the result will follow from Theorem 4.1.3.

First observe that $\mathcal{H}$ is $(2n)$-uniform, i.e. squares of Hamiltonian cycles have $2n$ edges (provided n is sufficiently large). As noted in [KNP21, Equation (2)], one can prove that $\mathcal{H}$ is $\tilde{q}$-spread with $\tilde{q} = \Theta(n^{-1/2})$. This implies that for all $A \subseteq E(K_n)$ and j with

$2n \geq |A| \geq j \geq n/4$ that

$$M_j(A) \leq \sum_{B \subseteq A,\ |B|=j} d(B) \leq 2^{2n} \cdot \tilde{q}^j |\mathcal{H}| \leq (2^8 \tilde{q})^j |\mathcal{H}|. \tag{4.3}$$

If $2n \geq |A| \geq 4\sqrt{n}$ with $d(A) > 0$, and if $n/4 \geq j \geq 4\sqrt{n}$, then we can conclude that

$$M_j(A) \leq \sum_{B \subseteq A,\ |B|=j} d(B) \leq (128e)^j \sum_{c=1}^{j} \binom{4n}{c} \cdot (n - \left\lceil \frac{j+c}{2} \right\rceil - 1)! \tag{4.4}$$

where the second inequality follows by first partitioning the sum over $B \subseteq A$ based off of the number of connected components of B, then using (4.1) to upper bound $d(B)$, then using (4.2) to upper bound the number of B with a given number of connected components. Noting that $|\mathcal{H}| = (n-1)!/2 \geq \sqrt{n}(\frac{n-1}{e})^{n-1}$ for n sufficiently large, and that $(n - \lceil \frac{j+c}{2} \rceil - 1)! \leq 2\sqrt{n}(\frac{n-1}{e})^{n-1-j/2-c/2}$ for n sufficiently large, we have

$$\frac{(n - \lceil \frac{j+c}{2} \rceil - 1)!}{|\mathcal{H}|} \leq 2e^{j/2+c/2}(n-1)^{-j/2-c/2} \leq 2e^j (n/2)^{-j/2-c/2}.$$

Thus (4.2) is at most

$$\begin{aligned}(128e)^j |\mathcal{H}| \sum_{c=1}^{j} \left(\frac{4en}{c}\right)^c \cdot 2e^j (n/2)^{-j/2-c/2} &= 2(128e^2(n/2)^{-1/2})^j |\mathcal{H}| \sum_{c=1}^{j} \left(\frac{32e^2 n}{c^2}\right)^{c/2} \\ &\leq 2(128e^{100}(n/2)^{-1/2})^j |\mathcal{H}|, \end{aligned} \tag{4.5}$$

where this last step used that the function $(\alpha/x^2)^x$ is maximized when $x = e\sqrt{\alpha}$, i.e. at $e^{e\sqrt{\alpha}}$, so taking $x = c/2$ and $\alpha = 8e^2 n$ gives

$$\sum_{c=1}^{j} \left(\frac{32e^2 n}{c^2}\right)^{c/2} \leq j \cdot e^{e^2\sqrt{8n}} \leq e^{98j},$$

where this last step critically used that $j = \Omega(\sqrt{n})$. Combining (4.3) and (4.5), we conclude for all $2n \geq |A| \geq j \geq 4\sqrt{n}$ and $d(A) > 0$ that for some $q = \Theta(n^{-1/2})$, we have

$$M_j(A) \leq q^j |\mathcal{H}|.$$

Similarly for $A \subseteq V$ and j with $4\sqrt{n} \geq |A| \geq j \geq 1$ and $d(A) > 0$, we have

$$M_j(A) \leq \sum_{B \subseteq A,\ |B|=j} d(B) \leq (128e)^j \sum_{c=1}^{j} \binom{8\sqrt{n}}{c} (n - \left\lceil \frac{j+c}{2} \right\rceil - 1)!,$$

and essentially the same reasoning as before gives that this is at most $q^j|\mathcal{H}|$. Thus $\mathcal{H}$ is $(q; 2n, 4\sqrt{n}, 1)$-spread, giving the result. $\square$

We emphasize in the proceeding proof that one cannot prove $M_j(A) \leq q^j|\mathcal{H}|$ with $q = \Theta(n^{-1/2})$ for arbitrary $2n \geq |A| \geq j \geq 1$. Roughly speaking, this is because the expression $\sum_{B \subseteq A : |B|=j} d(B)$ has too many terms when $|A| \gg \sqrt{n} \gg j$. Indeed, when $j = 1$ one can check that this sum equals $|A| \cdot \frac{2n}{\binom{n}{2}} |\mathcal{H}| \approx |A| n^{-1} \cdot |\mathcal{H}|$, so there is no hope of succeeding here with $q = \Theta(n^{-1/2})$ unless $|A| = O(n^{1/2})$. This is why it is crucial that we split our bounds for $M_j(A)$ into separate ranges depending on $|A|, j$, and this partially motivates the definition of being $(q; r_1, \ldots, r_\ell)$-spread.

4.3 Proof of Theorem 4.1.3

Our approach borrows heavily from Kahn, Narayanan, and Park [KNP21]. We break our proof into three parts: the main reduction lemma, auxiliary lemmas to deal with some special cases, and a final subsection proving the theorem.

4.3.1 The Main Lemma

We briefly sketch our approach for proving Theorem 4.1.3. Let $\mathcal{H}$ be a hypergraph with vertex set V. We first choose a random set $W_1 \subseteq V$ of size roughly $q|V|$. If W_1

contains an edge of $\mathcal{H}$ then we would be done, but most likely we will need to try and add in an additional random set W_2 of size $q|V|$ and repeat the process. In total then we are interested in finding the smallest I such that $W_1 \cup \cdots \cup W_I$ contains an edge of $\mathcal{H}$ with relatively high probability. One way to guarantee that I is small would be if we had $|S \setminus W_1|$ small for most $S \in \mathcal{H}$ (i.e., most vertices of most edges $S \in \mathcal{H}$ are covered by W_1), and then that W_2 covered most of the vertices of most $S \setminus W_1$, and so on.

The condition that, say, $|S \setminus W_1|$ is small for most $S \in \mathcal{H}$ turns out to be too strong a condition to impose. However, if $\mathcal{H}$ is sufficiently spread, then we can guarantee a weaker result: for most $S \in \mathcal{H}$, there is an $S' \subseteq S \cup W_1$ such that $|S' \setminus W_1|$ is small. We can then discard S and focus only on S', and by iterating this repeatedly we obtain the desired result.

To be more precise, given a hypergraph $\mathcal{H}$, we say that a pair of sets (S, W) is *k-good* if there exists $S' \in \mathcal{H}$ such that $S' \subseteq S \cup W$ and $|S' \setminus W| \leq k$, and we say that the pair is *k-bad* otherwise. The next lemma shows that $(q; r, k)$-spread hypergraphs have few k-bad pairs with $S \in \mathcal{H}$ and W a set of size roughly $q|V|$. In the lemma statement we adopt the notation that $\binom{V}{m}$ is the set of subsets of V of size m.

Lemma 4.3.1. *Let $\mathcal{H}$ be an r-uniform n-vertex hypergraph on V which is $(q; r, k)$-spread. Let $C \geq 4$ and define $p = Cq$. If $pn \geq 2r$ and $p \leq \frac{1}{2}$, then*

$$\left|\left\{(S, W) : S \in \mathcal{H},\ W \in \binom{V}{pn},\ (S, W) \text{ is } k\text{-bad}\right\}\right| \leq 3(C/2)^{-k/2}|\mathcal{H}|\binom{n}{pn}.$$

Proof. Throughout this lemma we make frequent use of the identity

$$\binom{a-c}{b-c} \Big/ \binom{a}{b} = \binom{b}{c} \Big/ \binom{a}{c},$$

which follows from the simple combinatorial identity $\binom{a}{b}\binom{b}{c} = \binom{a}{c}\binom{a-c}{b-c}$.

For $t \leq r$, define

$$\mathcal{B}_t = \{(S, W) : S \in \mathcal{H},\ W \in \binom{V}{pn},\ (S, W) \text{ is } k\text{-bad},\ |S \cap W| = t\}.$$

Observe that the quantity we wish to bound is $\sum_t |\mathcal{B}_t|$, so it suffices to bound each term of this sum. From now on we fix some t and define

$$w = pn - t.$$

At this point we need to count the number of elements in $\mathcal{B}_t$, and there are several natural approaches that could be used. One way would be to first pick any $S \in \mathcal{H}$ and then count how many W satisfy $(S, W) \in \mathcal{B}_t$. Another approach would be to pick any set Z of size $r + w$ (which will be the size of $S \cup W$ since $|S \cap W| = t$) and then bound how many $S, W \subseteq Z$ have $(S, W) \in \mathcal{B}_t$. For some pairs the first approach is more efficient, and for others the second is. In particular, the second approach will be more effective whenever $Z = S \cup W$ contains few elements of $\mathcal{B}_t$.

With this in mind, we say that a set Z is *pathological* if

$$|\{S \in \mathcal{H} : S \subseteq Z,\ (S, Z \setminus S) \text{ is } k\text{-bad}\}| > N,$$

where

$$N := (C/2)^{-k/2}|\mathcal{H}|\binom{n-r}{w} \Big/ \binom{n}{w+r} = (C/2)^{-k/2}|\mathcal{H}|\binom{w+r}{r} \Big/ \binom{n}{r}.$$

We say that a pair (S, W) is *pathological* if the set $S \cup W$ is pathological and that (S, W) is *non-pathological* otherwise.

Claim 4.3.2. *The number of $(S, W) \in \mathcal{B}_t$ which are non-pathological is at most*

$$\binom{n}{r+w} N \binom{r}{t} = (C/2)^{-k/2} |\mathcal{H}| \binom{r}{t} \binom{n-r}{w}.$$

Proof. We identify each of the non-pathological pairs (S, W) by specifying $S \cup W$, then S, then $S \cap W$.

Observe that $S \cup W$ is a non-pathological set of size $r + w$, and in particular there are at most $\binom{n}{r+w}$ ways to make this first choice. Fix such a non-pathological set Z of size $r + w$. Observe that if (S, W) is k-bad with $S \cup W = Z$, then $(S, Z \setminus S)$ is also k-bad. Because Z is non-pathological, there are at most N choices for S such that $(S, Z \setminus S)$ is k-bad. Given S, there are at most $\binom{r}{t}$ choices for $S \cap W$. Multiplying the number of choices at each step gives the stated result. □

Claim 4.3.3. *The number of $(S, W) \in \mathcal{B}_t$ which are pathological is at most*

$$2(C/2)^{-k/2} |\mathcal{H}| \binom{r}{t} \binom{n-r}{w}$$

Proof. We identify these pairs by first specifying $S \in \mathcal{H}$, then $S \cap W$, then $W \setminus S$.

Note that S and $S \cap W$ can be specified in at most $|\mathcal{H}| \cdot \binom{r}{t}$ ways, and from now on we fix such a choice of S and $S \cap W$. It remains to specify $W \setminus S$, which will be some element of $\binom{V \setminus S}{w}$. Thus it suffices to count the number of $W' \in \binom{V \setminus S}{w}$ such that (S, W') is both k-bad and pathological.

For $W' \in \binom{V \setminus S}{w}$, define

$$\mathcal{S}(W') = |\{S' \in \mathcal{H} : S' \subseteq (S \cup W'),\ |S' \cap S| \geq k\}|.$$

Observe that if (S, W') is k-bad, then every edge $S' \subseteq (S \cup W')$ has $|S' \cap S| \geq k$ (since $|S' \cap S| \geq |S' \setminus W'|$), so the W' we wish to count satisfy

$$\mathcal{S}(W') = |\{S' \in \mathcal{H} : S' \subseteq (S \cup W')|.$$

If (S, W') is pathological, then this latter set has size at least N. In total, if $\mathbf{W}'$ is chosen uniformly at random from $\binom{V \setminus S}{w}$, then

$$\Pr[(S, \mathbf{W}') \text{ is } k\text{-bad and pathological}] \leq \Pr[\mathcal{S}(\mathbf{W}') \geq N] \leq \frac{\mathbb{E}[\mathcal{S}(\mathbf{W}')]}{N}, \tag{4.6}$$

where this last step used Markov's inequality. It remains to upper bound $\mathbb{E}[\mathcal{S}(\mathbf{W}')]$.

Let

$$m_j(S) = |\{S' \in \mathcal{H} : |S \cap S'| = j\}|,$$

and observe that for any S' with $|S \cap S'| = j$, the number of $W' \in \binom{V \setminus S}{w}$ with $S' \subseteq S \cup W'$ is exactly $\binom{n-2r+j}{w-r+j}$. With this we see that

$$\mathbb{E}[\mathcal{S}(\mathbf{W}')] = \sum_{j \geq k} m_j(S) \frac{\binom{n-2r+j}{w-r+j}}{\binom{n-r}{w}} = \sum_{j \geq k} m_j(S) \frac{\binom{w}{r-j}}{\binom{n-r}{r-j}} = \frac{\binom{w+r}{r}}{\binom{n}{r}} \sum_{j \geq k} m_j(S) \frac{\binom{w}{r-j}}{\binom{n-r}{r-j}} \cdot \frac{\binom{n}{w+r}}{\binom{n-r}{w}}. \tag{4.7}$$

Because $\mathcal{H}$ is $(q; r, k)$-spread, we have for each $j \geq k$ in the sum that

$$m_j(S) \leq M_j(S) \leq q^j |\mathcal{H}|. \tag{4.8}$$

For integers x, y, define the falling factorial $(x)_y := x(x-1)\cdots(x-y+1)$. With this we have

$$\frac{\binom{w}{r-j}}{\binom{n-r}{r-j}} \cdot \frac{\binom{n}{w+r}}{\binom{n-r}{w}} = \frac{(w)_{r-j}}{(n-r)_{r-j}} \cdot \frac{(n)_r}{(w+r)_r} \le \left(\frac{w}{n-r}\right)^{r-j} \cdot \left(\frac{n-r}{w}\right)^{r} = \left(\frac{w}{n-r}\right)^{-j} \le (Cq/2)^{-j}, \tag{4.9}$$

where the first inequality used $w \le pn \le \frac{1}{2}n \le n-r$, and the second inequality used

$$w = pn - t \ge pn - r \ge pn/2 = Cqn/2.$$

Combining (4.7), (4.8), and (4.9) shows that

$$\mathbb{E}[\mathcal{S}(\mathbf{W}')] \le \frac{\binom{w+r}{r}}{\binom{n}{r}}|\mathcal{H}|(C/2)^{-k} \cdot \sum_{j\ge k}(C/2)^{k-j} \le \frac{\binom{w+r}{r}}{\binom{n}{r}}|\mathcal{H}|(C/2)^{-k} \cdot 2,$$

where this last step used $C \ge 4$. Plugging this into (4.6) shows that the number of $W' \in \binom{V\setminus S}{w}$ such that (S, W') is k-bad and pathological is at most

$$2(C/2)^{-k}|\mathcal{H}|\frac{\binom{w+r}{r}}{\binom{n}{r}N} \cdot \binom{n-r}{w} = 2(C/2)^{-k/2} \cdot \binom{n-r}{w}.$$

Combining this with the fact that there were $|\mathcal{H}| \cdot \binom{r}{t}$ ways of choosing S and $S \cap W$ gives the claim. □

In total $|\mathcal{B}_t|$ is at most the sum of the bounds from these two claims. Using this and $w = pn - t$ implies

$$\sum_{t\le r}|\mathcal{B}_t| \le \sum_{t\le r} 3(C/2)^{-k/2}|\mathcal{H}|\binom{r}{t}\binom{n-r}{pn-t}$$
$$= 3(C/2)^{-k/2}|\mathcal{H}|\binom{n}{pn},$$

giving the desired result. □

4.3.2 Auxiliary Lemmas

To prove Theorem 4.1.3, we need to consider two special cases. The first is when $\mathcal{H}$ is r-uniform with r relatively small. In this case the following lemma gives effective bounds.

Lemma 4.3.4 ([FKNP21])**.** *Let $\mathcal{H}$ be a q-spread r-bounded hypergraph on V and $\alpha \in (0, 1)$ such that $\alpha \geq 2rq$. If W is a set of size $\alpha|V|$ chosen uniformly at random from V, then the probability that W does not contain an element of $\mathcal{H}$ is at most*

$$2e^{-\alpha/(2rq)}.$$

The other special case we consider is the following.

Lemma 4.3.5. *Let $\mathcal{H}$ be an r-uniform $(q; r, 1)$-spread hypergraph on V and $\alpha \in (0, 1)$ such that $\alpha \geq 4q$. If W is a set of size $\alpha|V|$ chosen uniformly at random from V, then the probability that W does not contain an edge of $\mathcal{H}$ is at most*

$$4q\alpha^{-1} + 2e^{-\alpha|V|/4}.$$

Proof. Let W' be a random set of V obtained by including each vertex independently and with probability $\alpha/2$. Let $X = |\{S \in \mathcal{H} : S \subseteq W'\}|$ and define $m_j(S)$ to be the number of $S' \in \mathcal{H}$ with $|S \cap S'| = j$. Note that $\mathbb{E}[X] = (\alpha/2)^r |\mathcal{H}|$ and that

$$\begin{aligned}\mathrm{Var}(X) &\leq (\alpha/2)^{2r} \sum_{S\in\mathcal{H}} \sum_{S'\in\mathcal{H},\ S\cap S'\neq\emptyset} (\alpha/2)^{-|S\cap S'|} = (\alpha/2)^{2r} \sum_{S\in\mathcal{H}} \sum_{j=1}^{r} (\alpha/2)^{-j} \cdot m_j(S) \\ &\leq (\alpha/2)^{2r} \sum_{S\in\mathcal{H}} \sum_{j=1}^{r} (\alpha/2)^{-j} \cdot q^j |\mathcal{H}| = (\alpha/2)^{2r} \sum_{j=1}^{r} (\alpha/2q)^{-j} |\mathcal{H}|^2 \\ &= \mathbb{E}[X]^2 (\alpha/2q)^{-1} \sum_{j=1}^{r} (\alpha/2q)^{1-j} \leq 4\mathbb{E}[X]^2 q\alpha^{-1},\end{aligned}$$

where the second inequality used that $\mathcal{H}$ being $(q;r,1)$-spread implies $m_j(S) \leq q^j|\mathcal{H}|$ for all $S \in \mathcal{H}$ and $j \geq 1$, and the last inequality used $\alpha/2q \geq 2$. By Chebyshev's inequality we have

$$\Pr[X = 0] \leq \mathrm{Var}(X)/\mathbb{E}[X]^2 \leq 4q\alpha^{-1}.$$

Lastly, observe that

$$\begin{aligned}\Pr[W \text{ contains an edge of } \mathcal{H}] &\geq \Pr[W' \text{ contains an edge of } \mathcal{H} \big| |W'| \leq \alpha|V|] \\ &\geq \Pr[W' \text{ contains an edge of } \mathcal{H}] - \Pr[W' > \alpha|V|].\end{aligned}$$

By the Chernoff bound (see for example [AS04]) we have $\Pr[|W'| > \alpha|V|] \leq 2e^{-\alpha|V|/4}$. Note that W' contains an edge of $\mathcal{H}$ precisely when $X > 0$, so the result follows from our analysis above. $\square$

We conclude this subsection with a small observation.

Lemma 4.3.6. *If $\mathcal{H}$ is an r_1-uniform $(q; r_1, \ldots, r_\ell)$-spread hypergraph on V, then $r_1 \leq eq|V|$.*

Proof. Let $m = \max_{S \in \mathcal{H}} d(S)$, i.e. this is the maximum multiplicity of any edge in $\mathcal{H}$. Then for any $S \in \mathcal{H}$ with $d(S) = m$, we have

$$m = M_{r_1}(S) \leq q^{r_1} |\mathcal{H}| \leq q^{r_1} \cdot m \binom{|V|}{r_1} \leq m(eq|V|/r_1)^{r_1},$$

proving the result. $\square$

4.3.3 Putting the Pieces Together

We now prove a technical version of Theorem 4.1.3 with more explicit quantitative bounds. Theorem 4.1.3 will follow shortly (but not immediately) after proving this.

Theorem 4.3.7. *Let $\mathcal{H}$ be an r_1-uniform $(q; r_1, \ldots, r_\ell, 1)$-spread hypergraph on V and let $C \geq 8$ be a real number. If W is a set of size $2C\ell q|V|$ chosen uniformly at random from V, then*

$$\Pr[W \text{ contains an edge of } \mathcal{H}] \geq 1 - 6\ell^2 (C/4)^{-r_\ell/2} - 40(C\ell)^{-1}, \tag{4.10}$$

and for any i with $4r_i \leq C\ell$ we have

$$\Pr[W \text{ contains an edge of } \mathcal{H}] \geq 1 - 6\ell^2 (C/4)^{-r_i/2} - 2e^{-C\ell/4r_i}. \tag{4.11}$$

Proof. Define $p := Cq$ and $n := |V|$. We can assume $p \leq \frac{1}{2}$, as otherwise the result is trivial. Let $W_1, \ldots W_{\ell-1}$ be chosen independently and uniformly at random from $\binom{V}{pn}$. Throughout this proof we let $r_{\ell+1} = 1$.

Let $\mathcal{H}_1 = \mathcal{H}$ and let $\phi_1 : \mathcal{H}_1 \to \mathcal{H}$ be the identity map. Inductively assume we have defined $\mathcal{H}_i$ and $\phi_i : \mathcal{H}_i \to \mathcal{H}$ for some $1 \leq i < \ell$. Let $\mathcal{H}'_i \subseteq \mathcal{H}_i$ be all the edges

$S \in \mathcal{H}_i$ such that (S, W_i) is r_{i+1}-good with respect to $\mathcal{H}_i$. Thus for each $S \in \mathcal{H}'_i$, there exists an $S' \in \mathcal{H}_i$ such that $S' \subseteq S \cup W_i$ and $|S' \setminus W_i| \le r_{i+1}$. Choose such an S' for each $S \in \mathcal{H}'_i$ and let A_S be any subset of S of size exactly r_{i+1} that contains $S' \setminus W_i$ (noting that $S' \setminus W_i \subseteq S$ since $S' \subseteq S \cup W_i$). Finally, define $\mathcal{H}_{i+1} = \{A_S : S \in \mathcal{H}'_i\}$ and $\phi_{i+1} : \mathcal{H}_{i+1} \to \mathcal{H}$ by $\phi_{i+1}(A_S) = \phi_i(S)$.

Intuitively, $\phi_i(A)$ is meant to correspond to the "original" edge $S \in \mathcal{H}$ which generated A. More precisely, we have the following.

Claim 4.3.8. *For $i \le \ell$, the maps ϕ_i are injective and $A \subseteq \phi_i(A)$ for all $A \in \mathcal{H}_i$.*

Proof. This claim trivially holds at $i = 1$. Inductively assume the result has been proved through some value i. Observe that in the process for generating $\mathcal{H}_{i+1}$, we have implicitly defined a bijection $\psi : \mathcal{H}'_i \to \mathcal{H}_{i+1}$ through the correspondence $\psi(S) = A_S$.

By construction of ϕ_{i+1}, we have $\phi_{i+1}(A) = \phi_i(\psi^{-1}(A))$, so ϕ_{i+1} is injective since ϕ_i was inductively assumed to be injective and ψ is a bijection. Also be construction we have $A \subseteq \psi^{-1}(A)$, and by the inductive hypothesis we have $\psi^{-1}(A) \subseteq \phi_i(\psi^{-1}(A)) = \phi_{i+1}(A)$. This completes the proof. $\square$

For $i < \ell$, we say that W_i is *successful* if $|\mathcal{H}_{i+1}| \ge (1 - \frac{1}{2\ell})|\mathcal{H}_i|$. Note that $|\mathcal{H}_{i+1}| = |\mathcal{H}'_i|$, so this is equivalent to saying that the number of r_{i+1}-bad pairs (S, W_i) with $S \in \mathcal{H}_i$ is at most $\frac{1}{2\ell}|\mathcal{H}_i|$.

Claim 4.3.9. *For $i \le \ell$, if $W_1, \dots, W_{i-1}$ are successful, then $\mathcal{H}_i$ is $(2q; r_i, \dots, r_\ell, 1)$-spread.*

Proof. For a hypergraph $\mathcal{H}'$, we let $M_j(A;\mathcal{H}')$ denote the number of edges of $\mathcal{H}'$ intersecting A in at least j vertices. By Claim 4.3.8, if $\{A_1,\ldots,A_t\}$ are the set of edges of $\mathcal{H}_i$ which intersect some set A in at least j vertices, then $\{\phi_i(A_1),\ldots,\phi_i(A_t)\}$ is a set of t distinct edges of $\mathcal{H}$ intersecting A in at least j vertices. Thus for all sets A and integers j we have $M_j(A;\mathcal{H}_i)\le M_j(A;\mathcal{H})$.

If A is contained in an edge A' of $\mathcal{H}_i$, then by Claim 4.3.8 A is contained in the edge $\phi_i(A')$ of $\mathcal{H}$. Thus $d_{\mathcal{H}_i}(A)>0$ implies $d_{\mathcal{H}}(A)>0$. By assumption of $\mathcal{H}$ being $(q;r_1,\ldots,r_\ell,1)$-spread, if A is a set with $r_{i'}\ge|A|\ge r_{i'+1}$ for some integer i' such that $d_{\mathcal{H}_i}(A)>0$, and if j is an integer satisfying $j\ge r_{i'+1}$, then our previous observations imply

$$M_j(A;\mathcal{H}_i)\le M_j(A;\mathcal{H})\le q^j|\mathcal{H}|. \tag{4.12}$$

Because each of $W_1,\ldots,W_{i-1}$ were successful, we have

$$|\mathcal{H}_i|\ge\left(1-\frac{1}{2\ell}\right)^i|\mathcal{H}|\ge\left(1-\frac{1}{2\ell}\right)^\ell|\mathcal{H}|\ge\frac{1}{2}|\mathcal{H}|,$$

where in this last step we used that $(1-1/(2x))^x$ is an increasing function for $x\ge1$. Plugging $|\mathcal{H}|\le2|\mathcal{H}_i|$ into (4.12) shows that $\mathcal{H}_i$ is $(2q;r_i,\ldots,r_\ell,1)$-spread as desired. $\square$

Claim 4.3.10. *For $i<\ell$,*

$$\Pr[W_i \textit{ is not successful} \mid W_1,\ldots,W_{i-1} \textit{ are successful}]\le 6\ell(C/4)^{-r_{i+1}/2}.$$

Proof. By construction $\mathcal{H}_i$ is r_i-uniform. Conditional on $W_1,\ldots,W_{i-1}$ being successful, Claim 4.3.9 implies that $\mathcal{H}_i$ is in particular $(2q;r_i,r_{i+1})$-spread. By hypothesis we have

$p \leq \frac{1}{2}$ and $C/2 \geq 4$, and by Lemma 4.3.6 applied to $\mathcal{H}$ we have $2r_i \leq pn$ since $C \geq 2e$. Thus we can apply Lemma 4.3.1 to $\mathcal{H}_i$ (using $C/2$ instead of C), which shows that the expected number of r_{i+1}-bad pairs (S, W_i) is at most $3(C/4)^{-r_{i+1}/2}|\mathcal{H}_i|$. By Markov's inequality, the probability of there being more than $\frac{1}{2\ell}|\mathcal{H}_i|$ total r_{i+1}-bad pairs is at most $6\ell(C/4)^{-r_{i+1}/2}$, giving the result. □

We are now ready to prove the result. Let W and W' be sets of size $2\ell pn$ and ℓpn chosen uniformly at random from V. Observe that for any $1 \leq i \leq \ell$, the probability of W containing an edge of $\mathcal{H}$ is at least the probability of $W_1 \cup \cdots \cup W_{i-1} \cup W'$ containing an edge of $\mathcal{H}$, and this is at least the probability that W' contains an edge of $\mathcal{H}_i$ (since every edge of $\mathcal{H}_i$ is an edge of $\mathcal{H}$ after removing vertices that are in $W_1 \cup \cdots \cup W_{i-1}$), so it suffices to show that this latter probability is large for some i.

By Proposition 4.1.2(a) and Claim 4.3.9, the hypergraph $\mathcal{H}_i$ will be $(2q)$-spread if $W_1, \ldots, W_{i-1}$ are all successful. If i is such that $C\ell \geq 4r_i$, then by Claim 4.3.10 and Lemma 4.3.4 the probability that $W_1, \ldots, W_{i-1}$ are all successful and W' contains an edge of $\mathcal{H}_i$ is at least

$$1 - 6\ell^2(C/4)^{-r_i/2} - 2e^{-C\ell/4r_i},$$

giving (4.11).

Alternatively, the probability that W' contains an edge of $\mathcal{H}_\ell$ can be computed using Lemma 4.3.5, which gives that the probability of success is at least

$$1 - 6\ell^2(C/4)^{-r_\ell/2} - 16(C\ell)^{-1} - 2e^{-C\ell qn/4}.$$

Using $qn \geq e^{-1}r_1 \geq 1/3$ from Lemma 4.3.6 together with $e^{-x} \leq x^{-1}$ gives (4.10) as desired. $\square$

We now use this to prove Theorem 4.1.3.

Proof of Theorem 4.1.3. There exists a large constant K' such that if[1] $r_\ell \geq K' \log(\ell+1)$, then the result follows from (4.10). If this does not hold and if $r_1 > K' \log(\ell+1)$, then there exists some $I \geq 2$ such that $r_{I-1} > K' \log(\ell+1) \geq r_I$. If $r_I = K' \log(\ell+1)$, then the result follows from (4.11) with $i = I$ provided C is sufficiently large in terms of K'. Otherwise we define a new sequence of integers $r'_1, \ldots, r'_{\ell+1}$ with $r'_i = r_i$ for $i < I$, $r'_I = K' \log(\ell+1)$, and $r'_i = r_{i-1}$ for $i > I$. It is not hard to see that $\mathcal{H}$ is $(q; r'_1, \ldots, r'_{\ell+1}, 1)$-spread, so the result follows[2] from (4.11) with $i = I$.

It remains to deal with the case $r_1 \leq K' \log(\ell+1)$. Because $\ell \leq r_1$, this can only hold if $r_1 \leq K''$ for some large constant K''. In this case we can apply Lemma 4.3.4 to give the desired result by choosing K_0 sufficiently large in terms of K''. $\square$

4.4 Concluding Remarks

With a very similar proof one can prove the following non-uniform analog of Theorem 4.1.3.

[1] We consider $\log(\ell+1)$ as opposed to $\log(\ell)$ to guarantee that this is a positive number for all $\ell \geq 1$.

[2] The bound of (4.11) now uses $\ell+1$ instead of ℓ throughout because we are working with the r'_i sequence, but this does not affect the final result.

Theorem 4.4.1. *Let $\mathcal{H}$ be a $(q; r_1, \ldots, r_\ell, 1)$-spread hypergraph on V and define $s = \min_{S\in\mathcal{H}} |S|$. Assume that there exists a K such that $r_1 \leq Kq|V|$, and such that for all i with $r_i > s$ we have $\log r_i \leq K r_{i+1}$. Then there exists a constant K_0 depending only on K such that if $r_\ell \leq \max\{s, K_0 \log(\ell+1)\}$ and $C \geq K_0$, then a set W of size $C\ell q|V|$ chosen uniformly at random from V satisfies*

$$\Pr[W \text{ contains an edge of } \mathcal{H}] \geq 1 - \frac{K_0}{C\ell}.$$

Observe that if $\mathcal{H}$ is r_1-uniform then this reduces to Theorem 4.1.3 with the additional constraint that $r_1 \leq Kq|V|$ for some K. By Lemma 4.3.6, this extra condition is always satisfied for uniform hypergraphs with $K = e$. We note that Theorem 4.4.1 together with Proposition 4.1.2(b) implies Theorem 4.1.1. We briefly describe the details on how to prove this.

Sketch of Proof. We first adjust the statement and proof of Lemma 4.3.1 to allow $\mathcal{H}$ to be r-bounded. To do this, we partition $\mathcal{H}$ into the uniform hypergraphs $\mathcal{H}_{r'} = \{S \in \mathcal{H} : |S| = r'\}$, and word for word the exact same proof[3] as before shows that the number of k-bad pairs using $S \in \mathcal{H}_{r'}$ is at most $3(C/2)^{-k/2}|\mathcal{H}|\binom{n}{pn}$. We then add these bounds over all r' to get the same bound as in Lemma 4.3.1 multiplied by an extra factor of r. With regards to the other lemmas, one no longer needs Lemma 4.3.6 due to the $r_1 \leq Kq|V|$ hypothesis, and Lemmas 4.3.4 and 4.3.5 are fine as is (in particular, Lemma 4.3.5 still requires $\mathcal{H}$ to be uniform).

[3]The $\mathcal{H}_{r'}$ hypergraphs may not be spread, but they still have the property that $m_j(S) \leq q^j|\mathcal{H}|$ for all $S \in \mathcal{H}_{r'} \subseteq \mathcal{H}$, and this is the only point in the proof where we used that $\mathcal{H}$ is spread.

For the main part of the proof, instead of choosing A_S to be a subset of S of size exactly r_i, we choose it to have size at most r_i and at least $\min\{r_i, s\}$. With this $\mathcal{H}_i$ will be uniform if $r_i \le s$, and otherwise when we apply the non-uniform version of Lemma 4.3.1 our error term will have an extra factor of $r_i \le e^{Kr_{i+1}}$, with this inequality holding by our hypothesis for $r_i > s$. This term will be insignificant compared to $(C/2)^{-r_{i+1}/2}$ provided C is large in terms of K.

If $r_\ell \le K' \log(\ell + 1)$ for some large K' depending on K, then as in the proof of Theorem 4.1.3 we can assume $r_I = K' \log(\ell + 1)$ for some I and conclude the result as before. Otherwise $r_\ell \le s$ by hypothesis, so $\mathcal{H}_\ell$ will be uniform and we can apply Lemma 4.3.5 to conclude the result. $\square$

Another extension can be made by not requiring the same "level of spreadness" throughout $\mathcal{H}$.

Definition 2. *Let $0 < q_1, \dots, q_{\ell-1} \le 1$ be real numbers and $r_1 > \cdots > r_\ell$ positive integers. We say that a hypergraph $\mathcal{H}$ on V is $(q_1, \dots, q_{\ell-1}; r_1, \dots, r_\ell)$-spread if $\mathcal{H}$ is non-empty, r_1-bounded, and if for all $1 \le i < \ell$, every $A \subseteq V$ with $d(A) > 0$ and every integer j satisfying $r_i \ge |A| \ge j \ge r_{i+1}$ has*

$$M_j(A) = |\{S \in \mathcal{H} : |A \cap S| \ge j\}| \le q_i^j |\mathcal{H}|.$$

Different levels of spread was also considered in [ALWZ20]. Here one can prove the following.

Theorem 4.4.2. *Let $\mathcal{H}$ be a $(q_1, \ldots, q_\ell; r_1, \ldots, r_\ell, 1)$-spread hypergraph on V and define $s = \min_{S \in \mathcal{H}} |S|$. Assume that there exists a K such that for all i we have $r_i \leq K q_i |V|$, and that for all i with $r_i > s$ we have $\log r_i \leq K r_{i+1}$. Then there exists a constant K_0 depending only on K such that if $r_\ell \leq \max\{s, K_0 \log(\ell+1)\}$ and if $C \geq K_0$, then a set W of size $C \sum q_i |V|$ chosen uniformly at random from V satisfies*

$$\Pr[W \textit{ contains an edge of } \mathcal{H}] \geq 1 - \frac{K_0 \log(\ell+1)}{CL},$$

where $L := \sum_i q_i / \max_i q_i$.

Note that $\sum q_i \leq \ell \max q_i$, so we have $L \leq \ell$ with equality if $q_i = q_j$ for all i, j.

Sketch of Proof. We now choose our random sets W_i to have sizes $Cq_i|V|$ and W' to have size $C \sum q_i |V| = C(L \cdot \max q_i)|V|$. With this any of the $\mathcal{H}_i$ could be at worst $(2 \max q_i)$-spread if each $\mathcal{H}_{i'}$ was successful, so in this case when we apply Lemma 4.3.4 with W' we end up getting a probability of roughly $1 - e^{-CL/r_i}$ of containing an edge. From this quantity we should subtract roughly $\ell^2 C^{-r_i}$, since this is the probability that some $\mathcal{H}_{i'}$ is unsuccessful. If $r_i = K' \log(\ell+1)$ for some large constant K' then this gives the desired bound. Otherwise we can basically assume $r_\ell > K' \log(\ell+1)$ and apply Lemma 4.3.5 to $\mathcal{H}_\ell$ to get a probability of roughly $1 - (CL)^{-1}$, which also gives the result after subtracting $\ell^2 C^{-r_\ell}$ to account for some $\mathcal{H}_{i'}$ being unsuccessful. $\square$

Recently Frieze and Marbach [FM21] developed a variant of Theorem 4.1.1 for rainbow structures in hypergraphs. We suspect that straightforward generalizations of our proofs and those of [FM21] should give an analog of Theorem 4.1.3 (as well as Theorems 4.4.1 and 4.4.2) for the rainbow setting.

This chapter contains material from: S. Spiro, *"A Smoother Notion of Spread Hypergraphs"*, submitted (2021). The dissertation author was the primary investigator and author of this paper.

References

[ALWZ20] Ryan Alweiss, Shachar Lovett, Kewen Wu, and Jiapeng Zhang, *Improved bounds for the sunflower lemma*, Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, 2020, pp. 624–630.

[AS04] Noga Alon and Joel H Spencer, *The probabilistic method*, John Wiley & Sons, 2004.

[BCW21] Tolson Bell, Suchakree Chueluecha, and Lutz Warnke, *Note on sunflowers*, Discrete Mathematics **344** (2021), no. 7, 112367.

[BHJ57] David Blackwell and JL Hodges Jr, *Design for the control of selection bias*, The Annals of Mathematical Statistics (1957), 449–460.

[BL20] József Balogh and Lina Li, *On the number of linear hypergraphs of large girth*, Journal of Graph Theory **93** (2020), no. 1, 113–141.

[BMS15] József Balogh, Robert Morris, and Wojciech Samotij, *Independent sets in hypergraphs*, Journal of the American Mathematical Society **28** (2015), no. 3, 669–709.

[BNS19] József Balogh, Bhargav Narayanan, and Jozef Skokan, *The number of hypergraphs without linear cycles*, Journal of Combinatorial Theory, Series B **134** (2019), 309–321.

[BS11] József Balogh and Wojciech Samotij, *The number of $K_{s,t}$-free graphs*, Journal of the London Mathematical Society **83** (2011), no. 2, 368–388.

[BT87] Béla Bollobás and Arthur G Thomason, *Threshold functions*, Combinatorica **7** (1987), no. 1, 35–38.

[CDGM81] FRK Chung, Persi Diaconis, RL Graham, and CL Mallows, *On the permanents of complements of the direct sum of identity matrices*, Advances in Applied Mathematics **2** (1981), no. 2, 121–137.

[CDH+21] Alexander Clifton, Bishal Deb, Yifeng Huang, Sam Spiro, and Semin Yoo, *Continuously increasing subsequences of random multiset permutations*, arXiv preprint arXiv:2110.10315 (2021).

[CFSZ20] David Conlon, Jacob Fox, Benny Sudakov, and Yufei Zhao, *The regularity method for graphs with few 4-cycles*, arXiv preprint arXiv:2004.10180 (2020).

[CG16] David Conlon and William Timothy Gowers, *Combinatorial theorems in sparse random sets*, Annals of Mathematics (2016), 367–454.

[Ciu98] Mihai Ciucu, *No-feedback card guessing for dovetail shuffles*, Annals of Applied Probability **8** (1998), no. 4, 1251–1269.

[CT17] Jan Corsten and Tuan Tran, *Balanced supersaturation for degenerate hypergraphs*, arXiv preprint arXiv:1707.03788 (2017).

[DG81] Persi Diaconis and Ronald Graham, *The analysis of sequential experiments with feedback to subjects*, The Annals of Statistics **9** (1981), no. 1, 3–23.

[DGH01] Persi Diaconis, Ronald Graham, and Susan P Holmes, *Statistical problems involving permutations with restricted positions*, Lecture Notes-Monograph Series (2001), 195–222.

[DGHS21] Persi Diaconis, Ron Graham, Xiaoyu He, and Sam Spiro, *Card guessing with partial feedback*, Combinatorics, Probability and Computing (2021), 1–20.

[DGS20] Persi Diaconis, Ron Graham, and Sam Spiro, *Guessing about guessing: Practical strategies for card guessing with feedback*, arXiv preprint arXiv:2012.04019 (2020).

[Dia78] Persi Diaconis, *Statistical problems in esp research*, Science **201** (1978), no. 4351, 131–136.

[Efr71] Bradley Efron, *Forcing a sequential experiment to be balanced*, Biometrika **58** (1971), no. 3, 403–417.

[EFR86] Paul Erdős, Peter Frankl, and Vojtech Rödl, *The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent*, Graphs and Combinatorics **2** (1986), no. 1, 113–121.

[EL05] SN Ethier and David A Levin, *On the fundamental theorem of card counting, with application to the game of trente et quarante*, Advances in applied probability **37** (2005), no. 1, 90–107.

[EP21] Alberto Espuny Díaz and Yury Person, *Spanning F-cyles in random graphs*, arXiv preprint arXiv:2106.10023 (2021).

[Erd47] Paul Erdös, *Some remarks on the theory of graphs*, Bulletin of the American Mathematical Society **53** (1947), no. 4, 292–294.

[ES83] Paul Erdős and Miklós Simonovits, *Supersaturated graphs and hypergraphs*, Combinatorica **3** (1983), no. 2, 181–192.

[FKNP21] Keith Frankston, Jeff Kahn, Bhargav Narayanan, and Jinyoung Park, *Thresholds versus fractional expectation-thresholds*, Annals of Mathematics **194** (2021), no. 2, 475–495.

[FKP15] Florent Foucaud, Michael Krivelevich, and Guillem Perarnau, *Large subgraphs without short cycles*, SIAM Journal on Discrete Mathematics **29** (2015), no. 1, 65–78.

[FM21] Alan Frieze and Trent G Marbach, *Rainbow thresholds*, arXiv preprint arXiv:2104.05629 (2021).

[FMS20] Asaf Ferber, Gweneth McKinley, and Wojciech Samotij, *Supersaturated sparse graphs and hypergraphs*, International Mathematics Research Notices **2020** (2020), no. 2, 378–402.

[Für94] Zoltán Füredi, *Random ramsey graphs for the four-cycle*, Discrete Mathematics **126** (1994), no. 1-3, 407–410.

[GL12] Ervin Győri and Nathan Lemons, *Hypergraphs with no cycle of a given length*, Combinatorics, Probability & Computing **21** (2012), no. 1-2, 193.

[GT08] Ben Green and Terence Tao, *The primes contain arbitrarily long arithmetic progressions*, Annals of Mathematics (2008), 481–547.

[HO21] Jimmy He and Andrea Ottolini, *Card guessing and the birthday problem for sampling without replacement*, arXiv preprint arXiv:2108.07355 (2021).

[JKV08] Anders Johansson, Jeff Kahn, and Van Vu, *Factors in random graphs*, Random Structures & Algorithms **33** (2008), no. 1, 1–28.

[JŁR00] Svante Janson, Tomasz Łuczak, and Andrzej Rucinski, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000. MR 1782847

[KKS98] Yoshiharu Kohayakawa, Bernd Kreuter, and Angelika Steger, *An extremal problem for random graphs and the number of graphs with large even-girth*, Combinatorica **18** (1998), no. 1, 101–120.

[KNP21] Jeff Kahn, Bhargav Narayanan, and Jinyoung Park, *The threshold for the square of a hamilton cycle*, Proceedings of the American Mathematical Society **149** (2021), no. 08, 3201–3208.

[KO12] Daniela Kühn and Deryk Osthus, *On pósa's conjecture for random graphs*, SIAM Journal on Discrete Mathematics **26** (2012), no. 3, 1440–1457.

[KQ21] William Kuszmaul and Qi Qi, *The multiplicative version of azuma's inequality, with an application to contention analysis*, arXiv preprint arXiv:2102.05077 (2021).

[Liu21] Pengda Liu, *On card guessing game with one time riffle shuffle and complete feedback*, Discrete Applied Mathematics **288** (2021), 270–278.

[LV03] Felix Lazebnik and Jacques Verstraëte, *On hypergraphs of girth five*, The Electronic Journal of Combinatorics **10** (2003), no. 1, R25.

[Man07] W Mantel, *Problem 28*, Wiskundige Opgaven **10** (1907), no. 60-61.

[MS16] Robert Morris and David Saxton, *The number of $C_{2\ell}$-free graphs*, Advances in Mathematics **298** (2016), 534–580.

[MY20] Dhruv Mubayi and Liana Yepremyan, *Random Turán theorem for hypergraph cycles*, arXiv preprint arXiv:2007.10320 (2020).

[NSV20] Jiaxi Nie, Sam Spiro, and Jacques Verstraëte, *Triangle-free subgraphs of hypergraphs*, arXiv preprint arXiv:2004.10992 (2020).

[Peh10] Lerna Pehlivan, *No feedback card guessing for top to random shuffles*, arXiv preprint arXiv:1006.1321 (2010).

[PR17] Guillem Perarnau and Bruce Reed, *Existence of spanning $\mathcal{F}$-free subgraphs with large minimum degree*, Combinatorics, Probability and Computing **26** (2017), no. 3, 448–467.

[Pro91] Michael Proschan, *A note on blackwell and hodges (1957) and diaconis and graham (1981)*, The Annals of Statistics (1991), 1106–1108.

[Rao19] Anup Rao, *Coding for sunflowers*, arXiv preprint arXiv:1909.04774 (2019).

[Rom15] Dan Romik, *The surprising mathematics of longest increasing subsequences*, no. 4, Cambridge University Press, 2015.

[RS78] Imre Z Ruzsa and Endre Szemerédi, *Triple systems with no six points carrying three triangles*, Combinatorics (Proceedings of the Fifth Hungarian Colloquium Keszthely, 1976), volume II, Colloquium Mathemtical Society, János Bolyai **18** (1978), 939–945.

[Sch16] Mathias Schacht, *Extremal results for random discrete structures*, Annals of Mathematics (2016), 333–365.

[Spi21] Sam Spiro, *Online card games*, arXiv preprint arXiv:2106.11866 (2021).

[ST15] David Saxton and Andrew Thomason, *Hypergraph containers*, Inventiones mathematicae **201** (2015), no. 3, 925–992.

[SU83] Francisco J Samaniego and Jessica M Utts, *Evaluating performance in continuous experiments with feedback to subjects*, Psychometrika **48** (1983), no. 2, 195–209.

[SV20] Sam Spiro and Jacques Verstraëte, *Relative Tur\ 'an numbers for hypergraph cycles*, arXiv preprint arXiv:2012.11061 (2020).

[SV21] ———, *Relative turán problems for uniform hypergraphs*, SIAM Journal on Discrete Mathematics **35** (2021), no. 3, 2170–2191.

[Sze75] Endre Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith **27** (1975), no. 299-345, 5.

[TV15] Craig Timmons and Jacques Verstraëte, *A counterexample to sparse removal*, European Journal of Combinatorics **44** (2015), 77–86.