Title
Independent Sets in Hypergraphs

Author
Nie, Jiaxi

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Independent Sets in Hypergraphs

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

in

Mathematics

by

Jiaxi Nie

Committee in charge:

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

2022

The Dissertation of Jiaxi Nie is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

DEDICATION

To Ruiqin.

EPIGRAPH

We'll continue tomorrow—If I live.

Paul Erdős

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

First of all, I must thank my advisor Jacques Verstraëte. I would not have become the mathematician I am today without his guidance, support, and encouragement. Thank you, Jacques, for sharing with me exciting and important research problems, offering me many suggestions on reading, writing, and presenting, being the Chair of the committee, supporting me to travel, and offering me the GSR positions.

I want to thank Fan Chung, Shachar Lovett, Andrew Suk, and Alexander Vardy for serving on the committee. I would also like to thank my collaborators: Calum Buchanan, Alexander Clifton, Dan Cranston, Eric Culver, Xiaoyu He, Jason O'Neill, Puck Rombach, Sam Spiro, Jacques Verstraëte, Alexandra Wesolek, and Mei Yin. Thank you for your hard work and helpful discussions that lead to fruitful outcomes.

Finally, I want to thank my parents, sisters, and Ruiqin for their unconditional love and support.

Chapter 2, in full, is a version of the material as it appears in "Randomized greedy algorithm for independent sets in regular uniform hypergraphs with large girth", Random Structures & Algorithms 59.1 (2021): 79-95, co-authored with Jacques Verstraëte. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is a version of the material as it appears in "Ramsey Numbers for Nontrivial Berge Cycles", SIAM Journal on Discrete Mathematics 36.1 (2022): 103-113, co-authored with Jacques Verstraëte. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in full, is currently being prepared for submission for publication of the material, co-authored with Jacques Verstraëte. The dissertation author was the primary investigator and author of this paper.

Chapter 5, in full, is a version of the material as it appears in "Triangle-free Subgraphs of Hypergraphs", Graphs and Combinatorics 37.6 (2021): 2555-2570, co-authored with Sam Spiro, Jacques Verstraëte. The dissertation author was the primary investigator and author of this paper.

VITA

2016	B. S. in Mathematics, Nankai University
2022	Ph. D. in Mathematics, University of California San Diego

PUBLICATIONS

Jiaxi Nie, Jacques Verstraëte, Ramsey Numbers for Nontrivial Berge Cycles. SIAM Journal on Discrete Mathematics 36.1 (2022): 103-113.

Jiaxi Nie, Sam Spiro, and Jacques Verstraëte, Triangle-free Subgraphs of Hypergraphs. Graphs and Combinatorics 37.6 (2021): 2555-2570.

Jiaxi Nie, Jacques Verstraëte, Randomized greedy algorithm for independent sets in regular uniform hypergraphs with large girth. Random Structures & Algorithms 59.1 (2021): 79-95.

ABSTRACT OF THE DISSERTATION

Independent Sets in Hypergraphs

by

Jiaxi Nie

Doctor of Philosophy in Mathematics

University of California San Diego, 2022

Professor Jacques Verstraëte, Chair

In this dissertation, we study problems concerning independent sets in hypergraphs. We try to determine the minimum independence number of a hypergraph in a certain family of hypergraphs.

In Chapter 1, we introduce history of this kind of problems.

In Chapter 2, we analyze the outcome of a randomized greedy algorithm for independent sets in regular uniform hypergraphs with large girth. We show that the expected size of the output independent sets can be expressed as the solution of some differential equations and show that it is concentrated around its mean.

In Chapter 3, we establish an upper bound for the Ramsey numbers for nontrivial Berge

cycles of even length and prove that this bound is essentially tight if a conjecture of Erdös and Simonovits is true.

In Chapter 4, we show an upper bound for the Ramsey numbers for 3-uniform loose 4-cycle, improving previous result on this problem.

In Chapter 5, we find a lower bound for the maximum size of a loose triangle-free subgraph in a uniform hypergraph in terms of its maximum degree. Moreover, we determine the magnitude of the maximum size of a loose triangle-free subgraph in the uniform random hypergraph.

Chapter 1

Introduction

A *hypergraph* H is a pair $(V(H), E(H))$, where $V = V(H)$ is a set of *vertices*, and $E = E(H)$ is a set of nonempty subsets of V called *edges*. An *r-uniform hypergraph* (or r-graph for short) is a hypergraph whose edges have size r. When $r = 2$, this gives the usual definition of graph. Let v be a vertex of H. The *degree* of v in H, denoted by $d(v)$, is the number of edges containing v. A hypergraph is *d-regular* if all of its vertices have degree d. An *independent set* of H is a set of vertices which does not contain any edge of H. The size of the largest independent set in H is called the *independence number* of H, usually denoted by $\alpha(H)$.

In general, we ask the following type of questions:

Question 1. *For $r \geq 2$, let $\mathscr{H}$ be a family r-graphs. What is* $\min_{G \in \mathscr{H}} \alpha(G)$*?*

In other words, what is the largest number N such that each r-graph $G \in \mathscr{H}$ contains an independent set of size N? This question is closely related to many Ramsey-type and Turán-type problems: If $\mathscr{H}$ is the set of $\mathscr{F}$-free r-graphs on n vertices for a family of r-graphs $\mathscr{F}$, then it is a Ramsey-type problem; If $\mathscr{H}$ is the set of r-graphs on n vertices with average degree d (possibly with some extra conditions like triangle-free, has large girth, or regular etc.), then it is a Turán-type problem.

1.1 Turán's Theorem

Let $r \geq 2$, and let F be an r-graph, then the *Turán number* $\text{ex}(F,n)$ is the maximum number of edges in an F-free r-graph on n vertices. Let the *Turán graph* $T(t,n)$ be a complete t-partite graph where each part has size $\lfloor n/t \rfloor$ or $\lceil n/t \rceil$. Turán determine the Turán number for cliques and show that the Turán graphs are the only extremal graphs.

Theorem 1.1.1 (Turán [74])**.** *For integers $n \geq t \geq 2$, as $n \to \infty$,*

$$\text{ex}(K_{t+1},n) = e(T(t,n)) \sim (1-\frac{1}{t})\binom{n}{2}.$$

Equality holds only for the Turán graphs.

An equivalent way to phrase the Turán's Theorem is the following:

Theorem 1.1.1 (Turán [74])**.** *Let G be an n-vertex graph with average degree d. Then*

$$\alpha(G) \geq \frac{n}{d+1}. \tag{1.1}$$

Equality holds only when d is an integer and G is a disjoint union of cliques K_{d+1}.

Therefore, Turán 's Theorem in fact answers Question 1 when $\mathscr{H}$ is the set of n-vertex graphs with average degree d. Here is a simple proof given by Caro [14] and Wei [77].

Proof. Taking a random ordering of the vertices of the graph, let I be the set of vertices v such that v is larger than all of its neighbors. It is easy to check that I is an independent set and

$$\mathbb{E}[|I|] \geq \sum_{v \in V} \frac{1}{d(v)+1} \geq \frac{n}{d+1}.$$

Suppose equality holds for G, then the second inequality implies that G is d-regular. Let v be any vertex of G and let $N(v)$ be the set of neighborhoods of v. Let $G' = G \setminus (\{v\} \cup N(v))$. Since the expected size of I is constant, we have $\mathbb{E}[I \cap G' | v \in I] = \frac{n-d-1}{d+1}$. This implies that G' is still d-regular, which means that there is no edge between G' and $\{v\} \cup N(v)$. Therefore, G is a disjoint union of cliques K_{d+1}. □

1.2 Triangle-free graphs

If $\mathscr{H}$ in Question 1 is the set of triangle-free graphs on n vertices with average degree d, then the seminal paper by Ajtai, Komlós and Szemerédi [2] improved Turán's Theorem by a factor of order $\log d$, which was later further improved by Shearer [70].

Theorem 1.2.1 (Shearer [70])**.** *Let G be a triangle-free graph on n-vertices with average degree $d \geq 2$. Then*

$$\alpha(G) \geq \frac{d \log d - d + 1}{(d-1)^2} \cdot n. \tag{1.2}$$

This theorem can be proved by analyzing a randomized greedy algorithm. The algorithm starts with $X = \emptyset$ and $Y = V(G)$. In each round, it randomly picks a vertex v in Y, puts it in X, deletes v and its neighbors from Y, and then repeats until Y is empty. In the end the algorithm outputs an independent set X.

Sketch proof. Let $f(d) = \frac{d \log d - d + 1}{(d-1)^2}$, and let X be the output of the randomized greedy algorithm on G. Since f is convex, it suffices to show that

$$\mathbb{E}[X] \geq \sum_{v \in V(G)} f(d(v)).$$

We will show this by induction on the number of vertices. Let v be the first vertex selected by the greedy algorithm and let $G' = G \setminus (\{v\} \cup N(v))$. Then by inductive assumption we have

$$\mathbb{E}[|X|] = 1 + \mathbb{E}[|X \cap G'|] \geq 1 + \mathbb{E}[\sum_{v \in V(G')} f(d(v))] \geq \sum_{v \in V(G)} f(d(v)).$$

The last inequality is due to the form of f. □

1.3 Graphs with large girth

The *girth* of a graph is the length of a shortest cycle in the graph, graphs of infinity girth contain no cycle. By definition, the graphs with girth at least 4 are triangle-free graphs. For graphs with large girth, Shearer [71] improved (1.2), and Lauer and Wormald [50] showed that there exist a function $\delta = \delta(d,g)$ such that $lim_{g\to\infty}\delta(d,g) = 0$ for fixed d and if G is a d-regular graph of girth g, then

$$\alpha(G) \geq \frac{1}{2}(1-(d-1)^{-\frac{2}{d-2}})n - \delta n.$$

By analyzing the performance of a randomized greedy algorithm, Gamarnik and Goldberg [36] prove the same bound, with an explicit form for δ. The algorithm iteratively selects a vertex uniformly randomly from all remaining vertices of the hypergraph and adds it to the independent set so far, and then deletes all remaining vertices that form an edge with the set of selected vertices, and repeat until no vertices remain. It is convenient to let

$$\varepsilon = \varepsilon(d,g) = \frac{d(d-1)^{\lfloor\frac{g-3}{2}\rfloor}}{(\lfloor\frac{g-1}{2}\rfloor)!}.$$

Note that for each fixed d, $\varepsilon(d,g) \to 0$ as $g \to \infty$.

Theorem 1.3.1 (Gamarnik and Goldberg [36])**.** *Let integers $d \geq 3$ and $g \geq 4$, and let G be a d-regular graph on n vertices with girth g, and let $\mathscr{I}$ be the independent set generated by the greedy algorithm. Then*

$$\Big(\frac{1-(d-1)^{-2/(d-2)}}{2} - \varepsilon\Big)n \leq \mathbb{E}[|\mathscr{I}|] \leq \Big(\frac{1-(d-1)^{-2/(d-2)}}{2} + \varepsilon\Big)n, \tag{1.3}$$

The bounds are effective when $g \gg d$ and, in particular, Theorem 1.3.1 shows

$$\alpha(G) \geq \Big(\frac{1-(d-1)^{-2/(d-2)}}{2} - \varepsilon\Big)n. \tag{1.4}$$

We also observe that when g is sufficiently large relative to d, this bound agrees with (1.2) asymptotically as $d \to \infty$, since

$$1-(d-1)^{-2/(d-2)} = 1-\exp\left(-\frac{2\log(d-1)}{d-2}\right) \sim \frac{2\log d}{d}.$$

On the other hand, the best asymptotic upper bound for this problem is $\frac{2\log d}{d}n$ coming from random regular graphs [12, 55].

1.4 Independent sets in hypergraphs

Turán's Theorem determined the Turán numbers of complete graphs. However, little is known for the Turán numbers of complete hypergraphs. Erdős offered 500 dollars for determining

$$t_{r,k} := \lim_{n\to\infty} \frac{\mathrm{ex}(K_k^r, n)}{\binom{n}{r}}$$

for even one single $k > r > 2$, and he offered 1000 dollars for solving the whole set of problems. For the smallest case when $r = 3$ and $k = 4$, Turán [74] made the following conjecture

Conjecture I (Turán [74])**.**

$$t_{3,4} = \frac{5}{9}. \tag{1.5}$$

For the lower bound, Kostochka [49] gave several different constructions. Chung and Lu [18] proved the current best upper bound.

Theorem 1.4.1 (Chung and Lu [18])**.**

$$t_{3,4} \leq \frac{3+\sqrt{17}}{12} = 0.5936\ldots. \tag{1.6}$$

As in the graph case, this problem is equivalent to determining the minimum independence number of a hypergraph with given order and average degree, that is, answering Question 1 when $\mathscr{H}$ is the set of r-graphs on n vertices with average degree d. Using random sampling and deletion, we can easily obtain the following result.

Proposition 1.4.2. *Let $r \geq 2$, $d \geq 1$ and let H be an $(r+1)$-graph on n vertices of average degree d. Then*

$$\alpha(H) \geq \frac{r}{r+1} n d^{-\frac{1}{r}}. \tag{1.7}$$

Proof. Let X be a subset of $V(H)$ whose elements are chosen independently with probability $p = d^{-1/r}$. We can get an independent set by deleting a vertex for each edge of H contained in X. Then the expected size of such independent set is at least

$$pn - p^{r+1}\frac{nd}{r+1} = \frac{r}{r+1} n d^{-\frac{1}{r}}.$$

□

A better lower bound is given by Caro and Tuza [16], which can be viewed as a hypergraph analog of Turán's Theorem.

Theorem 1.4.3 (Caro and Tuza [16])**.** *Let G be an $(r+1)$-uniform hypergraph on n vertices with average degree d. Then*

$$\alpha(G) \geq \frac{d!}{\prod_{i=1}^{d}(i+\frac{1}{r})} \cdot n. \tag{1.8}$$

The same bound can also be obtained by extending the Caro-Wei proof for independent sets in graphs: taking a random ordering of the vertices of the hypergraph, let I be the set of vertices v such that for every edge e containing v, v is not the smallest vertex in e. Then it can be shown via elementary combinatorial methods that

$$\mathbb{E}[|I|] \geq \sum_{v \in V} \frac{d(v)!}{\prod_{i=1}^{d(v)}(i+\frac{1}{r})} \geq \frac{d!}{\prod_{i=1}^{d}(i+\frac{1}{r})} \cdot n. \tag{1.9}$$

The same algorithm can be implemented via the following random process, which provide a different (and possibly easier) way to analyze the outcome (see for example in Dutta, Mubayi and Subramanian [25]):

1. Equip each vertex with i.i.d. weight from the uniform distribution on [0,1]. Then with probability 1, all vertices will have distinct weights.

2. Select all the vertices that are not the smallest-weighted vertex in any edge that contains it. These vertices form an independent set.

If we select vertices in a more careful way – iteratively select the vertex with largest weight, i.e., select the vertex with largest weight, delete vertices that form an edge with the vertices selected thus far, and repeat – then this random process will be equivalent to the randomized greedy algorithm. In any case, a computation shows

$$\mathbb{E}[|I|] \geq n \int_0^1 (1-x^r)^d dx$$

which gives (1.9). These bounds are asymptotic to $\Gamma(1+\frac{1}{r})nd^{-\frac{1}{r}}$ as $d \to \infty$, where Γ here is the well-known gamma function that extends factorial function to complex numbers, defined as follow for complex number z with positive real part:

$$\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx.$$

1.5 Hypergraphs with small girth

To define *girth* in hypergraph, we first need to define what is a *cycle* in hypergraph. There are many different ways to define cycle in hypergraph–see, e.g., a talk by Sárközy [67]. Here we chose to work with the *Berge-cycle*. For $k \geq 3$, a *Berge k-cycle* is an r-uniform hypergraph with k edges $e_1, e_2, \dots, e_k$ such that there exist distinct vertices $v_1, v_2, \dots, v_k$ such that $\{v_k, v_1\} \in e_1, \{v_1, v_2\} \in e_2, \dots, \{v_{k-1}, v_k\} \in e_k$. When $k = 2$, this corresponds to $v_1, v_2 \in e_1 \cap e_2$. A hypergraph is *linear* if the intersection of any pair of edges has size at most 1. The *girth* of a hypergraph containing a Berge cycle is the smallest g such that the hypergraph contains a Berge g-cycle. In particular, the girth of a non-linear hypergraph is 2. Ajtai, Komlós, Pintz, Spencer and Szemerédi [1] established the following lower bound for $(r+1)$-uniform hypergraphs with girth $g \geq 5$, which improves (1.9) by a factor of order $(\log d)^{\frac{1}{r}}$.

Theorem 1.5.1 (Ajtai, Komlós, Pintz, Spencer and Szemerédi [1])**.** *For integer $r \geq 1$, real number d sufficiently large and integer n sufficiently large, let G be an n-vertex $(r+1)$-uniform hypergraph with average degree d and girth at least 5, then*

$$\alpha(G) \geq 0.36 \cdot 10^{-\frac{5}{r}} \left(\frac{\log d}{rd}\right)^{\frac{1}{r}} n. \tag{1.10}$$

Based on this theorem, Duke, Lefmann and Rödl [24] showed that the same bound(with different constant) holds for linear hypergraphs. Let $\Delta(n)$ be the area of the smallest triangle determined by three out of n points in a unit square, when the points are chosen to maximize this area. Heilbronn made the following conjecture.

Conjecture II (Heilbronn)**.** *As $n \to \infty$,*

$$\Delta(n) = O(\frac{1}{n^2}).$$

As an application of Theorem 1.5.1, Komlós,Pintz and Szemerédi [46] disproved this conjecture.

Theorem 1.5.2 (Komlós,Pintz and Szemerédi [46])**.** *As $n \to \infty$,*

$$\Delta(n) = \Omega(\frac{\log n}{n^2}). \tag{1.11}$$

They also gave the so far best upper bounds for $\Delta(n)$.

Theorem 1.5.3 (Komlós,Pintz and Szemerédi [46])**.** *As $n \to \infty$,*

$$\Delta(n) \leq n^{-\frac{8}{7}+o(1)} \tag{1.12}$$

1.6 Regular hypergraphs with large girth

In Chapter 2, we analyze a randomized greedy algorithm for independent sets in hypergraphs and extend the ideas of Gamarnik and Goldberg [36] to hypergraphs. First, it is convenient to define the following: Let $u(d,r)$ be the only positive real number that satisfies the following equation:

$$\sum_{n \geq 0} \binom{n+d-2}{d-2} \frac{u(d,r)^{rn+1}}{rn+1} = 1. \tag{1.13}$$

Define

$$\varepsilon = \varepsilon(g,d,r) = \frac{d(d-1)^{\lfloor \frac{g-3}{2} \rfloor}}{r \sum_{k=1}^{\lfloor \frac{g-1}{2} \rfloor} (k + \frac{1}{r})}. \tag{1.14}$$

Our main theorem is as follows:

Theorem 1.6.1. *For any integers $r \geq 1$, $d \geq 2$ and $g \geq 4$, let G be an $(r+1)$-uniform d-regular hypergraph with n vertices and girth g, let $\mathscr{I}$ be the independent set of G generated by the greedy algorithm. Let*

$$f(d,r) = u(d,r) - \frac{u(d,r)^{r+1}}{r+1}. \tag{1.15}$$

Then

$$(f(d,r) - \varepsilon)n \leq \mathbb{E}[|\mathscr{I}|] \leq (f(d,r) + \varepsilon)n, \tag{1.16}$$

In particular, due to the form of the quantity $\varepsilon = \varepsilon(g,d,r)$, this theorem is effective for $g \gg d$, and shows

$$\alpha(G) \geq (f(d,r) - \varepsilon)n. \tag{1.17}$$

For $r = 1$, this coincides with Theorem 1.3.1. Moreover, we prove that as $d \to \infty$,

$$f(d,r) \sim \left(\frac{\log d}{rd}\right)^{\frac{1}{r}}, \tag{1.18}$$

and so if g is large enough relative to d, then this slightly improves the constant in (1.10) asymptotically as $d \to \infty$. On the other hand, the random hypergraphs give an upper bound that is tight up to a constant when $n \gg d^g$.

Moreover, we show that the size of the independent set generated by the greedy algorithm concentrates around its mean asymptotically almost surely for linear hypergraphs with bounded degree (i.e. hypergraphs that are not necessarily regular):

Theorem 1.6.2. *For any integers $r \geq 1$ and $d \geq 2$, let G be an $(r+1)$-uniform linear hypergraph with maximum degree d on n vertices, $\mathscr{I}(G)$ be the independent set generated by the greedy algorithm, then for any positive function $b(n)$ with $b(n) \to \infty$ as $n \to \infty$, we have*

$$\mathbb{P}[||\mathscr{I}(G)| - \mathbb{E}[|\mathscr{I}(G)|]| > \sqrt{n}b(n)] \to 0, \text{ as } n \to \infty.$$

1.7 Cycle-complete Ramsey numbers

Let $\mathscr{F}$ be a family of r-graphs and $t \geq 1$. The Ramsey numbers $R(t, \mathscr{F})$ denote the minimum n such that every n-vertex r-graph contains either a hypergraph in $\mathscr{F}$ or an independent set of size t. If we let $\mathscr{H}$ be the set of $\mathscr{F}$-free r-graphs on n vertices, then the question of determining $\min_{G \in \mathscr{H}} \alpha(G)$ is equivalent to evaluating $R(t, \mathscr{F})$.

It is a notoriously difficult problem to determine even the order of magnitude of $R(t, C_k)$ – the cycle-complete graph Ramsey numbers. Kim [43] proved $R(t, C_3) = \Omega(t^2/\log t)$, which gives the order of magnitude of $R(t, C_3)$ when combined with the results of Ajtai, Komlós and Szemerédi [2] and Shearer [70]. The current best lower bounds on $R(t, C_3)$ are due to Fiz Pontiveros, Griffiths and Morris [62] and Bohman and Keevash [11], using the random triangle-free process, which determines $R(t, C_3)$ up to a small constant factor, comparing to the current best upper bounds obtained by Shearer [70].

Theorem 1.7.1 (Shearer [70]; Fiz Pontiveros, Griffiths, and Morris [62]; Bohman and Keevash [11])**.**

$$(\frac{1}{4}-o(1))\frac{t^2}{\log t} \le R(t,C_3) \le (1+o(1))\frac{t^2}{\log t}. \tag{1.19}$$

The case $R(t,C_4)$ is the subject of a notorious conjecture of Erdős [17]

Conjecture III (Erdős [17])**.** *For some $\varepsilon > 0$, as $t \to \infty$,*

$$R(t,C_4) = o(t^{2-\varepsilon}) \tag{1.20}$$

The best bounds for this problem are the following, where the lower bounds are obtained via probabilistic method by Spencer [72], and the upper bounds are proved by Szemerédi in an unpublished form [27].

Theorem 1.7.2 (Spencer [72]; Szemerédi)**.** *As $t \to \infty$,*

$$\Omega\left(\left(\frac{t}{\log t}\right)^{\frac{3}{2}}\right) = R(t,C_4) \le O\left(\left(\frac{t}{\log t}\right)^{2}\right) \tag{1.21}$$

Below we list the current best results for $R(t,C_k)$.

Theorem 1.7.3 (Caro, Li, Rousseau and Zhang [15])**.** *For $k \ge 2$, as $t \to \infty$,*

$$R(t,C_{2k}) = O\left(\left(\frac{t}{\log t}\right)^{k/(k-1)}\right), \tag{1.22}$$

Theorem 1.7.4 (Sudakov [73])**.** *For $k \ge 1$, as $t \to \infty$,*

$$R(t,C_{2k+1}) = O\left(\frac{t^{(k+1)/k}}{\log^{1/k} t}\right), \tag{1.23}$$

Theorem 1.7.5 (Bohman and Keevash [10]). *For $k \geq 4$, as $t \to \infty$,*

$$R(t, C_k) = \Omega\left(\frac{t^{(k-1)/(k-2)}}{\log t}\right). \tag{1.24}$$

Theorem 1.7.6 (Mubayi and Verstraëte [58]). *For odd integer $k \geq 5$, as $t \to \infty$,*

$$R(t, C_k) = \Omega\left(\frac{t^{(k-1)/(k-2)}}{\log^{2/(k-2)} t}\right). \tag{1.25}$$

Theorem 1.7.7 (Mubayi and Verstraëte [58]). *As $t \to \infty$,*

$$R(t, C_5) \geq (1 + o(1)) t^{11/8}, \tag{1.26}$$

$$R(t, C_7) \geq (1 + o(1)) t^{11/9}. \tag{1.27}$$

The lower bounds by Bohman and Keevash [10] are obtained by analyzing the C_k-free process. For odd $k \geq 5$, these lower bounds are slightly improved by Mubayi and Verstraëte [58] using random sampling on C_l-free pseudorandom graphs. Moreover, for $k = 5, 7$, they improved the previous bounds significantly using random block construction.

1.8 Cycle-complete hypergraph Ramsey numbers

For $k, r \geq 3$, an r-uniform *loose k-cycle*, denoted C_k^r, is an r-graph with a cyclic list of edges $e_1, e_2, \ldots, e_k$ such that consecutive edges intersect in exactly one vertex and nonconsecutive sets are disjoint.

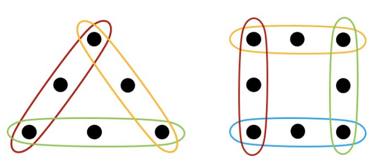

Figure 1.1. 3-uniform Loose cycles of length 3 and 4.

The extremal problem for C_3^3 is closely connected to the extremal problem for three-term arithmetic progressions in sets of integers. Specifically, Ruzsa and Szemerédi [66] made the connection that if Γ is an abelian group and $A \subseteq \Gamma$ has no three term arithmetic progression, then the tripartite linear 3-graph $H(A,\Gamma)$ whose parts are equal to Γ and where $(\gamma, \gamma+a, \gamma+2a)$ is an edge if $a \in A$ – in other words, the edges are three-term progressions whose common difference is in A – is C_3^3-free and has $|A||\Gamma|$ edges. Ruzsa and Szemerédi [66] showed that every n-vertex triangle-free linear triple system has $o(n^2)$ edges, and applying this to $H(A,\Gamma)$ one obtains Roth's Theorem [64] that $|A| = o(|\Gamma|)$. A construction of Behrend [8] gives in $\mathbb{Z}/n\mathbb{Z}$ a set A without three-term progressions of size $n/\exp(O(\sqrt{\log n}))$, and so $H(A,\mathbb{Z}/n\mathbb{Z})$ has $n^{2-o(1)}$ edges in this case.

Ramsey type problems for loose cycles in r-graphs have been studied extensively [13, 19, 21, 34, 38, 39, 47, 48, 56, 58]. For $r \geq 3$ and $k = 3$, Kostochka, Mubayi and Verstraëte [47] showed the following:

Theorem 1.8.1 (Kostochka, Mubayi and Verstraëte [47])**.** *For all $r \geq 3$, as $t \to \infty$,*

$$O\left(\frac{t^{\frac{3}{2}}}{(\log t)^{\frac{3}{4}}}\right) \leq R(t, C_3^r) \leq O\left(t^{\frac{3}{2}}\right). \tag{1.28}$$

They believed the lower bound is closer to the truth and proposed the following conjecture

Conjecture IV (Kostochka, Mubayi and Verstraëte [47])**.** *For all $r \geq 3$, as $t \to \infty$,*

$$R(t, C_3^r) = o\left(t^{\frac{3}{2}}\right). \tag{1.29}$$

They also made the following conjecture for $k \geq 3$:

Conjecture V. *For $r,k \geq 3$,*

$$R(t, C_k^r) = t^{\frac{k}{k-1}+o(1)}. \tag{1.30}$$

The conjecture is true for $k = 3$ due to (1.29). Towards this conjecture, Méroueh [60] showed the following upper bounds, improving earlier results of Collier-Cartaino, Graber and

Jiang [19]:

Theorem 1.8.2 (Méroueh [60])**.** *As $t \to \infty$, for $k \geq 3$,*

$$R(t, C_k^3) = O\left(t^{1+\frac{1}{\lfloor (k+1)/2 \rfloor}}\right), \tag{1.31}$$

and for $r \geq 4$ and odd integer $k \geq 5$,

$$R(t, C_k^r) = O\left(t^{1+\frac{1}{\lfloor k/2 \rfloor}}\right). \tag{1.32}$$

In particular, he showed that $R(t, C_4^3) = O(t^{3/2})$. In Chapter 4, we proved a stronger upper bound for $R(t, C_4^3)$, which matches Conjecture V when $r = 3$ and $k = 4$.

Theorem 1.8.3. *As $t \to \infty$, there exists constant $c > 0$ such that,*

$$R(t, C_4^3) < t^{\frac{4}{3}} \exp\left((1+o(1))\frac{8\sqrt{3}}{9}\sqrt{\log t}\right). \tag{1.33}$$

In Chapter 3, we consider an extension of loose cycle-complete hypergraph Ramsey numbers to Berge cycles in hypergraphs: Recall that for $k \geq 2$, an r-uniform *Berge k-cycle* is a family of r-sets $e_1, e_2, \ldots, e_k$ such that $e_1 \cap e_2, e_2 \cap e_3, \ldots, e_k \cap e_1$ has a system of distinct representatives. We say that a Berge cycle is *nontrivial* if $e_1 \cap e_2 \cap \cdots \cap e_k = \emptyset$. Let $\mathscr{B}_k^r$ denote the family of r-uniform nontrivial Berge k-cycles all of whose sets have size r. In particular, we let $\mathscr{B}_k = \mathscr{B}_k^3$.

In support of Conjecture V, we prove the following result for 3-uniform nontrivial Berge cycles of even length:

Theorem 1.8.4. *For $k \geq 3$, and t large enough,*

$$R(t, \mathscr{B}_{2k}) \leq t^{\frac{2k}{2k-1}} \exp\left(4\sqrt{\log t}\right). \tag{1.34}$$

The lower bound for $R(t,\mathscr{B}_k^r)$ is closely related to the following notoriously difficult conjecture.

Conjecture VI (Erdős and Simonovits [30])**.** *There exist graphs on n vertices of girth more than* $2k$ *with* $\Theta(n^{1+1/k})$ *edges.*

This conjecture remains open, except when $k \in \{2,3,5\}$, largely due to the existence of generalized polygons [9, 53, 75]. Towards this conjecture, Lazebnik, Ustimenko and Woldar [51] gave the densest known construction, which has $\Omega(n^{1+2/(3k-2)})$ edges. We prove the following theorem relating this conjecture to lower bounds on Ramsey numbers for nontrivial Berge cycles:

Theorem 1.8.5. *Let* $k \geq 2$*,* $r \geq 3$*. If Conjecture VI is true, then as* $t \to \infty$*,*

$$R(t,\mathscr{B}_k^r) = \Omega\left(\left(\frac{t}{\log t}\right)^{\frac{k}{k-1}}\right). \tag{1.35}$$

This shows that if the Erdős-Simonovits Conjecture is true, then Theorem 1.8.4 is tight up to a $t^{o(1)}$ factor. Indeed, following the proof of Theorem 1.8.5, the known construction of Lazebnik, Ustimenko and Woldar [51] would give a weaker lower bound of $\Omega((t/\log t)^{(3k-2)/(3k-4)})$.

Let B_k be the family of all 3-uniform Berge k-cycles without nontriviality. Random graphs together with the Lovász local lemma give $R(t,B_k) \geq t^{(2k-2)/(2k-3)-o(1)}$, see [3] for similar computation. We prove the following theorem, which gives a substantially better lower bound for B_4 if the Erdős-Simonovits Conjecture is true.

Theorem 1.8.6. *If Conjecture VI is true for* $k = 4$*, then as* $t \to \infty$*,*

$$R(t,B_4) = \Omega\left(\left(\frac{t}{\sqrt{\log t}}\right)^{16/13}\right). \tag{1.36}$$

We proposed the following 3-uniform analog of the Erdős-Simovits conjecture:

Conjecture VII. *There exist* $\{B_2,B_3,\ldots,B_k\}$*-free 3-graphs on n vertices with* $n^{1+1/\lfloor k/2\rfloor-o(1)}$ *edges.*

This is true for $k = 3$ due to Ruzsa and Szemeredi [33]. The proof of Theorem 1.8.6 makes use of the fact that there exist n-vertex $\{B_2, B_3, B_4\}$-free 3-graphs with $\Omega(n^{3/2})$ edges, that is, Conjecture VII is true for $k = 4$, which is due to Lazebnik and Verstraëte [52]. More generally, following the proof of Theorem 1.8.6, if Conjecture VII is true, then we have $R(t, \{B_2, B_3, \ldots, B_k\}) \geq t^{2k^2/(2k^2-k-2)-o(1)}$ for even integers $k \geq 4$ and $R(t, \{B_2, B_3, \ldots, B_k\}) \geq t^{2k(k-1)/(2k^2-3k-1)-o(1)}$ for odd integers $k \geq 3$, which are substantially better than the lower bounds obtained by random hypergraphs.

Theorem 1.8.6 is valid for all values of $k \geq 2$ and $r \geq 3$, while Theorem 1.8.4 only works for even values of k and $r = 3$. We believe that Theorem 1.8.4 should extend to odd values of k and all $r \geq 3$:

Conjecture VIII. *For all $r, k \geq 3$,*

$$R(t, \mathscr{B}_k^r) \leq t^{\frac{k}{k-1}+o(1)}. \tag{1.37}$$

1.9 Relative Turán numbers

The *Turán numbers* for a graph F are the quantities $\mathrm{ex}(n, F)$ denoting the maximum number of edges in an F-free n-vertex graph. The study of Turán numbers is a cornerstone of extremal graph theory, going back to Mantel's Theorem [54] and Turán's Theorem [74]. A more general problem involves studying the *relative Turán numbers* $\mathrm{ex}(G, F)$, which is the maximum number of edges in an F-free subgraph of a graph G. The problems of determining $\mathrm{ex}(G, F)$ can also be viewed as determining independence number of hypergraphs– consider a hypergraph whose vertices are edges of G and whose edges are copies of F. The study of $\mathrm{ex}(G, F)$ has attracted considerable attention in the literature. Foucaud, Krivelevich, and Perarnau [31] and Perarnau and Reed [61] studied these quantities as a function of the maximum degree of G.

In the case that F is a triangle, $\mathrm{ex}(G, F) \geq \frac{1}{2}e(G)$ for every graph G, which can be seen by taking a maximum cut of G, which is essentially tight. In the case $G = G_{n,p}$, the *Erdős-*

Rényi random graph, $\mathrm{ex}(G,F) \sim \frac{1}{2}p\binom{n}{2}$ with high probability provided p is not too small, and furthermore every maximum triangle-free subgraph is bipartite – see DeMarco and Kahn [23] and also Kohayakawa, Łuczak and Rödl [45] and di Marco, Hamm and Kahn [22] for related stability results.

More generally, if F has chromatic number $k \geq 3$, then by taking a maximum $(k-1)$-partite subgraph we find, for all H,

$$\mathrm{ex}(H,F) \geq (1 - \frac{1}{k-1})e(H),$$

which is best possible by the Erdős-Stone Theorem, which shows $\mathrm{ex}(K_n,F) \sim (1 - \frac{1}{k-1})e(K_n)$.

The study of F-free subgraphs of random graphs when F has chromatic number at least three is undertaken in seminal papers of Friedgut, Rödl and Schacht [33], Conlon and Gowers [20], and Schacht [69]. Let $m_2(G) = \max\{\frac{e(F)-1}{v(F)-2} : F \subset G,\ v(F) \geq 3\}$

Theorem 1.9.1 (Conlon and Gowers [20]; Schacht [69])**.** *Let G be a graph with at least two edges and suppose that $p \gg n^{-1/m_2(H)}$. Then*

$$\mathrm{ex}(G(n,p),H) = (1 - \frac{1}{\chi(H)-1} + o(1))p\binom{n}{2}$$

asymptotic almost surely (a.a.s.), that is, with probability tending to 1 as $n \to \infty$.

1.10 Triangle-free subgraphs of hypergraphs

In Chapter 5, we consider a generalization of the problem of determining $\mathrm{ex}(G,F)$ when F is a triangle to uniform hypergraphs. If G and F are r-graphs, then $\mathrm{ex}(G,F)$ denotes the maximum number of edges in an F-free subgraph of G. A *loose triangle* is a hypergraph T consisting of three edges e, f and g such that $|e \cap f| = |f \cap g| = |g \cap e| = 1$ and $e \cap f \cap g = \emptyset$. We write T^r for the loose r-uniform triangle. The Turán problem for loose triangles in r-graphs was essentially solved by Frankl and Füredi [32], who showed for each $r \geq 3$ that $\mathrm{ex}(n,T^r) = \binom{n-1}{r-1}$

for n is large enough, with equality only for the r-graph S_n^r of all r-sets containing a fixed vertex. We remark that the Turán problem for r-graphs is notoriously difficult in general, and the asymptotic behavior of $\mathrm{ex}(n, K_t^r)$ is a well-known open problem of Erdős [26] – the celebrated Turán conjecture states $\mathrm{ex}(n, K_4^3) \sim \frac{5}{9}\binom{n}{3}$.

We make use of the following theorem:

Theorem 1.10.1 (Ruzsa and Szemerédi [66]; Erdős, Frankl, and Rödl [28])**.** *For all n there exists an n-vertex r-graph which is linear, loose triangle-free, and which has $n^2 e^{-c\sqrt{\log n}}$ edges for some positive constant c.*

This theorem is an important ingredient for our first theorem in Chapter 5, giving a general lower bound on the number of edges in a densest triangle-free subgraphs of r-graphs:

Theorem 1.10.2. *Let $r \geq 3$ and let G be an r-graph with maximum degree $\triangle$. Then as $\triangle \to \infty$,*

$$\mathrm{ex}(G, T^r) \geq \triangle^{-\frac{r-2}{r-1} - o(1)} e(G).$$

If a positive integer t is chosen so that $\binom{t-1}{r-1} < \triangle \leq \binom{t}{r-1}$ and $t|n$, then the n-vertex r-graph G consisting of n/t disjoint copies of a clique K_t^r has maximum degree at most $\triangle$ whereas

$$\mathrm{ex}(G, T^r) = \binom{t-1}{r-1}\frac{n}{t} = \frac{r}{t} e(G) = O(\triangle^{-\frac{1}{r-1}}) \cdot e(G).$$

Here we used the result of Frankl and Füredi [32] that S_t^r is the extremal T^r-free subgraph of K_t^r for t large enough. Therefore for $r = 3$, Theorem 1.10.2 is sharp up to the $o(1)$ term in the exponent of $\triangle$. For $r \geq 4$, the best construction we have gives the following proposition:

Proposition 1.10.3. *For $r \geq 4$ there exists an r-graph G with maximum degree $\triangle$ such that as $\triangle \to \infty$,*

$$\mathrm{ex}(G, T^r) = O(\triangle^{-\frac{1}{2}}) \cdot e(G).$$

We leave it as an open problem to determine the smallest c such that $\mathrm{ex}(G, T^r) \geq$

$\triangle^{-c-o(1)} \cdot e(G)$ for every r-graph G of maximum degree $\triangle$. We conjecture the following for $r = 3$:

Conjecture IX. *For $\triangle \geq 1$, there exists a triple system G with maximum degree $\triangle$ such that as $\triangle \to \infty$, every T^3-free subgraph of G has $o(\triangle^{-1/2}) \cdot e(G)$ edges.*

1.11 Triangle-free subgraphs of random hypergraphs

Our next set of results concern random hosts. To this end, we say that a statement depending on n holds *asymptotically almost surely* (abbreviated a.a.s.) if the probability that it holds tends to 1 as n tends to infinity. Let $G^r_{n,p}$ denote random r-graph where edges of K^r_n are sampled independently with probability p. For the $r = 2$ case we simply write $G_{n,p}$.

A central conjecture of Kohayakawa, Łuczak and Rödl [45] was resolved independently by Conlon and Gowers [20] and by Schacht [69], and determines the asymptotic value of $\mathrm{ex}(G_{n,p}, F)$ whenever F has chromatic number at least three. The situation when F is bipartite is more complicated, partly due to the fact that the order of magnitude of Turán numbers $\mathrm{ex}(n,F)$ is not known in general – see Füredi and Simonovits [35] for a survey of bipartite Turán problems. The case of even cycles was studied by Kohayakawa, Kreuter and Steger [44] and Morris and Saxton [57] and complete bipartite graphs were studied by Morris and Saxton [57] and by Balogh and Samotij [7].

If F consists of two disjoint r-sets, then $\mathrm{ex}(n,F)$ is given by the celebrated Erdős-Ko-Rado Theorem [29], and $\mathrm{ex}(n,F) = \binom{n-1}{r-1}$. A number of researchers studied $\mathrm{ex}(G^r_{n,p}, F)$ in this case [4], with the main question being the smallest value of p such that an extremal F-free subgraph of $G^r_{n,p}$ consists of all r-sets on a vertex of maximum degree – $(1+o(1))p\binom{n-1}{r-1}$ edges. The same subgraphs are also T^r-free, however the extremal subgraphs in that case are denser and appear to be more difficult to describe. The second main result in Chapter 5 is as follows:

Theorem 1.11.1. *For all $n \geq 2$ and $p = p(n) \leq 1$ with $pn^3 \to \infty$ as $n \to \infty$, there exists a constant*

$c > 0$ *such that asymptotically almost surely*

$$\min\{(1-o(1))p\binom{n}{3}, p^{\frac{1}{3}}n^2e^{-c\sqrt{\log n}}\} \leq \mathrm{ex}(G^3_{n,p}, T^3) \leq \min\{(1+o(1))p\binom{n}{3}, p^{\frac{1}{3}}n^{2+o(1)}\},$$

and more precisely, for any constant $\delta > 0$, *when* $n^{-3/2+\delta} \leq p \leq n^{-\delta}$, *we have*

$$\mathrm{ex}(G^3_{n,p}, T^3) \leq p^{\frac{1}{3}}n^2(\log n)^c.$$

We believe that perhaps the lower bound is closer to the truth.

Since $G^3_{n,p}$ for $p > n^{-2+o(1)}$ has maximum degree $\triangle \sim p\binom{n-1}{2}$ asymptotically almost surely, Theorem 1.10.2 only gives $\mathrm{ex}(G^3_{n,p}, T^3) \geq p^{1/2-o(1)}n^2$ a.a.s. The upper bound in Theorem 1.11.1 employs the method of *containers* developed by Balogh, Morris and Samotij [5] and Saxton and Thomason [68].

We do not have tight bounds for $\mathrm{ex}(G^r_{n,p}, T^r)$ in general for all p and $r \geq 4$. Partial results and conjectures are discussed.

Balogh, Narayanan and Skokan [6] showed that the number of triangle-free n-vertex r-graphs is $2^{\Theta(n^{r-1})}$ using the method of containers. Note that a lower bound follows easily by counting all subgraphs of the r-graph S^r_n on n vertices consisting of all r-sets containing a fixed vertex. In Chapter 5, we adapt the method to count triangle-free hypergraphs with a specified number of edges. We let $N(r,m)$ denote the number of T^r-free r-graphs with n vertices and m edges. Analogs of Theorems 1.11.2 and 1.11.3 for graphs were proven by Balogh and Samotij [7].

Theorem 1.11.2. *Let* $n \geq 2$, $\varepsilon(n)$ *be a function such that* $\frac{\varepsilon(n)\log n}{\log\log n} \to \infty$ *as* $n \to \infty$. *Let* $\delta = \delta(n)$ *be a function such that* $\varepsilon(n) < \delta < 1/2 - \varepsilon(n)$ *and let* $m = n^{2-\delta}$. *Then*

$$N(3,m) \leq \left(\frac{n^2}{m}\right)^{3m+o(m)}.$$

The upper bound on $\mathrm{ex}(G^3_{n,p}, T^3)$ in Theorem 1.11.1 will follow from the bound on $N(3,m)$ in Theorem 1.11.2 by taking $m = p^{1/3-o(1)}n^2$, see details in Section 5.3.

Moreover, we are able to generalize Theorem 1.11.2 to r-graphs as follows:

Theorem 1.11.3. *Let $r \geq 4$, $n \geq 1$, $0 < \delta < 3/2$ and $m = n^{3-\delta}$. Then*

$$N(r,m) \leq \left(\frac{n^{r-1}}{m}\right)^{\left(1+\frac{2\delta}{3r-12+3\delta}\right)m+o(m)}.$$

When $r > 4$, let $m = n^{3+\delta}$ with δ some constant satisfying $0 < \delta < r-4$. Then we have

$$N(r,m) \leq \left(\frac{n^{r-1}}{m}\right)^{m+o(m)}.$$

This bound will also leads to an upper bound for $\mathrm{ex}(G^r_{n,p}, T^r)$ when $n^{-r+3/2+o(1)} \leq p \leq 1$, which is essentially tight for $p = p(n)$ with $n^{-r+4+o(1)} \leq p \leq 1$. However, there is a gap between the lower bound and upper bound in the range $n^{-r+3/2+o(1)} \leq p \leq n^{-r+4+o(1)}$.

Using the same techniques for the $r = 3$ case, we are able to show the following:

Theorem 1.11.4. *For $r \geq 4$ and $0 \leq x \leq r$ a constant, let $p = n^{-r+x}$ and define*

$$f_r(x) = \lim_{n\to\infty} \log_n \mathbb{E}[\mathrm{ex}(G^r_{n,p}, T^r)].$$

Then for $0 \leq x \leq 3/2$, $f_r(x) = x$; for $4 < x \leq r$, $f_r(x) = x-1$; and for $3/2 < x \leq 4$, we have

$$\max\{\frac{x+3r-6}{2r-3}, x-1\} \leq f_r(x) \leq \frac{3x+3}{5}.$$

We believe that the upper bound is perhaps closer to the truth and have the following conjecture.

Conjecture X. *For $r \geq 4$ and $0 \leq x \leq r$ a constant, let $p = n^{-r+x}$ and $f_r(x)$ as defined in Theorem 1.11.4. Then for $\frac{3}{2} < x \leq 4$,*

$$f_r(x) = \frac{3x+3}{5}$$

1.12 Notation and Terminology

For a hypergraph H, let $V(H)$ denote the vertex set of H, $v(H) = |V(H)|$, and let $E(H)$ denote the edge set of H, $|H| = |E(H)|$. If all edges of H have size r, we say H is an *r-uniform hypergraph*, or an *r-graph* for short. For $v \in V(H)$, let $d_H(v) = |\{e \in H : v \in e\}|$ be the *degree of v in H*. The average degree of H is denoted by $d(H)$ and the maximum degree of H is denoted by $\Delta(H)$. For $u, v \in V(H)$, let $d_H(u,v) = |\{w : \{u,v,w\} \in E(H)\}|$ denote the *codegree of the pair $\{u,v\}$ in H*, and $\Delta_2(H) = \max\{d_H(u,v) : \{u,v\} \subset V(H)\}$. A hypergraph H is *linear* if $\Delta_2(H) \leq 1$. Let ∂H denote the shadow of H, namely, the set of pairs of vertices contained in at least one edge of H. For positive functions $f(n)$ and $g(n)$, $g(n) = O(f(n))$ if there exists a constant $c > 0$ such that $g(n) < cf(n)$ for n large enough; and $g(n) = \Omega(f(n))$ if $f(n) = O(g(n))$. Moreover, $g(n) = o(f(n))$ if for any constant $c > 0$, $g(n) < cf(n)$ for n large enough; and $g(n) = \omega(f(n))$ if $f(n) = o(g(n))$.

The rest of this dissertation is structured as following: In Chapter 2, we prove Theorem 1.6.1 and Theorem 1.6.2. In Chapter 3, Theorem 1.8.4, Theorem 1.8.5, and Theorem 1.8.6 are established. In Chapter 4, we give the proof of Theorem 1.8.3. Finally, in Chapter 5, we prove Theorem 1.10.2, Theorem 1.11.1, and Theorem 1.11.2.

Chapter 2, Chapter 3, Chapter 4, and Chapter 5 are written to be self contained. As a result, there are repetitions of some basic lemmas across different Chapters.

Chapter 2

Randomized Greedy Algorithm in Hypergraphs with Large Girth

In this chapter, we extend the ideas of Gamarnik and Goldberg [36] to hypergraphs. First, we restate our main theorems. Recall that $u(d,r)$ is the only positive real number that satisfies the following equation:

$$\sum_{n\geq 0}\binom{n+d-2}{d-2}\frac{u(d,r)^{rn+1}}{rn+1}=1. \tag{2.1}$$

Define

$$\varepsilon=\varepsilon(g,d,r)=\frac{d(d-1)^{\lfloor\frac{g-3}{2}\rfloor}}{r\sum_{k=1}^{\lfloor\frac{g-1}{2}\rfloor}(k+\frac{1}{r})}. \tag{2.2}$$

Our main theorem is as follows:

Theorem 2.0.1 (Theorem 1.6.1). *For any integers $r\geq 1$, $d\geq 2$ and $g\geq 4$, let G be an $(r+1)$-uniform d-regular hypergraph with n vertices and girth g, let $\mathscr{I}$ be the independent set of G generated by the greedy algorithm. Let*

$$f(d,r)=u(d,r)-\frac{u(d,r)^{r+1}}{r+1}. \tag{2.3}$$

Then

$$(f(d,r)-\varepsilon)n\leq\mathbb{E}[|\mathscr{I}|]\leq(f(d,r)+\varepsilon)n, \tag{2.4}$$

We prove in Appendix that as $d \to \infty$,

$$f(d,r) \sim \left(\frac{\log d}{rd}\right)^{\frac{1}{r}}. \tag{2.5}$$

Moreover, we show that the size of the independent set generated by the greedy algorithm concentrate around its mean asymptotically almost surely for linear hypergraphs with bounded degree (i.e. hypergraphs that are not necessarily regular):

Theorem 2.0.2 (Theorem 1.6.1)**.** *For any integers $r \geq 1$ and $d \geq 2$, let G be an $(r+1)$-uniform linear hypergraph with maximum degree d on n vertices, $\mathscr{I}(G)$ be the independent set generated by the greedy algorithm, then for any positive function $b(n)$ with $b(n) \to \infty$ as $n \to \infty$, we have*

$$\mathbb{P}[||\mathscr{I}(G)| - \mathbb{E}[|\mathscr{I}(G)|]| > \sqrt{n}b(n)] \to 0, \text{ as } n \to \infty.$$

2.1 Preliminaries

Gamarnik and Goldberg [36] introduce two notions for graphs, the *influence-blocking subgraph* and *bonus function*. In this section, we generalize these notions to hypergraphs and discuss their properties. A *hypertree* is a linear hypergraph with no Berge cycle, and a *rooted hypertree* is a hypertree in which a special vertex called the *root* is singled out. In summary, we show that the performance of the greedy algorithm on hypergraphs with large girth is locally similar to its performance on a rooted hypertree – note that if a hypergraph has high girth, then for each vertex, its neighbourhood within finite distance looks like a hypertree. Hence, if we can show that the event of a vertex being selected into the independent set is mostly dependent on its neighbourhood within finite distance, then we can simplify the analysis of each vertex into the analysis of the root of a rooted tree. Then we analyze the probability of the root of a rooted hypertree being selected by the randomized greedy algorithm. For ease of analysis, we consider an equivalent way to do the randomized greedy algorithm as follows:

1. Equip each vertex with i.i.d. weight from the uniform distribution on $[0,1]$. Then with probability 1, all vertices will have distinct weights.

2. Iteratively select the vertex with largest weight from all remaining vertices of G, and add it to the independent set so far, and then delete all remaining vertices that form an edge with the selected vertices, and repeat until no vertices remain.

The strategy is to analyze the probability of each vertex being selected into the independent set.

2.1.1 Influence-blocking hypergraphs

Garmarnik and Goldberg [36] introduce *influence-blocking subgraphs*; here we extend this notion to hypergraphs. Suppose we already applied the first step of the greedy algorithm on G. That is, the vertices of G are now equipped with distinct weights. Let v be a vertex of G, e be an edge of G such that e contains v. We say *v defeats e* if there is another vertex v' in e such that the weight of v' is smaller than the weight of v. That is, v is not the smallest weighted vertex in e. Observe that if v defeats all the edges that contains it, then v must be selected into $\mathscr{I}(G)$, since it cannot be deleted according to the rule of the algorithm. In this case, the weight of any other vertex that is not in the neighbourhood of v will not influence the behaviour of v. This phenomenon can be generalized to sub-hypergraphs, which gives us the following definition:

Definition 1. *Let G be a hypergraph whose vertices are equipped with distinct weights. An induced sub-hypergraph H of G is called an* influence-blocking hypergraph *if for every vertex $v \in V(H)$, and $e \in E(G)\backslash E(H)$ with $v \in e$, v is not the vertex in e with smallest weight.*

If G is a hypergraph whose vertices are already equipped with distinct weights, then we also let $\mathscr{I}(G)$ denote the independent set of G generated by applying the second step of the greedy algorithm to G. Let v be a vertex of G, such that $v \notin \mathscr{I}(G)$. If e is an edge of G, such that $v \in e$ and $e \subset v \cup \mathscr{I}(G)$, then we say *$v$ is deleted by e*. The first property of influence-blocking hypergraphs is that the performance of the greedy algorithm inside this sub-hypergraph is not dependent on the performance of the algorithm outside this sub-hypergraph. This phenomenon is described by the following lemma, which is a straightforward modification of Lemma 5 in [36]:

Lemma 2.1.1. *Let G be a hypergraph whose vertices are equipped with distinct weights. Let H be an influence-blocking hypergraph of G. Then $\mathscr{I}(H) = \mathscr{I}(G) \cap V(H)$.*

Proof. Let $V(H) = \{v_1, v_2, \ldots, v_m\}$, such that $v_1 > v_2 > \cdots > v_m$ (where $v_i > v_j$ means the weight of v_i is larger than the weight of v_j). To prove the lemma, it suffices to show that $v_i \in \mathscr{I}(H)$ if and only if $v_i \in \mathscr{I}(G)$, for all i such that $1 \leq i \leq m$. We do that by induction. First, for $i = 1$, we have $v_1 \in \mathscr{I}(H)$. By the definition of influence-blocking hypergraph, v_1 cannot be deleted by edges not in H. Since v_1 has the largest weight among all vertices of H, so it cannot be deleted by edges in H either. Hence, we also have $v_1 \in \mathscr{I}(G)$. This completes the base case. Now suppose $1 < i \leq m$, and the argument holds for all integer less than i. If $v_i \notin \mathscr{I}(H)$, then v_i must be deleted by an edge $e \in E(H)$ such that $e \backslash v_i$ consists of vertices whose weights are larger than the weight of v_i. Then by the inductive assumption, v_i must be deleted by the same edge in the algorithm for G. Hence, we have $v_i \notin \mathscr{I}(G)$. If $v_i \in \mathscr{I}(H)$, then v_i cannot form an edge in H with vertices whose weights are larger than the weight of v_i. Hence, by inductive assumption, v_i cannot be deleted by edges in H. Also, by the definition of influence-blocking hypergraph, v_i cannot be deleted by edges not in H either. Therefore, we have $v_i \in \mathscr{I}(G)$. This completes the inductive step, and hence the proof of the lemma. □

The second property of the influence-blocking hypergraphs is that any subset of vertices can be extended to a unique minimal influence-blocking hypergraph, which is presented by the following lemma, which is a straightforward modification of Lemma 3 in [36]:

Lemma 2.1.2. *Let G be a hypergraph whose vertices are equipped with distinct weights. Let A be such that $A \subset V(G)$, then there exist a unique minimal influence-blocking hypergraph $\mathscr{B}_G(A)$ of G such that $A \subset V(\mathscr{B}_G(A))$. It can be simplified as $\mathscr{B}(A)$ if there is no ambiguity.*

Proof. Pick a set of vertices V_A as following: First, put all vertices of A into V_A. Then, we iteratively take edges that are not in A but whose smallest-weighted vertex is in A, and put all the vertices of such edges into V_A, and then repeat until no edge like this remains. Let $\mathscr{B}(A)$ be the sub-hypergraph of G induced by V_A. By definition, $\mathscr{B}(A)$ is an influence-blocking hypergraph of G, and is contained in any influence-blocking hypergraph of G that contains A. Hence, it is minimal. Also, by the process that it is generated, we can see that it is unique. □

Definition 2. *For any integers* $r, l \geq 1$*, an* $(r+1)$-uniform path of length l *connecting* v_0 *to* v_{lr} *is a hypergraph with vertices* $\{v_0, v_1, \ldots, v_{lr}\}$ *and edges* $e_k = \{v_{kr}, v_{kr+1}, \ldots, v_{(k+1)r}\}$ *for* $0 \leq k \leq l-1$*. If the vertices of a path are weighted and the smallest-weighted vertex in edge* e_k *is* v_{kr} *for all* $0 \leq k \leq l-1$*, then we say the weighted path is* increasing *from* v_0 *to* v_{lr}*.*

Note that the definition of path here is different from the definition of a *Berge path*, which is defined in a similar way as the Berge cycle.

The following lemma evaluate the probability that a path in a hypergraph is increasing when given a random total order:

Lemma 2.1.3. *For any integers* $r, l \geq 1$*, the number of ways to assign* $\{0, 1, \ldots, lr\}$ *as distinct weights to the vertices of an* $(r+1)$*-uniform paths of length* l *from* v_0 *to* v_{lr} *so that it is increasing is*

$$\frac{(lr+1)!}{\prod_{k=1}^{l}(kr+1)}. \tag{2.6}$$

Hence, for an $(r+1)$*-uniform path* P *of length* l*, if each vertex is equipped with i.i.d. weight from the uniform distribution on* $[0,1]$*, then*

$$\mathbb{P}[P \text{ is increasing from } v_0 \text{ to } v_{lr}] = \frac{1}{\prod_{k=1}^{l}(kr+1)}. \tag{2.7}$$

Proof. For simplicity, we only prove this for $r = 2$. In this case, we want to show that the number of proper weight assignments for paths of length l is $(2l)!! = \prod_{k=1}^{l}(2k)$. The idea of the proof for general case is exactly the same. Let a_l be the number of proper weight assignments for 3-uniform paths of length l with distinct weights from $\{0, 1, \ldots, 2l\}$. Let W_i be the weight of v_i. We prove $a_l = (2l)!!$ by induction. First, for $l = 1$, W_0 has to be 0, W_2 can be either 1 or 2. So

$a_1 = 2 = 2!!$. Now for $l \geq 2$, suppose the lemma is true for $l-1$. Then again, W_0 has to be 0. W_2 is less than all W_i with $i > 2$, so W_2 is at least the third smallest weight. As a result, $W_2 = 1$ or 2. When $W_2 = 1$, W_1 can be any number in $\{2,3,\ldots,2l\}$, and all the other vertices form a 3-uniform increasing path of length $l-1$. So the number of proper weight assignments of this kind is $(2l-1)a_{l-1}$. When $W_2 = 2$, W_1 has to be 1, and all the other vertices form a 3-uniform increasing path of length $l-1$, the number of proper weight assignments of this kind is a_{l-1}. Hence, by inductive assumption, we have $a_l = 2la_{l-1} = 2l \cdot (2l-2)!! = (2l)!!$. This completes the proof for $r = 2$. □

For any vertex v and any integer $h \geq 1$, let $N_h(v)$ be the set of vertices w such that there exist a path, as defined in Definition 2, connecting v to w, whose length is less or equal than h. When $h = 0$, let $N_0(v) = v$. The following lemma, which is a modification of Lemma 6 in [36], show that for any vertex v, the probability that the minimal influence-blocking hypergraph containing v is not a sub-hypergraph of $N_h(v)$ converges to 0 as $h \to \infty$.

Lemma 2.1.4. *For any integers $r \geq 1$ and $d \geq 2$, let G be any $(r+1)$-uniform linear hypergraph of maximum degree d, and suppose that the vertices are equipped with i.i.d. weights from the uniform distribution on $[0,1]$. Then for any vertex v and any $h \geq 0$,*

$$\mathbb{P}[\mathscr{B}(v) \not\subset N_h(v)] \leq \frac{d(d-1)^h}{r\prod_{k=1}^{h+1}(k+\frac{1}{r})}. \tag{2.8}$$

Proof. For any vertex v there exist at most $d(d-1)^h r^h$ distinct paths of length $h+1$ that connecting v to some vertex in $N_{h+1}(i) \backslash N_h(i)$. By definition, $\mathscr{B}(v) \not\subset N_h(v)$ if and only if at least one of these path is increasing. So by applying a union bound and equation (2.7), we have

$$\mathbb{P}[\mathscr{B}(v) \not\subset N_h(v)] \leq \frac{d(d-1)^h r^h}{\prod_{k=1}^{h+1}(kr+1)} = \frac{d(d-1)^h}{r\prod_{k=1}^{h+1}(k+\frac{1}{r})}.$$

□

2.1.2 Bonus function of hypergraphs

To analyze the probability of the root of a rooted hypertree being selected into the independent set, we use the following notion to establish a recursive equation, and hence by some analysis, a differential equation.

Consider the following *bonus function of hypergraphs*, which is extended from the bonus function of graphs introduced by Garmarnik and Goldberg [36]:

Definition 3. *Let T be a rooted hypertree, whose vertices are equipped with distinct positive weights. Let W_v be the weight of a vertex v, $DE(v)$ be the set of descending edges of v and I be the indicator function. Then the* bonus function of hypergraphs $S_T : V(T) \to \mathbb{R}$ *is defined by*

$$S_T(v) = \begin{cases} W_v, & v \text{ is leaf,} \\ W_v \prod_{e \in DE(v)} I(W_v > \min_{u \in e, u \neq v} \{S_T(u)\}), & \text{otherwise.} \end{cases}$$

Given a weighted rooted tree, the bonus function value of the root is exactly the weight of the root if the root is selected by the greedy algorithm, and is 0 if the the root is not selected, as shown by the following lemma:

Lemma 2.1.5. *Let T be a rooted hypertree, whose vertices are equipped with distinct positive weights. Let γ be the root of T, W_γ be the weight of γ, then we have*

$$S_T(\gamma) = W_\gamma I(\gamma \in \mathscr{I}(T)).$$

Proof. We prove by induction on the height of the tree. When the height is 0, this lemma is true. Now suppose T has height $h > 0$, and this lemma holds for all trees with height less than h. Let $e_k,\ 1 \leq k \leq d$, be all descending edges of the root γ. Then by definition of the bonus function, we have

$$S_T(\gamma) = W_\gamma \prod_{k=1}^{d} I(W_\gamma > \min_{v \in e_k, v \neq \gamma} \{S_T(v)\}).$$

So it suffices to show that

$$\prod_{k=1}^{d} I(W_\gamma > \min_{v\in e_k, v\neq\gamma}\{S_T(v)\}) = I(\gamma\in\mathscr{I}(T)).$$

Let T_v be the subtree of T with root v, such that T_v contains only the edges descending from v. If $W_\gamma > \min_{v\in e_k, v\neq\gamma}\{S_T(v)\}$ for all k such that $1\le k\le d$. For an arbitrary k, pick $v\in e_k$, $v\neq\gamma$, such that $W_\gamma > S_T(v)$, then there are two cases. Firstly, if $W_\gamma < W_v$, then we have $S_T(v)=0$. By inductive assumption, this implies $v\notin\mathscr{I}(T_v)$. Then by Lemma 2.1.1, since T_v is an influence-blocking hypergraph of T, we have $v\notin\mathscr{I}(T)$. This means that γ will not be deleted by e_k. Secondly, if $W_\gamma > W_v$. This also means that γ will not be deleted by e_k. This argument works for all $1\le k\le d$. Therefore, $\gamma\in\mathscr{I}(T)$.

On the other hand, if $W_\gamma < \min_{v\in e_k, v\neq\gamma}\{S_T(v)\}$ for some k, then W_γ must be the smallest-weighted vertex in e_k and $v\in\mathscr{I}(T_v)$ for all $v\in e_k$, $v\neq\gamma$. Since T_v is an influence-blocking hypergraph of T, by Lemma 2.1.1 we have $v\in\mathscr{I}(T)$ for all $v\in e_k$, $v\neq\gamma$. This implies that γ will be deleted by e_k. Therefore, $\gamma\notin\mathscr{I}(T)$. □

Let $T(d,h)$ be the $(r+1)$-uniform rooted hypertree such that all non-leaf vertices have d descending edges, and all leaves have depth h. Let $\tilde{T}(d,h)$ be the $(r+1)$-uniform rooted hypertree such that the root has d descending edges while all other non-leaf vertices have $d-1$ descending edges, and all leaves have depth h.

Let γ be the root of $T(d,h)$. Apply the first step of the greedy algorithm to $T(d,h)$, that is, randomly assign weights to $T(d,h)$. Let $F_{d,h}$ be the distribution function of $S_{T(d,h)}(\gamma)$. That is, $F_{d,h}(x)=\mathbb{P}[S_{T(d,h)}(\gamma)\le x]$. Similarly, let $\tilde{F}_{d,h}$ be the distribution function of $S_{\tilde{T}(d,h)}(\gamma)$. That is, $\tilde{F}_{d,h}(x)=\mathbb{P}[S_{\tilde{T}(d,h)}(\gamma)\le x]$. Note that by Lemma 2.1.5, we have

$$1-F_{d,h}(0)=\mathbb{P}[\gamma\in\mathscr{I}(T(d,h)] \tag{2.9}$$

$$1-\tilde{F}_{d,h}(0)=\mathbb{P}[\gamma\in\mathscr{I}(\tilde{T}(d,h)] \tag{2.10}$$

Also by definition of the bonus function of hypergraphs, $F_{d,h}$ and $\tilde{F}_{d,h}$ satisfy the following recursive equations for all $x \in [0,1]$:

$$F_{d,h}(x) = 1 - \int_x^1 \mathbb{P}[S_{T(d,h)}(\gamma) = W_\gamma | W_\gamma = t]dt = 1 - \int_x^1 [1-(1-F_{d,h-1}(t))^r]^d dt \qquad (2.11)$$

$$\tilde{F}_{d,h}(x) = 1 - \int_x^1 \mathbb{P}[S_{\tilde{T}(d,h)}(\gamma) = W_\gamma | W_\gamma = t]dt = 1 - \int_x^1 [1-(1-F_{d-1,h-1}(t))^r]^d dt \qquad (2.12)$$

In order to get a differential equation, we need to show that $F_{d,h}$ and $\tilde{F}_{d,h}$ converge as $h \to \infty$. We make use of the following lemma:

Lemma 2.1.6. *For any $x \in \mathbb{R}$ and integer $h \geq 0$, the following inequalities hold:*

$$(-1)^h F_{d,h}(x) \leq (-1)^h F_{d,h+1}(x) \qquad (2.13)$$

$$(-1)^h F_{d,h}(x) \leq (-1)^h F_{d,h+2}(x) \qquad (2.14)$$

Proof. We prove inequality (2.13) by induction. First, when $h = 0$, by definition we have $F_{d,0}(x) \leq F_{d,1}(x)$. Now for $h \geq 1$, suppose inequality (2.13) holds for $h-1$. Replace h by $h+1$ in equality (2.11) and consider its difference with the original equality, we have

$$F_{d,h+1}(x) - F_{d,h}(x) = \int_x^1 \left(\left(1-(1-F_{d,h-1}(t))^r\right)^d - \left(1-(1-F_{d,h}(t))^r\right)^d \right) dt$$

Using this equation, we can check that when $F_{d,h}(x) \geq F_{d,h-1}(x)$, we have $F_{d,h+1}(x) \leq F_{d,h}(x)$; and when $F_{d,h}(x) \leq F_{d,h-1}(x)$, we have $F_{d,h+1}(x) \geq F_{d,h}(x)$. Hence, by inductive assumption, we have $(-1)^h F_{d,h}(x) \leq (-1)^h F_{d,h+1}(x)$. This completes the proof for inequality (2.13). Same reasoning gives the proof for inequality (2.14). □

Corollary 2.1.7. *There exist functions $F_{d,even}(x) : \mathbb{R} \to [0,1]$ and $F_{d,odd}(x) : \mathbb{R} \to [0,1]$ such that the sequence of functions $\{F_{d,2k}(x)\}_{k\geq 0}$ converges pointwise to $F_{d,even}(x)$ and the sequence of functions $\{F_{d,2k+1}(x)\}_{k\geq 0}$ converges pointwise to $F_{d,odd}(x)$, and $F_{d,even}(x) \leq F_{d,odd}(x)$ for all*

$x \in \mathbb{R}$.

Proof. As a result of inequality (2.14), for any $x \in \mathbb{R}$, the sequence $\{F_{d,2k}(x)\}_{k\geq 0}$ is increasing and the sequence $\{F_{d,2k+1}(x)\}_{k\geq 0}$ is decreasing. Also, by inequality (2.13), both sequences are bounded. Hence, by the Monotone Convergence Theorem [65], they must converge, which implies the existence of $F_{d,even}(x)$ and $F_{d,odd}(x)$. The inequality can be obtained by considering the inequality (2.13) with $h = 2k$ and $k \to \infty$. □

Similar results as Lemma 2.1.6 and Corollary 2.1.7 for $\tilde{F}_{d,h}$ can also be obtained using the same idea, and we omit the details.

2.2 Expectation

In this section we prove Theorem 1.6.1. The following lemma, which is a modification of Theorem 7 in [36], provide an upper bound for the difference between the probability that a vertex v in a hypergraph G is selected and the probability that the root γ of a rooted hypertree is selected by the greedy algorithm, showing that the performance of the greedy algorithm on G is locally similar to that on a hypertree.

Lemma 2.2.1. *For any integers $r \geq 1$, $d \geq 2$ and $g \geq 4$, let G be an $(r+1)$-uniform d-regular hypergraph with girth g. Let $h_0 = \lfloor \frac{g-3}{2} \rfloor$, $T = \tilde{T}(d,h)$ with $h \geq h_0+1$, let γ be the root of T. Then for every vertex $v \in V(G)$,*

$$|\mathbb{P}[v \in \mathscr{I}(G)] - \mathbb{P}[\gamma \in \mathscr{I}(T)]| \leq \frac{d(d-1)^{h_0}}{r\prod_{k=1}^{h_0+1}(k+\frac{1}{r})}. \tag{2.15}$$

Proof. We apply the first step of greedy algorithm on G and T in the following way. We first give vertices of G i.i.d. weights from the uniform distribution on $[0,1]$. Observe that $N_{h_0+1}(v)$ is a $\tilde{T}(d,h_0+1)$ hypertree, so we can find an isomorphism f that maps $N_{h_0+1}(v)$ to $N_{h_0+1}(\gamma)$. Then we give the vertices in $N_{h_0+1}(\gamma)$ the same weight as their coimage in $N_{h_0+1}(v)$. Finally we give all remaining vertices in T i.i.d. weights from the uniform distribution on $[0,1]$. Then we apply the second step of greedy algorithm on both G and T to get $\mathscr{I}(G)$ and $\mathscr{I}(T)$. In this setting, we have the following estimate:

$$\begin{aligned}\mathbb{P}[v\in\mathscr{I}(G)] =& \mathbb{P}[v\in\mathscr{I}(G),\mathscr{B}_G(v)\subset N_{h_0}(v)]+\mathbb{P}[v\in\mathscr{I}(G),\mathscr{B}_G(v)\not\subset N_{h_0}(v)]\\ =& \mathbb{P}[\gamma\in\mathscr{I}(T),\mathscr{B}_T(\gamma)\subset N_{h_0}(\gamma)]+\mathbb{P}[v\in\mathscr{I}(G),\mathscr{B}_G(v)\not\subset N_{h_0}(v)] \quad \text{(Lemma 2.1.1)}\\ \le& \mathbb{P}[\gamma\in\mathscr{I}(T)]+\mathbb{P}[\mathscr{B}_G(v)\not\subset N_{h_0}(v)].\end{aligned}$$

This implies that

$$\begin{aligned}\mathbb{P}[v\in\mathscr{I}(G)]-\mathbb{P}[\gamma\in\mathscr{I}(T)] &\le \mathbb{P}[\mathscr{B}_G(v)\not\subset N_{h_0}(v)]\\ &\le \frac{d(d-1)^{h_0}}{r\prod_{k=1}^{h_0+1}(k+\frac{1}{r})}. \quad \text{(Lemma 2.1.4)}\end{aligned}$$

We complete the proof by repeating the reasoning above with the roles of $\mathbb{P}[v\in\mathscr{I}(G)]$ and $\mathbb{P}[\gamma\in\mathscr{I}(T)]$ reversed. □

Using similar idea as in the proof above, we can also show that the following limits exist:

Lemma 2.2.2. *For any fixed integer d, the limits $\lim_{h\to\infty}\mathbb{P}[\gamma\in\mathscr{I}(T(d,h))]$ and $\lim_{h\to\infty}\mathbb{P}[\gamma\in\mathscr{I}(\tilde{T}(d,h))]$ exist, where γ denote the root of the rooted hypertrees.*

Proof. We only present the proof of the existence of $\lim_{h\to\infty}\mathbb{P}[\gamma\in\mathscr{I}(\tilde{T}(d,h))]$. The proof of the existence of $\lim_{h\to\infty}\mathbb{P}[\gamma\in\mathscr{I}(T(d,h))]$ is similar and we omit the details. Let h, h' be positive integers with $h'>h$. Using the same idea as in the proof of Lemma 2.2.1, we can show that

$$|\mathbb{P}[\gamma\in\mathscr{I}(\tilde{T}(d,h))]-\mathbb{P}[\gamma\in\mathscr{I}(\tilde{T}(d,h'))]|\le\frac{d(d-1)^{h-1}}{r\prod_{k=1}^{h}(k+\frac{1}{r})}\to 0 \text{ as } h\to\infty.$$

So we conclude that the sequence $\{\mathbb{P}[\gamma\in\mathscr{I}(\tilde{T}(d,h))]\}_{h\ge1}$ is a Cauchy sequence and therefore has a limit. □

Now we are ready to show that $F_{d,h}(x)$ and $\tilde{F}_{d,h}(x)$ converge, and hence get the differential equations we need:

Lemma 2.2.3. *there exist functions $F_d(x)$ and $\tilde{F}_d(x)$such that $F_{d,h}(x)$ converges pointwise to $F_d(x)$ and $\tilde{F}_{d,h}(x)$ converges pointwise to $\tilde{F}_d(x)$ as $h \to \infty$. $F_d(x)$ and $\tilde{F}_d(x)$ satisfy the following equations:*

$$F_d(x) = 1 - \int_x^1 [1 - (1 - F_d(t))^r]^d dt, \tag{2.16}$$

$$\tilde{F}_d(x) = 1 - \int_x^1 [1 - (1 - F_{d-1}(t))^r]^d dt. \tag{2.17}$$

Proof. We only present the proof of the existence of F_d here. The proof of the existence of $\tilde{F}_d$ is similar and we omit the details. By Corollary 2.1.7, there exist $F_{d,even}(x)$ and $F_{d,odd}(x)$ such that $F_{d,2k}(x)$ converges pointwise to $F_{d,even}(x)$ and $F_{d,2k+1}(x)$ converges pointwise to $F_{d,odd}(x)$ as $k \to \infty$. Hence, to prove the existence of F_d, it suffices to show that $F_{d,even}(x) = F_{d,odd}(x)$ for all $x \in \mathbb{R}$. By Lemma 2.2.2, $\lim_{h\to\infty} \mathbb{P}[\gamma \in \mathscr{I}(T(d,h))]$ exists. Since $F_{d,h}(0) = 1 - \mathbb{P}[\gamma \in \mathscr{I}(T(d,h))]$, this implies that $\lim_{h\to\infty} F_{d,h}(0)$ exists. So we have

$$F_{d,even}(0) = \lim_{h\to\infty} F_{d,h}(0) = F_{d,odd}(0).$$

Now consider equation (2.11) with $h = 2k$, and let k go to infinity on both sides, and then use the Dominated Convergence Theorem [65], we have

$$F_{d,even}(x) = 1 - \int_x^1 [1 - (1 - F_{d,odd}(t))^r]^d dt.$$

Similarly, we also have

$$F_{d,odd}(x) = 1 - \int_x^1 [1 - (1 - F_{d,even}(t))^r]^d dt.$$

Take the derivative on both sides and then take the difference of these two equations, we have

$$F'_{d,even}(x) - F'_{d,odd}(x) = [1-(1-F_{d,odd}(x))^r]^d - [1-(1-F_{d,even}(x))^r]^d \geq 0,$$

where the inequality comes from the fact that $F_{d,even} \leq F_{d,odd}$ by Corollary 2.1.7. So for any fixed $x \in [0,1]$,

$$F_{d,even}(x) = F_{d,even}(0) + \int_0^x F'_{d,even}(t)dt \geq F_{d,odd}(0) + \int_0^x F'_{d,odd}(t)dt = F_{d,odd}(x).$$

This combined with the inequality $F_{d,even} \leq F_{d,odd}$, implies $F_{d,even} = F_{d,odd}$. This completes the proof of the existence of $F_d(x)$. Now consider equations (2.11) and (2.12), let $h \to \infty$ and then use the Dominated Convergence Theorem [65], we get the desired differential equations. □

Lemma 2.2.4. *For any integer $d \geq 3$, let $G_d(x) = 1 - F_{d-1}(x)$, then $G_d(x)$ satisfies the following equation:*

$$1 - \sum_{n \geq 0} \binom{n+d-2}{d-2} \frac{G_d(x)^{rn+1}}{rn+1} = x. \tag{2.18}$$

Proof. By equation (2.16), we have

$$G_d(x) = \int_x^1 (1 - G_d(t)^r)^{d-1} dt.$$

Taking derivatives on both sides, we have

$$G'_d(x) = -(1 - G_d(x)^r)^{d-1}.$$

Let $H_d(x) = \sum_{n \geq 0} \binom{n+d-2}{d-2} \frac{x^{rn+1}}{rn+1}$, it is not hard to check that $H'_d(x) = \frac{1}{(1-x^r)^{d-1}}$. So the equation above is equivalent to

$$(H_d(G_d(x)))' = -1.$$

Solving this equation, we obtain

$$\sum_{n\geq 0}\binom{n+d-2}{d-2}\frac{G_d(x)^{rn+1}}{rn+1}=-x+C$$

Let $x=1$, we have $0=-1+C$, which implies $C=1$. This completes the proof. $\square$

Lemma 2.2.5. *For any integer $d\geq 3$, let $\tilde{G}_d(x)=1-\tilde{F}_d(x)$, then we have the following equation:*

$$\tilde{G}_d(x)=G_d(x)-\frac{G_d(x)^{r+1}}{r+1} \tag{2.19}$$

Proof. By equation (2.17),

$$\tilde{G}_d(x)=\int_x^1(1-G_d(t)^r)^d dt$$

Consider changing the variable in the integral by letting $u=G_d(t)$. By equation (2.16), not hard to see $dt=-\frac{du}{(1-u^r)^{d-1}}$. Hence,

$$\tilde{G}_d(x)=-\int_{G_d(x)}^{G_d(1)}(1-u^r)du=G_d(x)-\frac{G_d(x)^{r+1}}{r+1}$$

$\square$

Now we are ready to prove Theorem 1.6.1.

Proof of Theorem 1.6.1. Applying inequality (2.15), we have

$$\begin{aligned}|\frac{\mathbb{E}[|\mathscr{I}(G)|]}{n}-\mathbb{P}[\gamma\in\mathscr{I}(\tilde{T}(d,h))]| &\leq \frac{1}{n}\sum_{v\in V(G)}|\mathbb{P}[v\in\mathscr{I}(G)]-\mathbb{P}[\gamma\in\mathscr{I}(\tilde{T}(d,h))]| \\ &\leq \frac{d(d-1)^{h_0}}{r\prod_{k=1}^{h_0+1}(k+\frac{1}{r})}\end{aligned}$$

Note that this inequality holds for all $h\geq h_0+1$. Let $h\to\infty$, we have

$$|\frac{\mathbb{E}[|\mathscr{I}(G)|]}{n}-\lim_{h\to\infty}\mathbb{P}[\gamma\in\mathscr{I}(\tilde{T}(d,h))]|\leq\frac{d(d-1)^{h_0}}{r\prod_{k=1}^{h_0+1}(k+\frac{1}{r})}.$$

Let $f(d,r) = \lim_{h\to\infty} \mathbb{P}[\gamma \in \mathscr{I}(\tilde{T}(d,h))] = \tilde{G}_d(0)$, then we have the required inequality (2.4). Let $u(d,r) = \lim_{h\to\infty} \mathbb{P}[\gamma \in \mathscr{I}(T(d-1,h))] = G_d(0)$. By Lemma 2.2.4, we know that $u(d,r)$ satisfy equation (2.1). By Lemma 2.2.5, we have

$$f(d,r) = \tilde{G}_d(0) = G_d(0) - \frac{G_d(0)^{r+1}}{r+1} = u(d,r) - \frac{u(d,r)^{r+1}}{r+1}.$$

This completes the proof. □

2.3 Concentration

In this section we prove Theorem 1.6.2. If two vertices u,v are far away from each other, then the indicator of the event that u is selected and the indicator of the event that v is selected by the greedy algorithm have small covariance. This phenomenon can also be used to give an upper bound for the variance of the algorithm.

Lemma 2.3.1. *For any integers $r \geq 1$ and $d \geq 2$, let G be an $(r+1)$-uniform linear hypergraph on n vertices with maximum degree d, then the variance satisfies:*

$$\mathrm{Var}[\mathscr{I}(G)] \leq 3d^2 r^2 e^{r^2(d-1)^3} n. \tag{2.20}$$

Proof. Let $V(G) = \{v_1, v_2, \ldots, v_n\}$, $X_i = I(v_i \in \mathscr{I}(G))$. Then

$$\begin{aligned}
\mathrm{Var}(\mathscr{I}(G)) &= \mathrm{Var}(\sum_{i=1}^{n} X_i) \\
&= \sum_{i=1}^{n} (\mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2) + \sum_{1\leq i\neq j\leq n} (\mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j]) \\
&\leq n + \sum_{1\leq i\leq n} \sum_{\delta\geq 1} \sum_{v_j \in N_\delta(v_i)\setminus N_{\delta-1}(v_i)} (\mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j]),
\end{aligned}$$

where the inequality uses the bound $(\mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2) \leq 1$. For any $1 \leq i \leq n$, we consider the sum

$$\sum_{\delta \geq 1} \sum_{v_j \in N_\delta(v_i) \setminus N_{\delta-1}(v_i)} (\mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j])$$

First, for any $\delta \geq 3$, let $h = \lfloor \frac{\delta-3}{2} \rfloor$, and let $A_{i,h}$ denote the event $\{\mathscr{B}(v_i) \not\subset N_h(v_i)\}$, $A_{i,h}^c$ denote the complement of the event $A_{i,h}$, that is $\{\mathscr{B}(v_i) \subset N_h(v_i)\}$. This event is only determined by the weights of vertices in $N_{h+1}(v_i)$. Notice that for every $v_j \in N_\delta(v_i) \setminus N_{\delta-1}(v_i)$, $N_{h+1}(v_i) \cap N_{h+1}(v_j) = \emptyset$. So $A_{i,h}^c$ and $A_{j,h}^c$ are independent. Then we have,

$$\begin{aligned}
\mathbb{E}[X_i X_j] &= \mathbb{P}[v_i \in \mathscr{I}(G), v_j \in \mathscr{I}(G)] \\
&= \mathbb{P}[v_i \in \mathscr{I}(G), v_j \in \mathscr{I}(G), A_{i,h}^c \cap A_{j,h}^c] + \mathbb{P}[v_i \in \mathscr{I}(G), v_j \in \mathscr{I}(G), A_{i,h} \cup A_{j,h}].
\end{aligned}$$

By Lemma 2.1.1 and the independence between $A_{i,h}^c$ and $A_{j,h}^c$, we have

$$\begin{aligned}
\mathbb{P}[v_i \in \mathscr{I}(G), v_j \in \mathscr{I}(G), A_{i,h}^c \cap A_{j,h}^c] &= \mathbb{P}[v_i \in \mathscr{I}(\mathscr{B}(v_i)), v_j \in \mathscr{I}(\mathscr{B}(v_j)), A_{i,h}^c \cap A_{j,h}^c] \\
&= \mathbb{P}[v_i \in \mathscr{I}(\mathscr{B}(v_i)), A_{i,h}^c]\mathbb{P}[v_j \in \mathscr{I}(\mathscr{B}(v_j)), A_{j,h}^c] \\
&\leq \mathbb{E}[X_i]\mathbb{E}[X_j].
\end{aligned}$$

On the other hand, by Lemma 2.1.4

$$\begin{aligned}
\mathbb{P}[v_i \in \mathscr{I}(G), v_j \in \mathscr{I}(G), A_{i,h} \cup A_{j,h}] &\leq \mathbb{P}[A_{i,h}] + \mathbb{P}[A_{j,h}] \\
&\leq \frac{2d(d-1)^h}{r \prod_{k=1}^{h+1}(k + \frac{1}{r})}
\end{aligned}$$

Hence,

$$\mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j] \leq \frac{2d(d-1)^h}{r \prod_{k=1}^{h+1}(k + \frac{1}{r})}$$

Since G has maximum degree d, we have $|N_\delta(v_i) \setminus N_{\delta-1}(v_i)| \leq d(d-1)^{\delta-1} r^\delta$.

In particular, for odd integer $\delta \geq 3$, we have $\delta = 2h+3$. So the sum

$$\begin{aligned}\sum_{odd\ \delta\geq 3}\sum_{v_j\in N_\delta(v_i)\backslash N_{\delta-1}(v_i)}(\mathbb{E}[X_iX_j]-\mathbb{E}[X_i]\mathbb{E}[X_j]) &\leq \sum_{h\geq 0} d(d-1)^{2h+2}r^{2h+3}\frac{2d(d-1)^h}{r(h+1)!}\\ &= \frac{2d^2}{d-1}\sum_{h\geq 1}\frac{r^{2h}(d-1)^{3h}}{h!}\\ &\leq 2d^2\sum_{h\geq 1}\frac{r^{2h}(d-1)^{3h}}{h!}\end{aligned}$$

For even integer $\delta \geq 3$, we have $\delta = 2h+4$. So the sum

$$\begin{aligned}\sum_{even\ \delta\geq 3}\sum_{v_j\in N_\delta(v_i)\backslash N_{\delta-1}(v_i)}(\mathbb{E}[X_iX_j]-\mathbb{E}[X_i]\mathbb{E}[X_j]) &\leq \sum_{h\geq 0} d(d-1)^{2h+3}r^{2h+4}\frac{2d(d-1)^h}{r(h+1)!}\\ &= 2d^2r\sum_{h\geq 1}\frac{r^{2h}(d-1)^{3h}}{h!}\end{aligned}$$

For $1 \leq \delta \leq 2$, use the bound $\mathbb{E}[X_iX_j]-\mathbb{E}[X_i]\mathbb{E}[X_j] \leq 1$, we have

$$\sum_{1\leq\delta\leq 2}\sum_{v_j\in N_\delta(v_i)\backslash N_{\delta-1}(v_i)}(\mathbb{E}[X_iX_j]-\mathbb{E}[X_i]\mathbb{E}[X_j]) \leq 2d^2r^2$$

Combine the three inequalities above, we have

$$\sum_{\delta\geq 1}\sum_{v_j\in N_\delta(v_i)\backslash N_{\delta-1}(v_i)}(\mathbb{E}[X_iX_j]-\mathbb{E}[X_i]\mathbb{E}[X_j]) \leq 2d^2r^2e^{r^2(d-1)^3}$$

So the variance

$$\mathrm{Var}(\mathscr{I}(G)) \leq n + 2d^2r^2e^{r^2(d-1)^3}n \leq 3d^2r^2e^{r^2(d-1)^3}n$$

□

Proof of Theorem 1.6.2. By Lemma 2.3.1, we know that for fix d and r, there exist a constant c

such that $\text{Var}(\mathscr{I}(G)) \le cn$. Hence, by Chebyshev's Inequality we have

$$\mathbb{P}[||\mathscr{I}(G)| - \mathbb{E}[|\mathscr{I}(G)|]| \ge \sqrt{n}b(n)] \le \frac{\text{Var}(|\mathscr{I}(G)|)}{b(n)^2 n} \to 0, \text{ as } n \to \infty.$$

□

Chapter 2, in full, is a version of the material as it appears in "Randomized greedy algorithm for independent sets in regular uniform hypergraphs with large girth", Random Structures & Algorithms 59.1 (2021): 79-95, co-authored with Jacques Verstraëte. The dissertation author was the primary investigator and author of this paper.

2.4 Appendix

We first collect some real number inequalities:

Proposition 2.4.1. *Let n, r, d be positive integers. Then*

1. *For $x \ge 0$,*
$$\int_0^x e^{t^r} dt = \sum_{n \ge 0} \frac{x^{rn+1}}{n!(rn+1)} \tag{2.21}$$

2. *For $n \le \sqrt{d}$,*
$$\left(1 + \frac{n}{d}\right)^n \le e^{\frac{n^2}{d}} \le e \tag{2.22}$$

3. *For $y \ge 0$,*
$$\left(\frac{y}{n}\right)^n \le e^{\frac{y}{e}} \tag{2.23}$$

Let $u_d = \lim_{h\to\infty} \mathbb{P}[\gamma \in \mathscr{I}(T(d,h))] = u(d+1,r)$. Note that u_d can be viewed as the probability of the root of $T(d,\infty)$ being selected by the greedy algorithm, while $f(d,r)$ can be viewed as the probability of the root of $\tilde{T}(d,\infty)$ being selected by the greedy algorithm.

Proposition 2.4.2. *$f(d,r) \sim (\frac{\log d}{rd})^{\frac{1}{r}}$ as $d \to \infty$.*

Proof. Let $g(d,u) = \sum_{n\geq 0} \binom{n+d-1}{n} \frac{u^{rn+1}}{rn+1}$. It is not hard to see that g is increasing with respect to u. By Lemma 2.2.4, we have $g(d,u_d) = 1$. Now for any $\varepsilon > 0$, let $u = ((\frac{1}{r})^{\frac{1}{r}} + \varepsilon)(\frac{\log d}{d})^{\frac{1}{r}}$, we have

$$\begin{aligned}
g(d,u) &\geq \sum_{n\geq 0} \frac{d^n}{n!} \frac{u^{rn+1}}{rn+1} \\
&= d^{-\frac{1}{r}} \sum_{n\geq 0} \frac{(ud^{\frac{1}{r}})^{rn+1}}{n!(rn+1)} \\
&= d^{-\frac{1}{r}} \int_0^{ud^{\frac{1}{r}}} e^{t^r} dt \qquad \text{(by 2.21)} \\
&\geq d^{-\frac{1}{r}} \int_{(\log d)^{\frac{1}{r}} (\frac{1}{r})^{\frac{1}{r}}}^{(\log d)^{\frac{1}{r}} [(\frac{1}{r})^{\frac{1}{r}} + \varepsilon]} e^{t^r} dt \\
&\geq d^{-\frac{1}{r}} (\varepsilon (\log d)^{\frac{1}{r}}) e^{\frac{\log d}{r}} \\
&= \varepsilon (\log d)^{\frac{1}{r}}.
\end{aligned}$$

This means that $g(d,u) \to \infty$ as $d \to \infty$, hence $u_d \leq [(\frac{1}{r})^{\frac{1}{r}} + \varepsilon](\frac{\log d}{d})^{\frac{1}{r}}$ when d is large enough.

On the other hand, for any $\varepsilon > 0$, let $u = c(\frac{\log d}{d})^{\frac{1}{r}}$, where $c = (\frac{1}{r} - \varepsilon)^{\frac{1}{r}}$, we have

$$\begin{aligned}
g(d,u) &\leq \sum_{n\geq 0} \left(\frac{e(n+d)}{n}\right)^n \frac{u^{rn+1}}{rn+1} \\
&= \sum_{n\geq 0} \frac{u}{rn+1} \left(\frac{e}{n} + \frac{e}{d}\right)^n (c^r \log d)^n
\end{aligned}$$

When $n \geq 4c^r e \log d$, and d is large enough, we have

$$\sum_{n \geq 4c^r e \log d} \frac{u}{rn+1} \left(\frac{e}{n} + \frac{e}{d}\right)^n (c^r \log d)^n \leq u \sum_{n \geq 4c^r e \log d} \left(\frac{2e}{4c^r e \log d}\right)^n (c^r \log d)^n$$

$$= u \sum_{n \geq 4c^r e \log d} \left(\frac{1}{2}\right)^n$$

$$\leq c \left(\frac{\log d}{d}\right)^{\frac{1}{r}} \to 0 \; as \; d \to \infty.$$

When $n \leq 4c^r e \log d$, and d is large enough, we have

$$\sum_{n \leq 4c^r e \log d} \frac{u}{rn+1} \left(\frac{e}{n} + \frac{e}{d}\right)^n (c^r \log d)^n \leq u \sum_{n \leq 4c^r e \log d} \left(1 + \frac{n}{d}\right)^n \left(\frac{c^r e \log d}{n}\right)^n$$

$$\leq ue \sum_{n \leq 4c^r e \log d} \left(\frac{c^r e \log d}{n}\right)^n \qquad \text{(by 2.22)}$$

$$\leq c \left(\frac{\log d}{d}\right)^{\frac{1}{r}} e(4c^r e \log d) e^{c^r \log d} \qquad \text{(by 2.23)}$$

$$= 4e^2 c^{r+1} (\log d)^{\frac{r+1}{r}} d^{-\varepsilon} \to 0 \; as \; d \to \infty.$$

This means that $g(d,u) \to 0$ as $d \to \infty$, hence $u_d \geq (\frac{1}{r} - \varepsilon)^{\frac{1}{r}} (\frac{\log d}{d})^{\frac{1}{r}}$ when d is large enough. These estimates imply that $u_d \sim (\frac{\log d}{rd})^{\frac{1}{r}}$, hence $u_d \to 0$ as $d \to \infty$. Recall that by Theorem 1.6.1,

$$f(d,r) = u(d,r) - \frac{u(d,r)^{r+1}}{r+1} = u_{d-1} - \frac{u_{d-1}^{r+1}}{r+1}$$

Therefore, $f(d,r) \sim (\frac{\log d}{rd})^{\frac{1}{r}}$ as $d \to \infty$. □

Chapter 3

Hypergraph Ramsey Numbers for Berge Cycles

Recall that for $k \geq 2$, a *Berge k-cycle* is a family of sets $e_1, e_2, \ldots, e_k$ such that $e_1 \cap e_2, e_2 \cap e_3, \ldots, e_k \cap e_1$ has a system of distinct representatives, and a Berge cycle is *nontrivial* if $e_1 \cap e_2 \cap \cdots \cap e_k = \emptyset$. The family of nontrivial Berge k-cycles all of whose sets have size r is denoted by $\mathscr{B}_k^r$. In particular, we let $\mathscr{B}_k = \mathscr{B}_k^3$. In support of Conjecture V, we prove the following result for nontrivial Berge cycles of even length:

Theorem 3.0.1 (Theorem 1.8.4)**.** *For $k \geq 3$, and t large enough,*

$$R(t, \mathscr{B}_{2k}) \leq t^{\frac{2k}{2k-1}} \exp\left(4\sqrt{\log t}\right). \tag{3.1}$$

It seems likely that Theorem 1.8.4 can be extended to r-uniform hypergraphs with $r \geq 4$, however when following the proof of Theorem 1.8.4, two obstacles arise. The first is that one requires supersaturation for Berge cycles in r-uniform hypergraphs for $r \geq 3$ (in other words, an r-uniform version of Lemma 3.2.1). A second obstacle is that an r-uniform analog of Lemma 3.2.2 is not straightforward: for instance if an edge e in an r-graph is contained in m Berge cycles of length $2k$, then the number of edges may be as low as $m^{1/(2k-1)}$: take a graph $2k$-cycle, and replace one edge with the hyperedge e, and each other edge with $m^{1/(2k-1)}$ hyperedges. We believe these technical obstacles may be overcome (some of the ideas in the recent paper of

Mubayi and Yepremyan [59] may apply).

The lower bound for $R(t, \mathscr{B}_k^r)$ is closely related to the following notoriously difficult conjecture.

Conjecture VI (Erdős and Simonovits [30]). *There exist graphs on n vertices of girth more than* $2k$ *with* $\Theta(n^{1+1/k})$ *edges.*

This conjecture remains open, except when $k \in \{2,3,5\}$, largely due to the existence of generalized polygons [9, 53, 75]. Towards this conjecture, Lazebnik, Ustimenko and Woldar [51] gave the densest known construction, which has $\Omega(n^{1+2/(3k-2)})$ edges. We prove the following theorem relating this conjecture to lower bounds on Ramsey numbers for nontrivial Berge cycles:

Theorem 3.0.2 (Theorem 1.8.5). *Let* $k \geq 2$, $r \geq 3$. *If Conjecture VI is true, then as* $t \to \infty$,

$$R(t, \mathscr{B}_k^r) = \Omega\left(\left(\frac{t}{\log t}\right)^{\frac{k}{k-1}}\right). \tag{3.2}$$

This shows that if the Erdős-Simonovits Conjecture is true, then Theorem 1.8.4 is tight up to a $t^{o(1)}$ factor. Indeed, following the proof of Theorem 1.8.5, the known construction of Lazebnik, Ustimenko and Woldar [51] would give a weaker lower bound of $\Omega((t/\log t)^{(3k-2)/(3k-4)})$.

Let B_k be the family of all 3-uniform Berge k-cycles. Random graphs together with the Lovász local lemma give $R(t, B_k) \geq t^{(2k-2)/(2k-3)-o(1)}$, see [3] for similar computation. We prove the following theorem, which gives a substantially better lower bound for B_4 if Conjecture VI is true.

Theorem 3.0.3 (Theorem 1.8.6). *If Conjecture VI is true for* $k = 4$, *then as* $t \to \infty$,

$$R(t, B_4) = \Omega\left(\left(\frac{t}{\sqrt{\log t}}\right)^{16/13}\right). \tag{3.3}$$

We prove Theorem 1.8.4 in Section 3.2, Theorem 1.8.5 in Section 3.3, and Theorem 1.8.6 in Section 3.4.

3.1 Preliminaries

We make use of the following elementary lemma, whose proof is a standard probabilistic argument, included for completeness:

Lemma 3.1.1. *Let $d \geq 1$, and let H be a 3-graph of average degree at most d. Then*

$$\alpha(H) \geq \frac{2v(H)}{3d^{\frac{1}{2}}}.$$

Proof. Let X be a subset of $V(H)$ whose elements are chosen independently with probability $p = d^{-1/2}$. We can get an independent set by deleting a vertex for each edge of H contained in X. Then the expected size of such independent set is at least

$$pv(H) - p^3|H| = pv(H) - \frac{p^3 dv(H)}{3} = \frac{2v(H)}{3d^{\frac{1}{2}}}.$$

Hence, there must exist an independent set of size at least the desired lower bound, which completes the proof. □

Lemma 3.1.2. *Let H be a 3-graph on n vertices, and $0 < \varepsilon < 1/2$. Then there exists an induced subgraph G of H satisfying the following properties:*

1. $v(G) \geq n^{1-\frac{2}{\log_2(\frac{1}{\varepsilon})}}$,
2. $\Delta(G) \leq \frac{d(G)}{\varepsilon}$.

Proof. Let $H = G^{(0)}$. We do the following for $i \geq 0$. If $\Delta(G^{(i)}) \leq d(G^{(i)})/\varepsilon$, we let $G = G^{(i)}$. Otherwise, iteratively delete vertices of $G^{(i)}$ with degree at least $d(G^{(i)})$. Each deleted vertex will result in the loss of at least $d(G^{(i)})$ edges. So we can delete at most

$$\frac{|G^{(i)}|}{d(G^{(i)})} = \frac{v(G^{(i)}) \cdot d(G^{(i)})}{3 \cdot d(G^{(i)})} = \frac{v(G^{(i)})}{3} < \frac{v(G^{(i)})}{2}$$

vertices in this step. Let $G^{(i+1)}$ be the subgraph induced by the remaining vertices. Then we have $v(G^{(i+1)}) \geq v(G^{(i)})/2$. If $\Delta(G^{(i+1)}) \leq d(G^{(i+1)})/\varepsilon$, then we let $G = G^{(i+1)}$. Otherwise, we have

$$d(G^{(i+1)}) \leq \varepsilon\Delta(G^{(i+1)}) < \varepsilon d(G^{(i)}).$$

Let $K = 2\log_{1/\varepsilon} n$. We must obtain an induced subgraph G with $\Delta(G) \leq d(G)/\varepsilon$ after at most K repetitions. Otherwise, after K repetitions, since the average degree decreases by at least a factor of ε after each repetition, the remaining graph $G^{(K)}$ will have no edge, which satisfies the condition $\Delta(G^{(K)}) \leq d(G^{(K)})/\varepsilon$. Suppose after $m \leq K$ repetitions we have the desired induced subgraph G with $\Delta(G) < d(G)/\varepsilon$. Since the number of vertices decreases by at most a factor of 2, we also have

$$v(G) \geq \frac{n}{2^m} \geq n^{1-\frac{2}{\log_2(\frac{1}{\varepsilon})}}.$$

This completes the proof. □

We use the following slightly weaker version of a lemma due to Méroueh [56]; the lemma is in fact valid for 3-graphs H with no loose k-cycles:

Lemma 3.1.3. *Let H be a $\mathscr{B}_k$-free 3-graph. Then there exists a subgraph H^* of H such that $|H^*| > |H|/(3k^2)$ and each edge of H^* contains a pair of codegree 1.*

Proof. Given a 3-graph G and a pair of vertices x, y, we say that $\{x,y\}$ is *G-light* if $d_G(x,y) < k$. Let $G_1 = H$, and let H_1 consist of all edges of G_1 containing a G_1-light pair, and let $G_2 = G_1 \backslash H_1$. For $i \geq 2$, let H_i consist of all edges of G_i containing a G_i-light pair, and let $G_{i+1} = G_i \backslash H_i$. Suppose for contradiction that G_k is not empty. Let $e_1 = \{v_1, v_2, v_3\}$ be an edge in G_k, then by definition, $\{v_2, v_3\}$ is not a G_{k-1}-light pair, and hence, there exists an edge $e_2 = \{v_2, v_3, v_4\}$ such that $v_4 \neq v_1$. For $2 \leq i \leq k-1$, let $e_i = \{v_i, v_{i+1}, v_{i+2}\}$ be an edge in G_{k+1-i}. By definition, $\{v_{i+1}, v_{i+2}\}$ is not a G_{k-i}-light pair, and hence, there exists an edge $e_{i+1} = \{v_{i+1}, v_{i+2}, v_{i+3}\}$ in G_{k-i} such that v_{i+3} is distinct from all v_j, $1 \leq j \leq i$. Therefore, we have a tight path of length k in $G_1 = H$, that is, a hypergraph consisting of $k+2$ distinct vertices v_i, $1 \leq i \leq k+2$, and k edges $e_i = \{v_i, v_{i+1}, v_{i+2}\}$, $1 \le i \le k$. This is also a nontrivial Berge k-cycle. Indeed, when k is even, $\{v_2, v_4, \dots, v_k, v_{k+1}, v_{k-1}, \dots, v_3\}$ forms a system of distinct representatives of $\{e_1 \cap e_2, e_2 \cap e_4, e_4 \cap e_6, \dots, e_{k-2} \cap e_k, e_k \cap e_{k-1}, e_{k-1} \cap e_{k-3}, \dots, e_3 \cap e_1\}$, and when k is odd, $\{v_2, v_4, \dots, v_{k+1}, v_k, v_{k-2}, \dots, v_3\}$ forms a system of distinct representatives of $\{e_1 \cap e_2, e_2 \cap e_4, e_4 \cap e_6, \dots, e_{k-3} \cap e_{k-1}, e_{k-1} \cap e_k, e_k \cap e_{k-2}, \dots, e_3 \cap e_1\}$. This results in a contradiction, since H is $\mathscr{B}_k$-free. Therefore, G_k must be empty, and hence H can be partitioned into $k-1$ subgraphs H_i, $1 \le i \le k-1$, such that each H_i consists of edges containing a G_i-light pair, which is also H_i-light. Let H' be a subgraph H_i with the most edges, then by the pigeonhole principle,

$$|H'| > \frac{|H|}{k}.$$

Now consider a graph J whose vertex set is the set of 3-edges of H', and two 3-edges of H' form an edge of J if they share an H'-light pair. It is easy to see that J has maximum degree at most $3k-6$. Then we can greedily take an independent set of J of size at least $v(J)/(3k-5)$, and this independent set correspond to a subgraph H^* of H' such that

$$|H^*| > \frac{|H'|}{3k-5} > \frac{|H|}{3k^2},$$

and each edge of H^* contains a pair of codegree 1. □

3.2 Upper bound of $R(t, \mathscr{B}_{2k})$

In this section we prove the upper bound in Theorem 1.8.4

A key ingredient of the proof of Theorem 1.8.4 is a supersaturation theorem for cycles in graphs: we make use of the following result proved by Simonovits [35] (see Morris and Saxton [57] for stronger supersaturation):

Lemma 3.2.1. *For every $n, k \ge 2$, there exist constants $\gamma, b_0 > 0$ such that for every $b \ge b_0$, any n-vertex graph G with at least $bn^{1+1/k}$ edges contains at least $\gamma b^{2k} n^2$ copies of C_{2k}.*

We next give a simple lemma which says that if a graph has many cycles of length $2k$ containing a fixed edge, then it has many edges.

Lemma 3.2.2. *Let G be a graph containing m cycles of length $2k$, each containing an edge $e \in G$. Then $|G| \geq m^{1/(k-1)}/2$.*

Proof. For each cycle C of length $2k$ containing e, let $M(C)$ be the perfect matching of C containing e. Fixing a matching $M \subset G$ of size k containing e, at most $(k-1)!2^{k-1}$ cycles C have $M(C) = M$. It follows that the number of distinct matchings $M \subset G$ of size k containing e is at least $m/(k-1)!\,2^{k-1}$, and therefore

$$\binom{|G|-1}{k-1} \geq \frac{m}{(k-1)!\,2^{k-1}}.$$

We conclude $|G|^{k-1} \geq m/2^{k-1}$ and therefore $|G| \geq m^{1/(k-1)}/2$. □

Now we are ready to prove Theorem 1.8.4.

Proof of Theorem 1.8.4. It suffices to show that for every large enough integer n, any $\mathscr{B}_{2k}$-free 3-graph H on n vertices must contain an independent set of size at least $n^{(2k-1)/(2k)-5/(2\sqrt{\log n})}$. By Lemma 3.1.2 with $\varepsilon = \exp(-\sqrt{\log_2 n})$, we find an induced subgraph H_0 of H with n_0 vertices, average degree d_0 and maximum degree D_0 such that $n_0 \geq n^{1-2/\sqrt{\log_2 n}}$ and $D_0 < d_0/\varepsilon$. By Lemma 3.1.3, there is a subgraph H_1 of H_0 with at least $|H_0|/(4k^2)$ edges such that each edge of H_1 contains a pair of codegree 1 in H_1. Let $\chi : V(H_1) \to \{1,2,3\}$ be a random 3-coloring and let H_2 consist of all triples in H_1 such that the pair of vertices of colors 1 and 2 has codegree 1 in H_1 and the last vertex in the triple has color 3. The probability that an edge in H_1 is also an edge in H_2 is at least $1/27$, and therefore the expected number of edges in H_2 is at least $|H_1|/27 \geq |H_0|/(108k^2)$. Fix a coloring so that $|H_2| \geq |H_0|/(108k^2)$. Consider the bipartite graph G comprising all pairs of vertices of colors 1 and 2 contained in an edge of H_2. Thus, $|G| = |H_2|$ and G has average degree $d_G \geq d_0/(108k^2)$. For convenience, let $b > 0$ be defined by $d_G = 2bn_0^{1/k}$ so $|G| = bn_0^{1+1/k}$. By Lemma 3.2.1, there exist constants $\gamma, b_0 > 0$ such that if

$b > b_0$, then G must contain at least $\gamma b^{2k} n_0^2$ copies of C_{2k}. Notice that we must have $1/\varepsilon > b_0$ when n is large enough. The proof is split into two cases.

Case 1. $b \geq 1/\varepsilon$. By the pigeonhole principle, there exists an edge e such that the number of C_{2k} containing e in G is at least

$$\frac{2k\gamma b^{2k} n_0^2}{|G|} = 2k\gamma b^{2k-1} n_0^{1-\frac{1}{k}}.$$

Let G' be the union of all $2k$-cycles in G containing e. Then by Lemma 3.2.2, for some constant c,

$$|G'| \geq cb^{2+\frac{1}{k-1}} n_0^{\frac{1}{k}} = \frac{1}{2} cb^{1+\frac{1}{k-1}} d_G \geq \frac{1}{216k^2} c\varepsilon^{-1-\frac{1}{k-1}} d_0 > D_0$$

provided n is large enough. Let C be a $2k$-cycle in G containing e. Then there exist edges $e_1 \cup \{v_1\}, e_2 \cup \{v_2\}, \ldots, e_{2k} \cup \{v_{2k}\}$ in H_2 where $e_1, e_2, \ldots, e_{2k} \in C$ and $v_1, v_2, \ldots, v_{2k}$ have color 3. Since H_2 is $\mathscr{B}_{2k}$-free, for some vertex z we have $v_1 = v_2 = \cdots = v_{2k} = z$. Since each cycle C in G' contain e, they must have the same z. Now the degree of z in H_2 is at least $|G'| > D_0$, which contradicts the fact that H_0 has maximum degree at most D_0.

Case 2. $b < 1/\varepsilon$. In this case, $d_G < 2n_0^{1/k}/\varepsilon$ and so $d_0 < (216k^2/\varepsilon)n_0^{1/k}$. By Lemma 3.1.1 on H_0,

$$\alpha(H) \geq \alpha(H_0) \geq \frac{2n_0}{3d_0^{\frac{1}{2}}} \geq \frac{2}{3}\left(\frac{216k^2}{\varepsilon}\right)^{-\frac{1}{2}} n_0^{\frac{2k-1}{2k}} \geq \frac{1}{9\sqrt{6}k} n^{\frac{2k-1}{2k} - \frac{5k-2}{2k\sqrt{\log_2 n}}} > n^{\frac{2k-1}{2k} - \frac{5}{2\sqrt{\log n}}}.$$

Now let $n = t^{\frac{2k}{2k-1} + \frac{4}{\sqrt{\log t}}}$. Clearly, $\log n > \frac{2k}{2k-1}\log t$. Hence, an n-vertex $\mathscr{B}_{2k}$-free 3-graph H contains an independent set of size

$$n^{\frac{2k-1}{2k} - \frac{5}{2\sqrt{\log n}}} = t^{(\frac{2k}{2k-1} + \frac{4}{\sqrt{\log t}})(\frac{2k-1}{2k} - \frac{5}{2\sqrt{\log n}})} > t$$

provided n is large enough. Therefore, we have $R(t, \mathscr{B}_{2k}) < t^{\frac{2k}{2k-1} + \frac{4}{\sqrt{\log t}}}$. □

In fact, by more careful computation, we can obtain a slightly better upper bound $R(t,\mathscr{B}_{2k}) < t^{\frac{2k}{2k-1}+\frac{c}{\sqrt{\log t}}}$, where $c > \frac{5k-2}{2k-1} \cdot \sqrt{\frac{(2k)\log 2}{2k-1}}$.

3.3 Lower bound of $R(t,\mathscr{B}_{2k})$

In this section we prove the lower bound in Theorem 1.8.5. We will use the following lemma to get a large bipartite subgraph with large minimum degree and small maximum degree:

Lemma 3.3.1. *Let $k \geq 3$, $c > 0$, and let G be an n-vertex graph of girth more than $2k$ with more than $2cn^{1+1/k}$ edges. Then there exists a bipartite subgraph G' of G such that $\delta(G') \geq cn^{1/k}$, $\Delta(G') \leq n^{1/k}/c^{k-1}$, and $v(G') \geq c^k n$.*

Proof. A maximum cut of G gives a bipartite subgraph with at least $cn^{1+1/k}$ edges. A subgraph G' of this bipartite subgraph of minimum degree at least $cn^{1/k}+1$ may be obtained by repeatedly removing vertices of degree at most $cn^{1/k}$. Let $\Delta := \Delta(G')$ be the maximum degree of G', and let v be a vertex of maximum degree, then the number of vertices at distance k from v is at least $\Delta c^{k-1}n^{(k-1)/k}$, since G has girth larger than $2k$. In particular, $\Delta c^{k-1}n^{(k-1)/k} \leq n$ and so $\Delta \leq n^{1/k}/c^{k-1}$. The number of vertices in G' is at least $c^k n$, since G' has minimum degree at least $cn^{1/k}+1$ and girth larger than $2k$. □

Let $r \geq 2$, a *star* with vertex set V is an r-graph on V consisting of all edges containing a fixed vertex of V, i.e., the edge set of a star is $\{e \subset V : |e| = r, v \in e\}$ for some vertex $v \in V$. Let integers $d \geq m$ and let $S_{d,m}$ be a d-vertex r-graph consisting of m vertex-disjoint stars of size $\lfloor d/m \rfloor$ or $\lceil d/m \rceil$.

Lemma 3.3.2. *Let integer $r \geq 2$, and let integers $d \geq m$.The probability that a uniformly chosen set of s vertices of $S_{d,m}$ is independent is at most*

$$\exp\left(-\frac{m(s-rm)}{2d}\right).$$

Proof. Let the vertex sets of these stars be $V_1, V_2, \ldots, V_m$. The probability that a uniformly chosen set of s_i vertices in V_i is independent in $S_{d,m}$ is at most $1 - s_i/\lceil d/m \rceil \leq 1 - ms_i/2d$ if $s_i \geq r$, and is 1 if $s_i < r$. Hence, this probability is at most $1 - m(s_i - r)/2d$ for $0 \leq s_i \leq d$. Therefore a uniformly chosen set $I \subset S_{d,m}$ of s vertices with $|I \cap V_i| = s_i$ is independent with probability at most

$$\prod_{i=1}^{m} \left(1 - \frac{m(s_i - r)}{2d}\right) \leq \exp\left(-\sum_{i=1}^{m} \frac{m(s_i - r)}{2d}\right) = \exp\left(-\frac{m(s - rm)}{2d}\right).$$

□

Now we are ready to prove Theorem 1.8.5.

Proof of Theorem 1.8.5. It suffices to show that for n large enough, there exists an n-vertex $\mathscr{B}_k^r$-free r-graph with independence number $O(n^{1-\frac{1}{k}} \log n)$. Let G be an n-vertex graph of girth more than $2k$ with $2cn^{1+1/k}$ edges for some positive constant c. By Lemma 3.3.1, there exists a bipartite subgraph G' of G with at least $N = c^k n$ vertices, minimum degree at least $cn^{1/k}$ and maximum degree at most $n^{1/k}/c^{k-1}$. Let X, Y be the parts of this bipartite graph where $|Y| \geq |X|$. Let $m = 8\log n/c^k$. We form an r-graph H with vertex set Y by placing a random copy of $S_{d(x),m}$ on the vertex set $N_{G'}(x)$, the neighborhood of x in G', independently for each $x \in X$. Since G' has girth more than $2k$, it is straightforward to check that H does not contain any nontrivial Berge k-cycle. We now compute the expected number of independent sets of size $t = rmn^{1-1/k}/c^{k+1}$ in H. Clearly, $\log t \geq (1 - 1/k)\log n$. If H has no independent set of size t with positive probability, then since $v(H) \geq N/2$, we find that

$$R(t, \mathscr{B}_k^r) \geq N/2 \geq \frac{c^k}{2}\left(\frac{c^{2k+1}t}{8r\log n}\right)^{\frac{k}{k-1}} \geq c_{k,r}\left(\frac{t}{\log t}\right)^{\frac{k}{k-1}},$$

for some positive constant $c_{k,r}$. This is enough to prove Theorem 1.8.5.

For an independent t-set I in H, $I \cap N_{G'}(x)$ is an independent set in $S_{d(x),m}$ for all $x \in X$.

Since these events are independent, setting $s(x) = |I \cap N_{G'}(x)|$, and applying Lemma 3.3.2 gives:

$$\begin{aligned}\mathbb{P}(I \text{ independent in } H) &\leq \prod_{x\in X} \exp\left(-\frac{m(s(x)-rm)}{2d(x)}\right) \\ &= \exp\left(-\sum_{x\in X}\frac{ms(x)}{2d(x)} + \sum_{x\in X}\frac{rm^2}{2d(x)}\right).\end{aligned}$$

For every $x \in X$, $cn^{1/k} \leq d(x) \leq n^{1/k}/c^{k-1}$ and therefore

$$\mathbb{P}(I \text{ independent in } H) \leq \exp\left(-\frac{c^{k-1}m\sum_{x\in X}s(x)}{2n^{1/k}} + \frac{|X|rm^2}{2cn^{1/k}}\right).$$

Now $\sum_{x\in X} s(x)$ is precisely the number of edges of G' between X and I. Since every vertex in I has degree at least $cn^{1/k}$, this number of edges is at least $cn^{1/k}t = rmn/c^k$. Consequently, using $|X| < n/2$,

$$\mathbb{P}(I \text{ independent in } H) \leq \exp\left(-\frac{c^k mt}{2} + \frac{c^k mt}{4}\right) = \exp\left(-\frac{c^k mt}{4}\right).$$

The expected number of independent sets of size t is at most

$$\binom{n}{t}\exp\left(-\frac{c^k mt}{4}\right) < \exp\left(t\log n - \frac{c^k mt}{4}\right) = \exp\left(-t\log n\right).$$

This is vanishing as $n \to \infty$, and the proof of Theorem 1.8.5 is complete. □

3.4 Lower bound of $R(t, B_4)$

In this section we prove the upper bound in Theorem 1.8.6. Lazebnik and Verstraëte [52] showed that there exist n-vertex B_4-free 3-graphs with $(1/6+o(1))n^{3/2}$ triples. More specifically, for n large enough, there exists a linear n-vertex B_4-free 3-graphs J_n with $n^{3/2}/10$ triples and maximum degree at most $n^{1/2}$. We want to find an upper bound for the probability that a random s-set is independent in J_n.

We make use of the following lemma.

Lemma 3.4.1. *Let n, s be integers such that $s < \sqrt{n}/2$. For n large enough, the probability that a uniformly chosen set of s vertices of J_n is independent is at most*

$$\exp\left(-\frac{s^3-216}{80n^{3/2}}\right).$$

When $s \geq \sqrt{n}/2$, the probability is at most $639/640$.

Proof. This is trivial when $s < 6$. When $6 < s < \sqrt{n}/2$, let X be the uniformly chosen s-set. For any edge $e \in E(J_n)$, let A_e be the event that $e \in X$. Then by inclusion-exclusion principle, for n large enough, the probability that X is not independent is at least

$$\begin{aligned}
&\sum_{e\in E(J_n)} \mathbb{P}(A_e) - \sum_{\{e,f\}\subset E(J_n)} \mathbb{P}(A_e \wedge A_f) \\
\geq & \frac{1}{\binom{n}{s}}\left(\frac{n^{3/2}}{10}\binom{n-3}{s-3} - n\binom{n^{1/2}}{2}\binom{n-5}{s-5} - \binom{n^{3/2}/10}{2}\binom{n-6}{s-6}\right) \\
\geq & \frac{s^3}{40n^{3/2}}\left(1-\frac{4s^3}{n^{3/2}}\right) \\
\geq & \frac{s^3}{80n^{3/2}}.
\end{aligned}$$

Therefore, for $s > 6$ and n large enough, the probability that X is independent is at most

$$1-\frac{s^3}{80n^{3/2}} \leq \exp\left(-\frac{s^3}{80n^{3/2}}\right) < \exp\left(-\frac{s^3-216}{80n^{3/2}}\right).$$

When $s \geq \sqrt{n}/2$, the probability is at most

$$1-\frac{(\sqrt{n}/2)^3}{80n^{3/2}} = \frac{639}{640}.$$

This completes the proof. $\square$

Now we are ready to prove Theorem 1.8.6.

Proof of Theorem 1.8.6. Let G be an n-vertex graph of girth more than 8 with $2c_1n^{5/4}$ edges for some positive constant c_1. By Lemma 4, there exists a bipartite subgraph G' of G with at least $N = c_1^4 n$ vertices, minimum degree at least $c_1 n^{1/4}$ and maximum degree at most $n^{1/4}/c_1^3$. Let X, Y be the parts of this bipartite graph where $|Y| \geq |X|$. We form a 3-graph H with vertex set Y by placing a random copy of $J_{d(x)}$ on the vertex set $N_{G'}(x)$, the neighborhood of x in G, independently for each $x \in X$. Since G has girth more than $2k$, it is straightforward to check that H does not contain any Berge 4-cycle. Let $m = 8c_1^{1/4}\sqrt{\log n}$, and let $t = mn^{13/16}$. Clearly, $\log t > 13 \log n/16$. If H has no independent sets of size t with positive probability, then since $v(H) \geq N/2$, we conclude that

$$R(t, B_4) \geq N/2 \geq \frac{c_1^4}{2}\left(\frac{t}{8c_1^{1/4}\sqrt{\log n}}\right)^{16/13} \geq c_2\left(\frac{t}{\sqrt{\log t}}\right)^{16/13},$$

for some positive constant c_2. This is enough to prove Theorem 1.8.6.

Let A be a t-set in Y, and let

$$X_A = \{x \in X : |N_{G'}(x) \cap A| \geq \sqrt{t}/2\}, \ \ \overline{X}_A = X \backslash A.$$

Case 1: When $|X_A| < n^{5/6}$. Since the induced bipartite subgraph of G' on $X_A \cup A$ has girth 8, the number of edges of G' between X_A and A is less than $(n^{5/6})^{5/4} = n^{25/24}$. If A is independent in H, then $N_{G'}(x) \cap A$ is also independent in $J_{d(x)}$ for all $x \in X$. Since these events are independent, setting $s(x) = |N_{G'}(x) \cap A|$, and applying Lemma 3.4.1 gives

$$\begin{aligned}\mathbb{P}(A \text{ independent in } H) &\leq \prod_{x \in \overline{X}_A} \exp\left(-\frac{s(x)^3 - 216}{80d(x)^{3/2}}\right)\\ &= \exp\left(-\sum_{x \in \overline{X}_A}\frac{s(x)^3}{80d(x)^{3/2}} + \sum_{x \in \overline{X}_A}\frac{27}{10d(x)^{3/2}}\right).\end{aligned}$$

For every $x \in X$, $c_1 n^{1/4} \le d(x) \le n^{1/4}/c_1^3$ and hence together with Jenson's inequality we have

$$\mathbb{P}(A \text{ independent in } H) \le \exp\left(-\frac{c_1^{9/2}\sum_{x\in\overline{X}_A} s(x)^3}{80n^{3/8}} + \frac{27|\overline{X}_A|}{10c_1^{3/2}n^{3/8}}\right)$$
$$\le \exp\left(-\frac{c_1^{9/2}(\sum_{x\in\overline{X}_A} s(x))^3}{80n^{3/8}|\overline{X}_A|^2} + \frac{27|\overline{X}_A|}{10c_1^{3/2}n^{3/8}}\right).$$

Note that $\sum_{x\in\overline{X}_A} s(x)$ is exactly the number of edges of G' between $\overline{X}_A$ and A, which is at least

$$tc_1n^{1/4} - n^{25/24} = (1-o(1))c_1mn^{17/16}.$$

Also note that $|\overline{X}_A| < N/2 = c_1^4 n/2$. Consequently,

$$\mathbb{P}(A \text{ independent in } H) \le \exp\left(-\frac{(1-o(1))m^3n^{13/16}}{20c_1^{1/2}} + \frac{27c_1^{5/2}n^{5/8}}{20}\right)$$
$$< \exp\left(-\frac{m^3n^{13/16}}{32c_1^{1/2}}\right).$$

Case 2: When $|X_A| \ge n^{5/6}$. Applying Lemma 6 gives

$$\mathbb{P}(A \text{ independent in } H) \le (639/640)^{|X_A|} \le \exp(-n^{5/6}/640) < \exp\left(-\frac{m^3n^{13/16}}{32c_1^{1/2}}\right).$$

In both cases we have $\mathbb{P}(A \text{ independent in } H) < \exp\left(-\frac{m^3n^{13/16}}{32c_1^{1/2}}\right)$. Therefore the expected number of independent sets of size t in H is at most

$$\binom{n}{t}\exp\left(-\frac{m^3n^{13/16}}{32c_1^{1/2}}\right) < \exp\left(mn^{13/16}\log n - \frac{m^3n^{13/16}}{32c_1^{1/2}}\right) = \exp\left(-mn^{13/16}\log n\right).$$

This is vanishing as $n \to \infty$, which completes the proof of Theorem 1.8.6.

□

Chapter 3, in full, is a version of the material as it appears in "Ramsey Numbers for

Nontrivial Berge Cycles", SIAM Journal on Discrete Mathematics 36.1 (2022): 103-113, co-authored with Jacques Verstraëte. The dissertation author was the primary investigator and author of this paper.

Chapter 4

Hypergraph Ramsey Numbers for Loose cycle

Recall that for $k, r \geq 3$, an r-uniform *loose k-cycle*, denoted C_k^r, is an r-graph with a cyclic list of edges $e_1, e_2, \ldots, e_k$ such that consecutive edges intersect in exactly one vertex and nonconsecutive sets are disjoint.

The following conjecture was proposed in [47]:

Conjecture V. *For $r, k \geq 3$,*

$$R(t, C_k^r) = t^{\frac{k}{k-1}+o(1)}. \tag{4.1}$$

Towards this conjecture, Méroueh [60] showed $R(t, C_k^3) = O(t^{1+1/\lfloor (k+1)/2 \rfloor})$ for $k \geq 3$ and $R(t, C_k^r) = O(t^{1+1/\lfloor k/2 \rfloor})$ for $r \geq 4$. In particular, he showed that $R(t, C_4^3) = O(t^{3/2})$. In this chapter, we proved a stronger upper bound for $R(t, C_4^3)$, which matches Conjecture V when $r = 3$ and $k = 4$.

Theorem 4.0.1 (Theorem 1.8.3)**.** *As $t \to \infty$, there exists constant $c > 0$ such that,*

$$R(t, C_4^3) < t^{\frac{4}{3}} \exp\left((1 + o(1)) \frac{8\sqrt{3}}{9} \sqrt{\log t} \right). \tag{4.2}$$

It seems likely that Theorem 1.8.3 can be extended to r-uniform hypergraphs with $r \geq 3$. In order to do this, it suffices to show that for any n-vertex 4-graph H, there exists an induced subgraph H' of H with $n^{1-o(1)}$ vertices such that $e(H') < n^{7/4+o(1)}$.

Also, we believe that the strategy we used to prove Theorem 1.8.3 should be extended to loose $2k$-cycles for $k \geq 3$. The obstacle is that we do not have variants of Lemma 4.2.4 and Lemma 4.2.5 for loose $2k$-cycles.

4.1 Preliminaries

We make use of the following elementary lemma, which can be considered as a 3-uniform version of Turán's Theorem, whose proof is a standard probabilistic argument:

Lemma 4.1.1. *Let $d \geq 1$, and let H be a 3-graph of average degree d. Then*

$$\alpha(H) \geq \frac{2v(H)}{3d^{\frac{1}{2}}}.$$

Proof. Let X be a subset of $V(H)$ whose elements are chosen independently with probability $p = d^{-1/2}$. We can get an independent set by deleting a vertex for each edge of H contained in X. Then the expected size of such independent set is at least

$$pv(H) - p^3|H| = pv(H) - \frac{p^3 dv(H)}{3} = \frac{2v(H)}{3d^{\frac{1}{2}}}.$$

Hence, there must exist an independent set of size larger than the required lower bound, which completes the proof. □

An obstacle in the proof of Theorem 1.8.3 is the presence of vertices of degree much larger than $d(H)$. We deal with these vertices using the following lemma:

Lemma 4.1.2. *Let H be an n-vertex 3-graph, and let $1 < t < n$. Then there exists an induced subgraph G of H such that $\Delta(G) \leq 2td(G)$ and $v(G) \geq n^{1-1/\log t}$.*

Proof. The idea is to repeatedly delete vertices with large degree until we obtain the desired induced subgraph. We set $H = G_0$, and run the following algorithm for integers $i \geq 0$:

1. If $\Delta(G_i) \leq 2td(G_i)$, set $G = G_i$ and STOP.

2. Set $G_i^{(0)} = G_i$, we do the following for $j \geq 0$: If there is a vertex $u_j \in V(G_i^{(j)})$ such that $u_j > 2d(G_i)$, let $G_i^{(j+1)} = G_i^{(j)} \backslash \{u_j\}$; otherwise, set $G_{i+1} = G_i^{(j)}$.

If the algorithm does not stop for some integer $i \geq 1$, then the average degree must decrease significantly. By definition we have $d(G_i) \leq \Delta(G_i)/(2t) < d(G_{i-1})/t$, which implies that $d(G_0) > t^i \cdot d(G_i)$.

Suppose the algorithm stop for some integer k. We claim that $d(G_{k-2}) > 1/2$. Otherwise, by definition $|G_{k-1}| = 0$, and hence $G = G_{k-1}$, which contradicts the selection of k. So we have

$$n^2/2 > d(G_0) > t^{k-2} \cdot d(G_{k-2}) = t^{k-2}/2,$$

which implies that $k < 2\log_t n + 2$.

On the other hand, the number of vertices decreases much slower comparing to the average degree. In step 2, by definition $|G_i^{(j+1)}| \leq |G_i^{(j)}| - 2d(G_i)$. This implies $|G_i^{(0)}| \geq |G_i^{(j)}| + 2j \cdot d(G_i) \geq 2j \cdot d(G_i)$ for each $j \geq 0$, and hence $j \leq |G_i^{(0)}|/2d(G_i) = v(G_i)/6$. So we have $v(G_{i+1}) \geq (\frac{5}{6})v(G_i)$, which implies that for n large enough

$$v(G) = v(G_k) \geq \left(\frac{5}{6}\right)^k n > \left(\frac{5}{6}\right)^{2\log_t n+2} n > n^{1-1/\log t}.$$

□

Definition 4. *Let H be a 3-graph. We say a pair $\{u,v\} \in V(H)$ is H-**heavy** if $d_H(u,v) \geq 6$. Otherwise, it is H-**light**. An edge of H is H-k-**heavy** if it contain k H-heavy pairs.*

An important property of the heavy pairs is that any non-uniform loose 4-cycle consisting of triples and heavy pairs can be extended to a 3-uniform loose 4-cycle C_4^3 (See picture in Figure 4.1):

Lemma 4.1.3. *Let H be a 3-graph. If there are* 4 *sets e_1, e_2, e_3, and e_4 such that e_i is either an edge of H or an H-heavy pair, and these* 4 *sets form a loose* 4*-cycle, then H must contain a copy* *of* C_4^3.

Proof. If $e_i = \{u,v\}$ is an H-heavy pair, then, since $d_H(u,v) \geq 6$, we can find a vertex w_i such that $e_i \cup \{w_i\}$ is an edge of H and $w_i \notin \cup_{j=1}^4 e_j$. Replace all the H-heavy pairs by their corresponding edges $e_i \cup \{w_i\}$ in H. These edges form a C_4^3. □

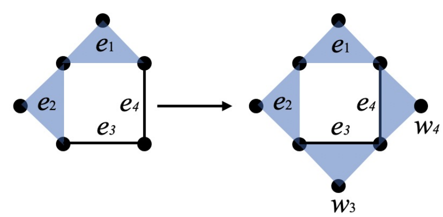

Figure 4.1. Non-uniform loose 4-cyle can be extended to a C_4^3.

Definition 5. *Let H be a 3-graph. A* bowtie *in H is a subgraph B on 5 vertices* $\{v,w,x,y,z\}$ *with a pair of edges* $\{\{v,w,x\},\{x,y,z\}\}$ *such that all pairs but* $\{w,x\}$ *and* $\{x,y\}$ *in* ∂B *are H-light (See Figure 4.2). We let* $\beta(H)$ *denote the number of bowties in H.*

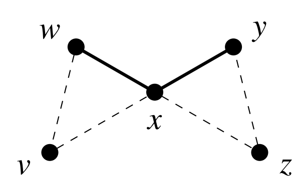

Figure 4.2. Heavy lines represent H-heavy pairs while dashed lines represent H-light pairs.

We make use of the following technical lemma:

Lemma 4.1.4. *Let H be an n-vertex* C_4*-free 3-graph, then*

$$\beta(H) < 10^5 n^2. \tag{4.3}$$

Proof. Let $\beta = \beta(H)$. Give each vertex of H a color uniformly at random from $\{1,2,3,4,5\}$, and let $C(v)$ denote the color of v. A bowtie $B = \{\{v,w,x\},\{x,y,z\}\}$, where $\{w,x\}$ and $\{x,y\}$ are H-heavy pairs, is properly colored if $C(v) = 1, C(w) = 2, C(x) = 3, C(y) = 4$ and $C(z) = 5$.

Note that the expected number of properly colored bowties is $2\beta/3125 > \beta/1600$, so we can fix a coloring such that the number is at least $\beta/1600$. For any vertex v with $C(v) = 1$, let $\mathscr{B}(v)$ denote the set of properly colored bowties containing v.

Then by Pigeonhole Principle, there exists a vertex v such that $|\mathscr{B}(v)| > \beta/(1600n)$. Let L be a graph on the vertices with color 2 or 3 such that a pair $\{w,x\}$ is an edge of L if and only if $\{v,w,x\}$ is an edge in a properly colored bowtie. By the definition of H-light pair, not hard to check that $\Delta(L) \leq 5$. Hence by Vizing's Theorem [76], the edge chromatic number of L is less than 6. So there exists a subset $\mathscr{B}'$ of $\mathscr{B}(v)$ with $|\mathscr{B}'| \geq |\mathscr{B}(v)|/6 > \beta/(10000n)$, such that the corresponding pairs $\{w,x\}$ of all bowties in $\mathscr{B}'$ form a matching. Let $S = \{z_1, z_2, \ldots, z_l\}$ be a set such that each z_i is contained in a bowtie $B \in \mathscr{B}'$ and $C(z_i) = 5$. For two different bowties $B_1 = \{\{v,w_1,x_1\},\{x_1,y_1,z_1\}\}$ and $B_2 = \{\{v,w_2,x_2\},\{x_2,y_2,z_2\}\}$ in $\mathscr{B}'$, we claim that if $z_1 = z_2$, then $\{x_1,w_1\} = \{x_2,w_2\}$. Otherwise, not hard to check that there would be a loose 4-cycle consisting of edges of H and H-heavy pairs (See pictures in Figure 4.3), and hence a C_4^3 by Lemma 4.1.3.

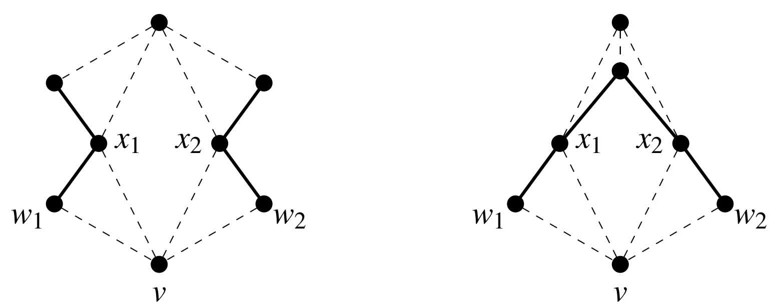

Figure 4.3. 2 possible ways to get a loose 4-cycle when $z_1 = z_2$ and $\{x_1,w_1\} \neq \{x_2,w_2\}$.

Note that $\{x_1,z_1\}$ is an H-light pair, so each $z_i \in S$ can be contained in at most 5 bowties in $\mathscr{B}'$. This implies $|S| \geq |\mathscr{B}'|/5$, and hence $n > |S| > 10^{-5}\beta/n$.

□

Next we introduce an extremal result on rainbow Turán numbers, which are closely related to the extremal problems for loose cycles. In an edge-colored graph, we say that a subgraph is *rainbow* if no two of its edges have the same color. The *rainbow Turán number* of a graph H, denoted $\mathrm{ex}^*(n,H)$, is the maximum number of edges in an n-vertex properly edge-colored graph that does not contain a rainbow H as a subgraph. Keevash, Mubayi, Sudakov and Verstraëte [42] showed that $\mathrm{ex}^*(n,C_{2k}) = \Omega(n^{1+1/k})$ for any integer $k \geq 2$. They conjectured that this is tight and verified it for $k \in \{2,3\}$. Recently, Janzer [40] established the conjecture.

Theorem 4.1.5. *For any integer $k \geq 2$,*

$$\mathrm{ex}^*(n,C_{2k}) = O(n^{1+\frac{1}{k}}). \tag{4.4}$$

The following lemma is a simple application of Theorem 4.1.5. We will use this lemma in the proofs of Theorem 1.8.3.

Lemma 4.1.6. *Let H be a linear 3-graph on n vertices. If H does not contain C_{2k}^3 as a subgraph, then the number of edges in H is $O(n^{1+1/k})$.*

Proof. We first take a random tripartition of H. color the vertices of H uniformly at random with three colors $\{1,2,3\}$. Let H' be the spanning subgraph of H containing all edges with distinct colors. Note that the probability that an edge in H is an edge in H' is $2/9$. So we can fix a coloring of H such that $|H'| \geq 2|H|/9$. Let G be the bipartite graph whose two parts are the set of vertices in H with color 1 and the set of vertices with color 2, and whose edges are pairs of vertices contained in some 3-edge in H'. We color each edge $\{x,y\}$ in G with the vertex z in H such that $\{x,y,z\}$ form a 3-edge in H'. Since H' is linear, we know that $|G| = |H'|$ and that G is properly edge-colored. Also, G does not contain a rainbow C_{2k}, because H' does not contain a C_{2k}^3. So $|G| = O(n^{1+1/k})$ by Theorem 4.1.5, and hence $|H| = O(n^{1+1/k})$. □

A key ingredient of the proof of Theorem 1.8.3 is a supersaturation theorem for cycles in graphs: we make use of the following result proved by Simonovits [35] (see Morris and Saxton [57] for stronger supersaturation):

Lemma 4.1.7. *For every $k \geq 2$, there exist constants $\gamma, b_0 > 0$ such that for every $b \geq b_0$ and integer $n \geq 1$, any n-vertex graph G with at least $bn^{1+1/k}$ edges contains at least $\gamma b^{2k} n^2$ copies of C_{2k}.*

We next give a simple lemma which says that if a graph has many cycles of length $2k$ containing a fixed edge, then it has many edges.

Lemma 4.1.8. *Let G be a graph containing m cycles of length $2k$, each containing an edge $e \in G$. Then $|G| \geq m^{1/(k-1)}/2$.*

Proof. For each cycle C of length $2k$ containing e, let $M(C)$ be the perfect matching of C containing e. Fixing a matching $M \subset G$ of size k containing e, at most $(k-1)!2^{k-1}$ cycles C have $M(C) = M$. It follows that the number of distinct matchings $M \subset G$ of size k containing e is at least $m/(k-1)!\,2^{k-1}$, and therefore

$$\binom{|G|-1}{k-1} \geq \frac{m}{(k-1)!\,2^{k-1}}.$$

We conclude $|G|^{k-1} \geq m/2^{k-1}$ and therefore $|G| \geq m^{1/(k-1)}/2$. □

4.2 Upper bound of $R(t, C_4^3)$

In this section, we prove Theorem 1.8.3. The key component of the proof of Theorem 1.8.3 is the following lemma:

Lemma 4.2.1 (Main Lemma)**.** *Let H be a C_4^3-free 3-graph on n vertices, then, when n is large enough, there exists an induced subgraph H^* of H on n_1 vertices such that $n_1 > (n/5)^{1-\frac{1}{(2\sqrt{\log n})}}$ and*

$$d(H^*) = O(\sqrt{n}\exp(\sqrt{\log n})). \tag{4.5}$$

Proof Ideas. Here we outline our strategy to prove the Main Lemma: First, we partition H into 2 subgraphs H_1 and H_2 such that none of them contain any 3-heavy edge. Then we show that the number of H_i-0-heavy edges is $O(n^{3/2})$ for both $i = 1,2$. After that, we show that there exists an induced subgraph H' of H such that the number of H_i-1-heavy edges in H' is $O(n^{3/2}\log n)$ for both $i = 1,2$. We then Use Lemma 4.1.2 to obtain an induced subgraph H^* of H on n_1 vertices with small maximum degree. Finally, we show that the number of H_i-2-heavy edges in H^* is $O(n^{3/2}\exp(\sqrt{\log n}))$ for both $i = 1,2$. Note that any edge in H is H_i-j-heavy for some $i \in \{1,2\}$ and $j \in \{0,1,2\}$. Therefore, H^* is the desired induced subgraph.

Before giving the formal proof of the Main Lemma, we introduce two lemmas for C^3_{2k}. In these lemmas (Lemma 4.2.2, 4.2.3), an H-heavy pair is a pair of vertices in H with codegree at least $4k-2$.

Lemma 4.2.2 is due to Méroueh [56].

Lemma 4.2.2 ([56]). *Let H be a C^3_{2k}-free 3-graph. Then H can be partitioned into $2k-2$ subgraphs H_1, $H_2 \ldots$, H_{2k-2} such that H_i does not contain any H_i-3-heavy edge for all $1 \le i \le 2k-2$.*

Proof. Let $G_1 = H$, and let H_1 consist of all edges of G_1 containing a G_1-light pair, and let $G_2 = G_1 \backslash H_1$. For $i \ge 2$, let H_i consist of all edges of G_i containing a G_i-light pair, and let $G_{i+1} = G_i \backslash H_i$. Suppose for contradiction that G_{2k-1} is not empty. Let $e_1 = \{v_1, v_2, v_3\}$ be an edge in G_{2k-1}, then by definition, $\{v_2, v_3\}$ is not a G_{2k-2}-light pair, and hence, there exists an edge $e_2 = \{v_2, v_3, v_4\}$ such that $v_4 \ne v_1$. For $2 \le i \le 2k-3$, let $e_i = \{v_i, v_{i+1}, v_{i+2}\}$ be an edge in G_{2k-1-i}. By definition, $\{v_{i+1}, v_{i+2}\}$ is not a G_{2k-2-i}-light pair, and hence, there exists an edge $e_{i+1} = \{v_{i+1}, v_{i+2}, v_{i+3}\}$ in G_{2k-2-i} such that v_{i+3} is distinct from all v_j, $1 \le j \le i$. Therefore, we have a $2k$-cycle $v_1 - v_3 - v_5 - \cdots - v_{2k-1} - v_{2k} - v_{2k-2} - \cdots - v_2 - v_1$ whose edges are all H-heavy pairs, which implies the existence of a C^3_{2k}. □

Lemma 4.2.3. *Let H be a 3-graph on n-vertices. If H does not contain C^3_{2k} as subgraph, then the number of H-0-heavy edges is $O(n^{1+1/k})$.*

Proof. Let $e_0(H)$ be the number of H-0-heavy edges in H. Let J be a graph whose vertices are the set of H-0-heavy 3-edges in H and whose edges are pair of 3-edges in H whose intersection is an H-light pair. By definition of H-light pair, not hard to check that $\Delta(J) \leq 6k-12$. So we can greedily take an independent set of J of size $v(J)/(6k-11)$. This independent set correspond to a linear spanning subgraph H' of H such that $|H'| \geq e_0(H)/(6k-11)$. Note that H' is C_{2k}^3-free. So by Lemma 4.1.6 we have $|H'| = O(n^{1+1/k})$, and hence $e_0(H) = O(n^{1+1/k})$. $\square$

We introduce two more lemmas specifically for C_4^3.

Lemma 4.2.4. *Let H be a C_4^3-free 3-graph on n-vertices. There exists a graph G on the same set of vertices as H such that $|G| < n$ and the number of H-1-heavy 3-edges that does not contain any edge of G is $O(\log n \cdot n^{3/2})$.*

Proof. Let H' be the spanning subgraph of H consisting of all H-1-heavy edges. We say that an H-1-heavy edge is "small" if the codegree of its H-heavy pair in H' is at most $10^3\sqrt{n}$, otherwise it is "large". Let H'_S be the spanning subgraph of H' consisting of all "small" edge and let $H'_L = H' \backslash H'_S$.

We first show that $|H'_S| < 10^3 n^{3/2} \log_2 n$.

Assume by contradiction $|H'_S| \geq 10^3 n^{3/2} \log_2 n$. Further partition H'_S according to the codegree of the H-heavy pair in each 3-edge. Specifically, for $0 \leq i \leq \log_2(10^3\sqrt{n})$, let H'_i be the spanning subgraph of H'_S with edge set

$$E(H'_i) := \{\{u,v,w\} \in H'_S : \{u,v\} \text{ is an } H\text{-heavy pair, } 2^{i-1} < d_{H'}(u,v) \leq 2^i\},$$

and let G_i be the graph consisting of the H-heavy pairs in $\partial H'_i$, that is,

$$E(G_i) := \{\{u,v\} \in \partial H'_i : \{u,v\} \text{ is an } H\text{-heavy pair}\}.$$

Then by Pigeonhole Principle, there exists an i such that

$$|H'_i| \geq \frac{|H'_S|}{\log_2 n} \geq 10^3 n^{3/2},$$

and hence $|G_i| > |H_i'|/(10^3\sqrt{n}) \geq n$. Recall that $\beta(H)$ is the number of bowties in H. Then by Jensen's inequality and the facts that $2|G_i|/n > 2$, we have

$$\beta(H) > \sum_{v \in V(H)} \binom{d_{G_i}(v)}{2} 2^{2i-2} > n\binom{2|G_i|/n}{2} 2^{2i-2} > \frac{|G_i|^2}{n} 2^{2i-2}.$$

Note that $|G_i| \geq |H_i'|/2^i$. So we have

$$\beta(H) > \frac{|H_i'|^2}{4n} \geq 10^6 n^2/4 > 10^5 n^2.$$

This contradicts to Lemma 4.1.4.

On the other hand, let G_L be the graph consisting of the H-heavy pairs in $\partial H_L'$. It suffices to show that $|G_L| < n$. Assume for contradiction that $|G_L| \geq n$. Then by Jensen's Inequality we have

$$\beta(H) > \sum_{v \in V(H)} \binom{d_{G_L}(v)}{2} (10^3\sqrt{n})^2 > n\binom{2|G_L|/n}{2} 10^6 n > 10^6 n^2 > 10^5 n^2.$$

This contradicts to Lemma 4.1.4. □

Lemma 4.2.5. *Let H be a C_4^3-free 3-graph on n-vertices. H' is an induced subgraph of H on n_0 vertices such that all edges in H' are H-2-heavy and $\Delta(H') \leq \exp(2\sqrt{\log n}) \cdot d(H')/4$. Then the number of H-2-heavy edges in H' is $O(\exp(\sqrt{\log n})n_0^{3/2})$*

Proof. Assume for contradiction that $|H'| \geq c_1 \exp(\sqrt{\log n})n_0^{3/2}$, where the constant $c_1 > 0$ to be determined. color the vertices of H' uniformly at random with colors $\{1,2,3\}$. An edge of H' is properly colored if all of its vertices are of different colors and its H-light pair has colors $\{1,2\}$. Note that the probability that an edge is properly colored is $2/27$. Let H'' be the spanning subgraph of H' containing all properly color H-2-heavy edges. We can fix a coloring such that $|H''| \geq 2|H'|/27$.

Consider the bipartite graph G comprising all pairs of vertices of colors 1 and 2 contained in an edge of H''. Note that for each edge xy of G, the codegree of xy in H'' is exactly 1.

Otherwise, if there are distinct vertices z_1, z_2 such that they both form an 3-edge in H'' with xy, then $x-z_1-y-z_2-x$ is a 4-cycle whose edges are all H-heavy pairs. This contradicts the fact that H is C_4^3-free. So we have $|G| = |H''|$. Let f be a function that maps $xy \in e(G)$ to the vertex z such that $\{x,y,z\}$ forms a 3-edge of H''. If there is a 4-cycle $v_1 - v_2 - v_3 - v_4 - v_1$ in G, then all of its edges are mapped to the same vertex by f. That is, there is a vertex z such that $f(v_1v_2) = f(v_2v_3) = f(v_3v_4) = f(v_4v_1) = z$. Otherwise, let $z_i = f(v_iv_{i+1})$ for $1 \leq i \leq 3$ and $z_4 = f(v_4v_1)$. If $z_1 = z_3$, without lost of generality, assume $z_2 \neq z_1$, then $z_2 - v_2 - z_1 - v_3 - z_2$ is a 4-cycle whose edges are all H-heavy pairs, contradiction. So we have $z_1 \neq z_3$. Similarly, $z_2 \neq z_4$. There are 3 cases left and they all result in a C_4^3 (See Figure 4.4).

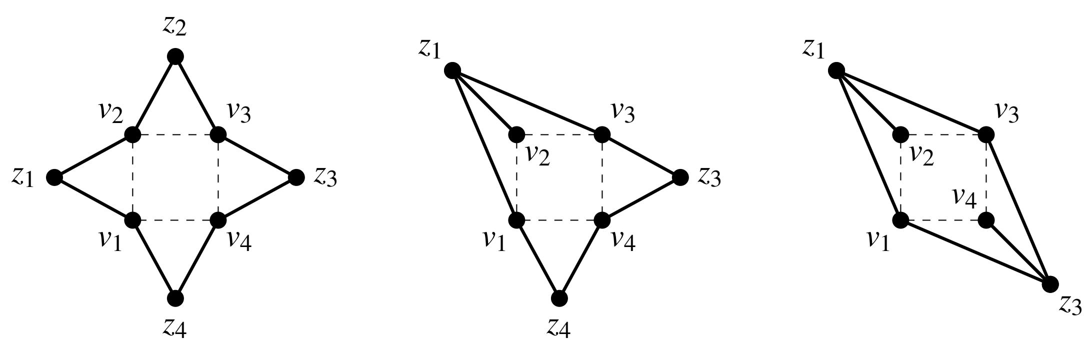

Figure 4.4. 3 ways to form a C_4^3

For convenience, let $b > 0$ be defined by $|G| = bn_0^{3/2}$. Recall that $|G| = |H''| > 2|H'|/27 > 2c_1n_0^{3/2}\exp(\sqrt{\log n})/27$, so we have $b > 2c_1\exp(\sqrt{\log n})/27$. By Lemma 4.1.7, there exist constants $\gamma, b_0 > 0$ such that if $b > b_0$, then G must contain at least $\gamma b^4 n_0^2$ copies of C_4. Note that $b > b_0$ for n large enough. By the Pigeonhole Principle, there exists an edge e such that the number of C_4 containing e in G is at least

$$\frac{4\gamma b^4 n_0^2}{|G|} = 4\gamma b^3 n_0^{1/2}.$$

Let G' be the union of all 4-cycles in G containing e. Then by Lemma 4.1.8, for some constant $c_2 > 0$, $|G'| \geq c_2 b^3 n_0^{1/2}$. There exists a vertex z such that for any edge $e \in E(G')$, $f(e) = z$. So we have

$$\Delta(H') \geq c_2 b^3 n_0^{1/2} = c_2 \frac{b^2|G|}{n_0} \geq \frac{2^3 c_1^2 c_2}{3^{10}} \exp(2\sqrt{\log n}) \cdot d'(H) > \Delta(H')$$

for some constant c_1, contradiction. Therefore, $|H'| = O((\log n)^{1/2} n_0^{3/2})$. □

Now we are ready to give the formal proof of the Main Lemma.

Proof of Lemma 4.2.1. First, by Lemma 4.2.2, we can partition H into 2 subgraphs H_1 and H_2 such that H_i does not contain any H_i-3-heavy edge for both $i = 1, 2$. By Lemma 4.2.3, we know that the number of H_i-0-heavy edges is $O(n^{3/2})$ for both $i = 1, 2$.

Then by Lemma 4.2.4 there are graphs G_1, G_2 such that $|G_1|, |G_2| < n$ and the number of H_i-1-heavy 3-edges that does not contain any edge of G_i is $O(n^{3/2} \log n)$. Using Turán's Theorem (Theorem 1.1.1), we have an induced subgraph H' of H on n_0 vertices such that $n_0 > n/5$ and $V(H')$ does not contain any edge in G_1 and G_2. So the number of H_i-1-heavy edges in H' is $O(n^{3/2} \log n)$ for both $i = 1, 2$.

We then use Lemma 4.1.2 with $t = \exp(2\sqrt{\log n})$ to get an induced subgraph H^* of H on n_1 vertices such that $n_1 > n_0^{1-1/(2\sqrt{\log n})} > (n/5)^{1-1/(2\sqrt{\log n})}$ and $\Delta(H^*) \leq \exp(2\sqrt{\log n}) \cdot d(H^*)$. Let H_i^* be the spanning subgraph of H^* consisting of all H_i-1-heavy edges in H^* for $i = 1, 2$. Note that $|H^*| \backslash |H_1^* \cup H_2^*| = O(n^{3/2} \log n)$. If $|H_1^* \cup H_2^*| < |H^*|/2$, then $|H^*| = O(n^{3/2} \log n)$. Otherwise, if $|H_1^* \cup H_2^*| \geq |H^*|/2$, assume without lost of generality that $|H_1^*| > |H_2^*|$, then by the Pigeonhole Principle,

$$\Delta(H_1^*) < \Delta(H^*) < \exp(2\sqrt{\log n}) d(H^*) < \exp(2\sqrt{\log n}) d(H_1^*)/4.$$

So by Lemma 4.2.5, $|H_1^*| = O(\exp(\sqrt{\log n}) n^{3/2})$. Since $|H_1^*| > |H^*|/4$, we have

$$|H^*| = O(\exp(\sqrt{\log n}) n^{3/2}).$$

□

Last but not least, we use the Main Lemma to prove Theorem 1.8.3.

Proof of Theorem 1.8.3. Let $n = t^{\frac{4}{3}+(1+o(1))\frac{8\sqrt{3}}{9\sqrt{\log t}}} = t^{\frac{4}{3}}\exp((8\sqrt{3}/9+o(1))\sqrt{\log t})$. Let H be a C_4^3-free graph on n vertices. By the Main Lemma (Lemma 4.2.1), there is an induced subgraph H^* of H on n_1 vertices such that $n_1 > (n/5)^{1-1/(2\sqrt{\log n})}$ and $d(H^*) = O(\exp(\sqrt{\log n})\sqrt{n})$. Then by Lemma 4.1.1 we have

$$\alpha(H) \geq \alpha(H^*) \geq \frac{2n_1}{3\sqrt{d(H^*)}} \geq n^{\frac{3}{4}-\frac{1+o(1)}{\sqrt{\log n}}} > t.$$

This completes the proof. □

Chapter 4, in full, is currently being prepared for submission for publication of the material, co-authored with Jacques Verstraëte. The dissertation author was the primary investigator and author of this paper.

Chapter 5

Triangle-free Subgraphs of Hypergraphs

Recall that the *Turán numbers* for a graph F with respect to another graph G are the quantities $\mathrm{ex}(G,F)$ denoting the maximum number of edges in an F-free subgraph of G. A *loose triangle* is a hypergraph T consisting of three edges e,f and g such that $|e\cap f| = |f\cap g| = |g\cap e| = 1$ and $e\cap f\cap g = \emptyset$. We write T^r for the loose r-uniform triangle. The Turán problem for loose triangles in r-graphs was essentially solved by Frankl and Füredi [32], who showed for each $r\geq 3$ that $\mathrm{ex}(n,T^r) = \binom{n-1}{r-1}$ for n is large enough, with equality only for the r-graph S_n^r of all r-sets containing a fixed vertex.

In this chapter, we give a general lower bound on the number of edges in a densest triangle-free subgraphs of r-graphs:

Theorem 5.0.1 (Theorem 1.10.2)**.** *Let $r\geq 3$ and let G be an r-graph with maximum degree $\triangle$. Then as $\triangle\to\infty$,*

$$\mathrm{ex}(G,T^r) \geq \triangle^{-\frac{r-2}{r-1}-o(1)}e(G).$$

For $r\geq 4$, the best construction we have gives the following proposition:

Proposition 5.0.2 (Proposition 1.10.3)**.** *For $r\geq 4$ there exists an r-graph G with maximum degree $\triangle$ such that as $\triangle\to\infty$,*

$$\mathrm{ex}(G,T^r) = O(\triangle^{-\frac{1}{2}})\cdot e(G).$$

Let $G^r_{n,p}$ denote random r-graph where edges of K^r_n are sampled independently with probability p. The second main result in this chapter is as follows:

Theorem 5.0.3 (Theorem 1.11.1)**.** *For all $n \geq 2$ and $p = p(n) \leq 1$ with $pn^3 \to \infty$ as $n \to \infty$, there exists a constant $c > 0$ such that asymptotically almost surely*

$$\min\{(1-o(1))p\binom{n}{3}, p^{\frac{1}{3}}n^2e^{-c\sqrt{\log n}}\} \leq \mathrm{ex}(G^3_{n,p}, T^3) \leq \min\{(1+o(1))p\binom{n}{3}, p^{\frac{1}{3}}n^{2+o(1)}\},$$

and more precisely, for any constant $\delta > 0$, when $n^{-3/2+\delta} \leq p \leq n^{-\delta}$, we have

$$\mathrm{ex}(G^3_{n,p}, T^3) \leq p^{\frac{1}{3}}n^2(\log n)^c.$$

In this chapter, we adapt the method of Balogh, Narayanan and Skokan [6] to count triangle-free hypergraphs with a specified number of edges. Let $N(r,m)$ denote the number of T^r-free r-graphs with n vertices and m edges.

Theorem 5.0.4 (Theorem 1.11.2)**.** *Let $n \geq 2$, $\varepsilon(n)$ be a function such that $\frac{\varepsilon(n)\log n}{\log\log n} \to \infty$ as $n \to \infty$. Let $\delta = \delta(n)$ be a function such that $\varepsilon(n) < \delta < 1/2 - \varepsilon(n)$ and let $m = n^{2-\delta}$. Then*

$$N(3,m) \leq \left(\frac{n^2}{m}\right)^{3m+o(m)}.$$

The upper bound on $\mathrm{ex}(G^3_{n,p}, T^3)$ in Theorem 1.11.1 will follow from the bound on $N(3,m)$ in Theorem 1.11.2 by taking $m = p^{1/3-o(1)}n^2$, see details in Section 5.3.

5.1 Deterministic host

In this section we prove Theorem 1.10.2 and Proposition 1.10.3

For graphs, Foucaud, Krivelevich and Perarnau [31] used certain random homomorphisms to obtain good lower bounds on $\mathrm{ex}(G,F)$. We briefly summarize these ideas. Let $\mathscr{M}(F)$ denote the family of graphs F' such that there exists a graph homomorphism $\phi : V(F) \to V(F')$ and such that ϕ induces a bijection from $E(F)$ to $E(F')$. Let H be an $\mathscr{M}(F)$-free graph with many edges, which we will use as a template for our subgraph of G. Specifically, we take a random mapping $\chi : V(G) \to V(H)$ and then constructs a subgraph $G' \subseteq G$ such that $uv \in E(G')$ if and only if $\chi(u)\chi(v) \in E(H)$ and such that $\chi(u)\chi(v) \neq \chi(u)\chi(w)$ for any other edge $uw \in E(G)$ (that is, we do not keep edges which are incident and map to the same edge). It is then proven in [31] that G' will be F-free because H is $\mathscr{M}(F)$-free, and that in expectation G' will have many edges provided H does.

For general r-graphs, it is not immediately clear how to extend these ideas in such a way that we can both construct a subgraph with many edges and such that the subgraph is F-free. Fortunately for T^r we are able to do this. In particular, for this case it turns out we can avoid a hypergraph analog of the family $\mathscr{M}(F)$ provided our template r-graph is linear. This is where the Ruzsa-Szemeredi construction of Theorem 1.10.1 plays its crucial role.

Proof of Theorem 1.10.2. Let t be an integer to be determined later. Let χ be a random map from $V(G)$ to $[t]$ and G_t be the r-graph on $[t]$ from Theorem 1.10.1. For ease of notation define $\chi(e) = \{\chi(v_1), \ldots, \chi(v_r)\}$ when $e = \{v_1, \ldots, v_r\}$. Let G' be the subgraph of G which contains the edge e if and only if

(1) $\chi(e)$ is an edge of G_t, and

(2) $\chi(e') \not\subset \chi(e)$ for any $e' \in E(G)$ with $|e \cap e'| = 1$.

We claim that G' is T^r-free. Indeed, let T be a T^r of G', say with edges e_1, e_2, e_3 and $e_i \cap e_j = \{x_{ij}\}$ for $i \neq j$. Because G_t is linear, if e, e' are (possibly non-distinct) edges of G_t, then $|e \cap e'|$ is either 0, 1, or r. Note that $\chi(e_i), \chi(e_j)$ are edges of G_t by (1). Because $e_i \cap e_j = \{x_{ij}\}$ for $i \neq j$, $\chi(x_{ij}) \in \chi(e_i) \cap \chi(e_j)$, and by (2) the size of this intersection is strictly less than r. Thus $\chi(e_i) \cap \chi(e_j) = \{\chi(x_{ij})\}$. Further, we must have, say, $\chi(x_{ij}) \neq \chi(x_{ik})$ for $k \neq i, j$. This is because (1) guarantees that $\chi(x)$ is a distinct element for each $x \in e_i$, so in particular this holds for $x_{ij}, x_{ik} \in e_i$. In total this implies $\chi(e_1), \chi(e_2), \chi(e_3)$ forms a T^r in G_t, a contradiction.

We wish to compute how large $e(G')$ is in expectation. Fix some $e \in E(G)$. The probability that e satisfies (1) is exactly $e(G_t)r!/t^r$. Let $\{e_1, \ldots, e_d\}$ be the edges in $E(G)$ with $|e_i \cap e| = 1$. Given that e satisfies (1), the probability that $\chi(e_1) \not\subset \chi(e)$ is exactly $1-(r/t)^{r-1}$. Note that for any $v \notin e \cup e_1$, the event $\chi(v) \in \chi(e)$ is independent of the event $\chi(e_1) \not\subset \chi(e)$, so we have

$$\Pr[\chi(v) \in \chi(e)|\ e \text{ satisfies (1)},\ \chi(e_1) \not\subset \chi(e)] = \frac{r}{t}.$$

On the other hand, if $v \in e_1 \setminus e$, then

$$\Pr[\chi(v) \in \chi(e)|\ e \text{ satisfies (1)},\ \chi(e_1) \not\subset \chi(e)] < \frac{r}{t},$$

as knowing some subset containing $\chi(v)$ is not contained in $\chi(e)$ makes it less likely that $\chi(v) \in \chi(e)$. By applying these observations to each vertex of $e_2 \setminus e$, we conclude that

$$\Pr[\chi(e_2) \not\subset \chi(e)|\ e \text{ satisfies (1)},\ \chi(e_1) \not\subset \chi(e)] \geq 1 - \left(\frac{r}{t}\right)^{r-1}.$$

By repeating this logic for each e_i, and using that $e(G_t) = t^{2-o(1)}$, we conclude that

$$\Pr[e \text{ satisfies (1), (2)}] \geq \frac{e(G_t)r!}{t^r}\left(1-\left(\frac{r}{t}\right)^{r-1}\right)^{r\triangle} = t^{2-r-o(1)}\left(1-\left(\frac{r}{t}\right)^{r-1}\right)^{r\triangle}.$$

By taking $t = r(r\triangle)^{1/(r-1)}$ and using that $(1-x^{-1})^x$ is a decreasing function in x, we conclude by linearity of expectation that

$$\mathbb{E}[e(G')] \geq \triangle^{-1+\frac{1}{r-1}-o(1)} \cdot e(G).$$

In particular, there exists some T^r-free subgraph of G with at least this many edges, giving the desired result. □

We close this section with a proof of Proposition 1.10.3.

Proof of Proposition 1.10.3. According to Rödl and Thoma [63], there exists an r-graph G with $\Theta(n^3)$ edges such that every three vertices is contained in at most one edge. Let G' be a T^r-free subgraph of G. Define G'' by deleting every edge of G' which contains two vertices that are contained in at most $2r$ edges. Note that $e(G') - e(G'') \leq 2r\binom{n}{2}$.

Assume G'' contains an edge $e = \{v_1, \ldots, v_r\}$. Because v_1, v_2 are contained in an edge of G'', there exist a set $E_{12} \subseteq E(G')$ of at least $2r+1$ many edges containing v_1 and v_2. As G contained at most one edge containing v_1, v_2, and v_3, any $e_{12} \neq e$ in E_{12} does not contain v_3. Fix such e_{12}. Because v_2, v_3 are contained in an edge of G'', there exists a set $E_{23} \subseteq E(G')$ of at least $2r+1 \geq r-1$ edges containing v_2, v_3. Because G contains at most one edge containing v_2, v_3, u_i for any $u_i \in e_{12} \backslash \{v_2\}$, we conclude that there exists some $e_{23} \in E_{23}$ such that $e_{12} \cap e_{23} = \{v_2\}$. Fix such e_{23}. Because v_1, v_3 are contained in an edge of G'', there exists a set $E_{13} \subseteq E(G')$ of at least $2r+1 \geq 2r-3$ edges containing v_1, v_3. Because G contains at most one edge containing v_1, v_3, u_i for any $u_i \in e_{12} \cup e_{23} \backslash \{v_1, v_3\}$, we conclude that there exists some $e_{13} \in E_{13}$ such that $e_{13} \cap e_{12} = \{v_1\}$ and $e_{13} \cap e_{23} = \{v_3\}$. These three edges form a T^r in G', a contradiction. We conclude that G'' contains no edges, and hence $e(G') \leq 2r\binom{n}{2}$ for any T^r-free subgraph G' of G. As G has maximum degree $\triangle = \Theta(n^2)$, we conclude that $\mathrm{ex}(G, T^r) = O(n^2) = O(\triangle^{-1/2}) \cdot e(G)$. □

We note that one can replace the G used in the above proof with an appropriate Steiner system to obtain a regular graph which serves as an upper bound. It has recently been proven by Keevash [41] and Glock, Kühn, Lo, and Osthus [37] that such Steiner systems exist whenever n satisfies certain divisibility conditions and is sufficiently large.

5.2 Random host: lower bound.

In this section we prove the lower bound in Theorem 1.11.1. As noted in the introduction, the bound of Theorem 1.10.2 is sharp for $r = 3$ by considering the disjoint union of cliques, so we cannot improve upon this bound in general. However, we are able to do better when G

contains few copies of T^r by using a deletion argument.

Proposition 5.2.1. *Let $R(G)$ denote the number of copies of T^r in the r-graph G. Then for some positive constant c and any integer $t \geq 1$,*

$$\mathrm{ex}(G, T^r) \geq (e(G)t^{2-r} - R(G)r^{3r}t^{5-3r})e^{-c\sqrt{\log t}}.$$

Proof. Let χ be a random map from $V(G)$ to $[t]$ and G_t the r-graph on $[t]$ from Theorem 1.10.1. For ease of notation, if $e = \{v_1, \ldots, v_r\}$ we define $\chi(e) := \{\chi(v_1), \ldots, \chi(v_r)\}$. Let G' be the subgraph of G which contains the edge e if and only if $\chi(e)$ is an edge of G_t.

We claim that $e_1, e_2, e_3 \in E(G')$ form a T^r in G' if and only if e_1, e_2, e_3 form a T^r in G and $\chi(e_1) = \chi(e_2) = \chi(e_3)$ is an edge of G_t. Indeed, the backwards direction is clear. Assume for contradiction that these edges form a T^r in G' and that $\chi(e_1) \neq \chi(e_2)$. Let x_{ij} for $i \neq j$ be such that $e_i \cap e_j = \{x_{ij}\}$. Because G_t is linear, if e, e' are (possibly non-distinct) edges of G_t, then $|e \cap e'|$ is either 0, 1, or r. Because each e_i is in $E(G')$, we have $\chi(e_i) \in E(G_t)$ by construction. In particular, as $e_1 \cap e_2 = \{x_{12}\}$ and $\chi(e_1) \neq \chi(e_2)$, we must have $\chi(e_1) \cap \chi(e_2) = \{\chi(x_{12})\}$. As e_3 contains an element in e_1 (namely x_{13}) and an element not in e_1 (namely x_{23}), we must have $\chi(e_1) \cap \chi(e_3) = \{\chi(x_{13})\}$. Similarly we have $\chi(e_2) \cap \chi(e_3) = \{\chi(x_{23})\}$. Because $\chi(e_i)$ is an r-set for each i, we have $\chi(x_{ij}) \neq \chi(x_{ik})$ for $\{i, j, k\} = \{1, 2, 3\}$. Thus $\chi(e_1), \chi(e_2), \chi(e_3)$ form a T^r in G_t, a contradiction.

The probability that a given $T^r \subseteq G$ maps onto a given $f \in E(G_t)$ is at most $(r/t)^{3r-3}$, since in particular each vertex of T^r must map onto one of the r vertices of f. By the claim above this is the only way that a T^r can appear in G', so by linearity of expectation we find

$$\mathbb{E}[R(G')] \leq \frac{e(G_t)r^{3r-3}}{t^{3r-3}} R(G).$$

Let $G'' \subseteq G'$ be a subgraph obtained by deleting an edge from each T^r of G'. By construction G'' is T^r-free. Since $e(G_t) \geq t^2 e^{-c\sqrt{\log t}}$ for some positive constant c, we conclude by linearity of expectation that

$$\begin{aligned}\mathrm{ex}(G,T^r) &\geq \mathbb{E}[e(G'')] \geq \mathbb{E}[e(G') - R(G')] \\ &= \frac{e(G_t)r!}{t^r}e(G) - \frac{e(G_t)r^{3r-3}}{t^{3r-3}}R(G) \\ &\geq (e(G)t^{2-r} - R(G)r^{3r}t^{5-3r})e^{-c_2\sqrt{\log t}}.\end{aligned}$$

□

Corollary 5.2.2. *For any integer $r \geq 3$, and function $p = p(n) \leq 1$ such that $p^{2/(2r-3)}n \geq 2$, we have*

$$\mathbb{E}[\mathrm{ex}(G^r_{n,p}, T^r)] \geq p^{\frac{1}{2r-3}}n^2 e^{-c\sqrt{\log n}},$$

for some constant $c > 0$.

Proof. Note for $n \geq 4$ that $\mathbb{E}[e(G^r_{n,p})] = p\binom{n}{r} \geq pn^r r^{-r}$, and that $\mathbb{E}[R(G^r_{n,p})] \leq p^3 n^{3r-3}$. Plugging these into the bound of Proposition 5.2.1 gives for some positive constant c_1

$$\mathbb{E}[\mathrm{ex}(G^r_{n,p}, T^r)] \geq (pn^r r^{-r} t^{2-r} - p^3 n^{3r-3} r^{3r} t^{5-3r})e^{-c_1\sqrt{\log t}}.$$

Taking $t = (2r^{4r})^{\frac{1}{2r-3}} p^{2/(2r-3)}n$, we conclude for some positive constant c_2 and sufficiently large n that

$$\mathbb{E}[\mathrm{ex}(G^r_{n,p}, T^r)] \geq p^{\frac{1}{2r-3}}n^2 e^{-c_2\sqrt{\log n}}.$$

□

To get the a.a.s. result of Theorem 1.11.1, we use Azuma's inequality (See for example in Alon and Spencer [3]) applied to the edge exposure martingale.

Lemma 5.2.3. *Let f be a function on r-graphs such that $|f(G) - f(H)| \leq 1$ whenever H is* *obtained from G by adding or deleting one edge. Then for any $\lambda > 0$,*

$$\Pr\left[|f(G^r_{n,p}) - \mathbb{E}[f(G^r_{n,p})]| > \lambda\sqrt{\binom{n}{r}}\right] < e^{-\frac{\lambda^2}{2}}.$$

Proof of Theorem 1.11.1(Lower Bounds). Let $\varepsilon(n) = e^{k\sqrt{\log n}}$, where $k > 0$ is some large enough constant. For $p \le n^{-3/2}/\varepsilon(n)$, it is not difficult to show that a.a.s. $G^3_{n,p}$ contains $o(pn^3)$ copies of T^3, and by deleting an edge from each of these loose cycles we see that $\mathrm{ex}(G^3_{n,p}, T^3) \ge (1 - o(1)p\binom{n}{3}$ a.a.s..

For $n^{-3/2}/\varepsilon(n) \le p \le n^{-3/2}\varepsilon(n)$, we do an extra round of random sampling on the edges of $G^r_{n,p}$ and keep each edge with probability $p' := \varepsilon(n)^{-2}$. The r-graph we obtained is equivalent to $G^r_{n,pp'}$, with $pp' \le n^{-3/2}/\varepsilon(n)$. Thus $\mathrm{ex}(G^3_{n,p}, T^3) \ge (1-o(1))pp'\binom{n}{3} = (1-o(1))p\binom{n}{3}/\varepsilon(n)^2$ a.a.s.. Using $p \ge n^{-3/2}/\varepsilon(n)$, we conclude that $\mathrm{ex}(G^3_{n,p}, T^3) \ge p^{1/3}n^2e^{-3k\sqrt{\log n}}$ a.a.s. in this range.

We now consider $p \ge n^{-3/2}\varepsilon(n)$. The bound in expectation follows from Corollary 5.2.2. To show that this result holds a.a.s., we observe that $f(G) = \mathrm{ex}(G, T^3)$ satisfies the conditions of Lemma 5.2.3. For ease of notation let $X_{n,p} = \mathrm{ex}(G^3_{n,p}, T^3)$ and let $B_{n,p} = p^{1/3}n^2e^{-c_2\sqrt{\log n}}$ be the lower bound for $\mathbb{E}[X_{n,p}]$ given in Corollary 5.2.2. Setting $\lambda = \frac{1}{2}B_{n,p}\binom{n}{3}^{-1/2}$ and applying Azuma's inequality, we find

$$\begin{aligned}\Pr\left[X_{n,p} < \frac{1}{2}B_{n,p}\right] &\le \Pr\left[X_{n,p} - \mathbb{E}[X_{n,p}] < -\lambda\binom{n}{3}^{\frac{1}{2}}\right] \\ &\le \Pr\left[\left|X_{n,p} - \mathbb{E}[X_{n,p}]\right| > \lambda\binom{n}{3}^{\frac{1}{2}}\right] \le \exp(-\frac{\lambda^2}{2}).\end{aligned}$$

Note that for $p \ge n^{-3/2}\varepsilon(n)$ we have $\lambda \ge e^{(k/3-c_2)\sqrt{\log n}} \to \infty$ as $n \to \infty$. So we conclude the a.a.s. result. □

5.3 Counting triangle-free hypergraphs with containers

In this section we use the method containers to prove Theorem 1.11.2. The method of containers developed by Balogh, Morris and Samotij [5] and Saxton and Thomason [68] is a powerful technique that has been used to solve a number of combinatorial problems. Roughly, the idea is for a suitable hypergraph H to find a family of sets $\mathscr{C}$ which contain every independent set of H, and in such a way that $|\mathscr{C}|$ is small and each $C \in \mathscr{C}$ contains few edges. For example, by letting H be the 3-uniform hypergraph where each edge is a K_3 in some graph G, we see that independent sets of H correspond to triangle-free subgraphs of G. The existence of containers then allows us to better understand how these subgraphs of G behave.

We proceed with the technical details of this approach. Given an r-graph $H = (V,E)$, let $v(H) = |V|$, $e(H) = |E|$, and let $\mathscr{P}(V)$ be the family of all subsets of V. For a set A of vertices in H, let $d(A)$ be the number of edges in H that contain A. Let $\bar{d}(H)$ be the average vertex degree of H, and let $\triangle_j(H) = \max_{|A|=j} d(A)$. In order to establish our upper bounds, we need to use the following container lemma for hypergraphs:

Lemma 5.3.1 (Balogh, Morris and Samotij [5])**.** *Let $r,b,l \in \mathbb{N}$, $\delta = 2^{-r(r+1)}$, and $H = (V,E)$ an r-graph such that*

$$\triangle_j(H) \le \left(\frac{b}{v(H)}\right)^{j-1} \frac{e(H)}{l},\ j \in \{1,2,\ldots,r\}.$$

Then there exists a collection $\mathscr{C}$ of subsets of V such that:

(1) *For every independent set I of H, there exists $C \in \mathscr{C}$ such that $I \subset C$.*

(2) *For every $C \in \mathscr{C}$, $|C| \le v(H) - \delta l$.*

(3) $|\mathscr{C}| \le \sum_{s=0}^{(k-1)b} \binom{v(H)}{s}$.

We will use this container lemma to give an upper bound for $N(r,m)$. The idea is to consider the 3-graph H with $V(H) = E(K_n^r)$ and $E(H)$ consisting of T^r in K_n^r. Notice that the container lemma requires upper bounds for the maximum codegrees of the hypergraph. In order to meet this requirement, we will use a balanced-supersaturation lemma for T^r:

Lemma 5.3.2 (Balogh, Narayanan and Skokan [6])**.** *For any integer $r \geq 3$, there exists $c = c(r)$ such that the following hold for all $n \in \mathbb{N}$. Given any r-graph G on $[n]$ with $e(G) = tn^{r-1}$, $t \geq 6(r-1)$, let $S = tn^{r-4}$ if $r \geq 4$ and $S = 1$ if $r = 3$. Then there exists a 3-graph H on $E(G)$, where each edge of H is a copy of T^r in G, such that:*

(1) $\bar{d}(H) \geq c^{-1}t^3S^2$.

(2) $\triangle_j(H) \leq ct^{5-2j}S^{3-j}$ *for each* $j = 1, 2$.

Using the previous two lemmas, we derive the following container lemma for T^3-free hypergraphs. Similar result for T^r-free hypergraphs can also be obtained using the same idea, and we omit the details.

Lemma 5.3.3. *Let $c_1 = c(3)$ be the constant obtained in Lemma 5.3.2 with $r = 3$. For any integer n and positive number t with $\max(12, c_1) \leq t \leq \binom{n}{3}/n^2$, there exists a collection $\mathscr{C}$ of subgraphs of K_n^3 such that for some constant c_2:*

(1) *For any T^3-free subgraph J of K_n^3, there exists $C \in \mathscr{C}$ such that $J \subset C$.*

(2) $|\mathscr{C}| \leq \exp\left(\frac{c_2(\log t)n^2}{\sqrt{t}}\right)$.

(3) *For every $C \in \mathscr{C}$, $e(C) \leq tn^2$.*

Proof. By Lemma 5.3.2 with $r = 3$, there exists a positive constant c_1 such that for any 3-graph G on $[n]$ with $e(G) = t_0n^2$, where $t_0 \geq t$, there exists a 3-graph H on $E(G)$ such that each edge of H is a copy of T^3 in G, such that:

(1) $\bar{d}(H) \geq c_1^{-1}t_0^3$.

(2) $\triangle_j(H) \leq c_1t_0^{5-2j}, j = 1, 2$. $\triangle_3(H) = 1$.

We can then use Lemma 5.3.1 on H with $l = t_0 n^2/(3c_1^2)$ and $b = n^2/\sqrt{c_1 t_0}$ to get a collection $\mathscr{C}$ of subgraphs of G such that they contain all T^3-free subgraphs of G, and for each $C \in \mathscr{C}$, $e(C) \leq (1-\varepsilon)t_0 n^2$ for some constant $\varepsilon > 0$. Also, we have

$$|\mathscr{C}| \leq \sum_{s=0}^{2b} \binom{t_0 n^2}{s} \leq \exp\left(\frac{c_2 (\log t_0) n^2}{\sqrt{t_0}}\right)$$

for some constant $c_2 > 0$.

We use the above argument on $G = K_n^3$ to get a family of containers $\mathscr{C}_1$. Notice that the containers of $\mathscr{C}_1$ are also 3-graph on $[n]$, so we can repeat this argument on each $C \in \mathscr{C}_1$ with more than tn^2 edges to get a new collection of containers $\mathscr{C}_2$. We do this repeatedly until all containers have less than tn^2 edges. Since in each step the number of edges will decrease by a constant $(1-\varepsilon)$, this process must stop after at most $\log(n/t)/\varepsilon$ steps. Let $t_0 = \binom{n}{3}/n^2$, $t_{k+1} = (1-\varepsilon)t_k$ for $k \geq 0$, and let M be the largest integer such that $t_M > t$. By definition of t_k, we have $t_{M-i} > (1-\varepsilon)^{-i} t$, and hence, there exists a constant $\delta = \delta(\varepsilon) > 0$ such that $\frac{\log t_{M-i}}{\sqrt{t_{M-i}}} < (1-\delta)^i \frac{\log t}{\sqrt{t}}$.

Then in the worst case, the number of containers we have in the end is less than

$$\begin{aligned}\prod_{i=0}^{M} \exp\left(\frac{c_2 (\log t_i) n^2}{\sqrt{t_i}}\right) &= \exp\left(\sum_{i=0}^{M} \frac{c_2 (\log t_i) n^2}{\sqrt{t_i}}\right) \\ &< \exp\left(c_2 n^2 \sum_{i=0}^{M} (1-\delta)^i \frac{\log t}{\sqrt{t}}\right) \\ &< \exp\left(\frac{c_2 (\log t) n^2}{\delta \sqrt{t}}\right).\end{aligned}$$

This completes the proof. $\square$

With the lemma above, we are ready to prove Theorem 1.11.2. The proof of Theorem 1.11.3 is essentially the same and we omit the details.

Proof of Theorem 1.11.2. Let $\mathscr{C}$ be a collection of containers and c a constant as in Lemma 5.3.3 with $t = n^{2\delta + \varepsilon_1(n)}$, where $\varepsilon_1(n) = \frac{2 \log\log n}{\log n}$. Since $\varepsilon(n) < \delta < 1/2 - \varepsilon(n)$, and $\varepsilon(n) = \omega(\frac{\log n}{\log\log n})$,

t must satisfies the condition in Lemma 5.3.3. By considering all subgraphs of each $C \in \mathscr{C}$ with $m = n^{2-\delta}$ edges, we conclude that

$$\begin{aligned} N(3,m) &\le \exp\left(\frac{c(\log t)n^2}{\sqrt{t}}\right) \cdot \binom{tn^2}{m} \\ &\le \exp\left(c\log t \cdot \frac{m}{\log n} + (1 + (3\delta + \varepsilon_1(n))\log n)\,m\right) \\ &\le \exp\left(\delta \log n \cdot m\left(3 + (2+o(1))\frac{\log\log n}{\delta \log n}\right)\right) \\ &\le \left(\frac{n^2}{m}\right)^{\left(3+(2+o(1))\frac{\log\log n}{\delta\log n}\right)m}. \end{aligned}$$

Since $\delta > \varepsilon(n) = \omega(\frac{\log n}{\log\log n})$, we have

$$N(3,m) = \left(\frac{n^2}{m}\right)^{3m+o(m)}.$$

□

5.4 Random host: upper bound

In this section we prove the upper bound of Theorem 1.11.1.

Proof of Theorem 1.11.1(Upper Bound). We will only present the proof of

$$\mathrm{ex}(G^3_{n,p}, T^3) \le \min\{(1+o(1))p\binom{n}{3}, p^{\frac{1}{3}}n^{2+o(1)}\}$$

for $p \le 1$. The proof of the more accurate upper bound in smaller range is essentially the same, with more careful and explicit computation for the $o(1)$ factor. For $p \le n^{-3/2+o(1)}$, the proof for upper bound is exactly the same as that for lower bound when $p \le n^{-3/2}\varepsilon(n)$. We now consider $n^{-3/2+\varepsilon(n)} \le p \le n^{-\varepsilon(n)}$ for some small function $\varepsilon(n) = o(1)$ to be determined. Our goal is to show

$$\Pr[ex(G^3_{n,p}, T^3) \ge m] \to 0, \text{ as } n \to \infty,$$

for some $m = p^{1/3}n^{2+o(1)}$. Let X_m be the expected number of T^3-free subgraphs in $G^3_{n,p}$ with m edges. By Theorem 1.11.2, when $n^{3/2+\varepsilon_1(n)} \leq m \leq n^{2-\varepsilon_1(n)}$ for some function $\varepsilon_1(n) = o(1)$, there exists a function $\varepsilon_2(n) = o(1)$ such that the expectation of X_m satisfies

$$\begin{aligned}\mathbb{E}[X_m] &= N(3,m)\cdot p^m \\ &\leq \left(\frac{n^2}{m}\right)^{m(3+\varepsilon_2(n))} p^m \\ &= \left(\left(\frac{n^2}{m}\right)^{(3+\varepsilon_2(n))} p\right)^m.\end{aligned}$$

We can let $m = n^2p^{1/3-\varepsilon_3(n)}$ for some small function $\varepsilon_3(n) = o(1)$ such that

$$\left(\frac{n^2}{m}\right)^{(3+\varepsilon_2(n))} p < 1.$$

Also we can pick some suitable $\varepsilon(n)$, so that $n^{3/2+\varepsilon_1(n)} \leq m \leq n^{2-\varepsilon_1(n)}$. Thus we have $\mathbb{E}[X_m] \to 0$ as $n \to \infty$. Then by Markov's inequality, we have

$$\Pr[ex(G^3_{n,p}, T^3) \geq m] = \Pr[X_m \geq 1] \leq \mathbb{E}[X_m] \to 0, \ as\ n \to \infty.$$

So a.a.s. we have

$$ex(G^3_{n,p}, T^3) < m = p^{\frac{1}{3}}n^{2+o(1)}.$$

Finally for $p \geq n^{-o(1)}$, we have $\mathrm{ex}(G^3_{n,p}, T) < \mathrm{ex}(K^3_n, T) = \Theta(n^2) = p^{1/3}n^{2+o(1)}$ a.a.s. □

Chapter 5, in full, is a version of the material as it appears in "Triangle-free Subgraphs of Hypergraphs", Graphs and Combinatorics 37.6 (2021): 2555-2570, co-authored with Sam Spiro, Jacques Verstraëte. The dissertation author was the primary investigator and author of this paper.

Bibliography

[1] M. Ajtai, J. Komlós, J. Pintz, J. Spencer, and E. Szemerédi. Extremal uncrowded hypergraphs. *Journal of Combinatorial Theory, Series A*, 32(3):321–335, 1982.

[2] M. Ajtai, J. Komlós, and E. Szemerédi. A note on ramsey numbers. *Journal of Combinatorial Theory, Series A*, 29(3):354–360, 1980.

[3] N. Alon and J. H. Spencer. *The probabilistic method*. John Wiley & Sons, 2004.

[4] J. Balogh, T. Bohman, and D. Mubayi. Erdős–ko–rado in random hypergraphs. *Combinatorics, Probability and Computing*, 18(5):629–646, 2009.

[5] J. Balogh, R. Morris, and W. Samotij. The method of hypergraph containers. In *Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018*, pages 3059–3092. World Scientific, 2018.

[6] J. Balogh, B. Narayanan, and J. Skokan. The number of hypergraphs without linear cycles. *Journal of Combinatorial Theory, Series B*, 134:309–321, 2019.

[7] J. Balogh and W. Samotij. The number of $k_{s,t}$-free graphs. *Journal of the London Mathematical Society*, 83(2):368–388, 2011.

[8] F. Behrend. On sequences of numbers not divisible one by another. *Journal of the London Mathematical Society*, 1(1):42–44, 1935.

[9] C. T. Benson. Minimal regular graphs of girths eight and twelve. *Canadian Journal of Mathematics*, 18:1091–1094, 1966.

[10] T. Bohman and P. Keevash. The early evolution of the h-free process. *Inventiones mathematicae*, 181(2):291–336, 2010.

[11] T. Bohman and P. Keevash. Dynamic concentration of the triangle-free process. *Random Structures & Algorithms*, 58(2):221–293, 2021.

[12] B. Bollobás. The independence ratio of regular graphs. *Proceedings of the American Mathematical Society*, pages 433–436, 1981.

[13] B. Bollobás and E. Győri. Pentagons vs. triangles. *Discrete Mathematics*, 308(19):4332–4336, 2008.

[14] Y. Caro. New results on the independence number. Technical report, Technical Report, Tel-Aviv University, 1979.

[15] Y. Caro, Y. Li, C. C. Rousseau, and Y. Zhang. Asymptotic bounds for some bipartite graph: complete graph ramsey numbers. *Discrete Mathematics*, 220(1):51–56, 2000.

[16] Y. Caro and Z. Tuza. Improved lower bounds on k-independence. *Journal of Graph Theory*, 15(1):99–107, 1991.

[17] F. Chung. Open problems of paul erdos in graph theory. *Journal of Graph Theory*, 25(1):3–36, 1997.

[18] F. Chung and L. Lu. An upper bound for the turán number $t_3(n,4)$. *Journal of Combinatorial Theory, Series A*, 87(2):381–389, 1999.

[19] C. Collier-Cartaino, N. Graber, and T. Jiang. Linear turán numbers of linear cycles and cycle-complete ramsey numbers. *Combinatorics, Probability and Computing*, 27(3):358–386, 2018.

[20] D. Conlon and W. T. Gowers. Combinatorial theorems in sparse random sets. *Annals of Mathematics*, 184(2):367–454, 2016.

[21] S. Das, C. Lee, and B. Sudakov. Rainbow turán problem for even cycles. *European Journal of Combinatorics*, 34(5):905–915, 2013.

[22] B. Demarco, A. Hamm, and J. Kahn. On the triangle space of a random graph. *Journal of Combinatorics*, 4(2):229–249, 2013.

[23] B. DeMarco and J. Kahn. Mantel's theorem for random graphs. *Random Structures & Algorithms*, 47(1):59–72, 2015.

[24] R. A. Duke, H. Lefmann, and V. Rödl. On uncrowded hypergraphs. *Random Structures & Algorithms*, 6(2-3):209–212, 1995.

[25] K. Dutta, D. Mubayi, and C. Subramanian. New lower bounds for the independence number of sparse graphs and hypergraphs. *SIAM Journal on Discrete Mathematics*, 26(3):1134–1147, 2012.

[26] P. Erdős. On the combinatorial problems which i would most like to see solved. *Combinatorica*, 1(1):25–42, 1981.

[27] P. Erdös, R. J. Faudree, C. C. Rousseau, and R. H. Schelp. On cycle—complete graph ramsey numbers. *Journal of Graph Theory*, 2(1):53–64, 1978.

[28] P. Erdös, P. Frankl, and V. Rödl. The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent. *Graphs and Combinatorics*, 2(1):113–121, 1986.

[29] P. ERDÓS, C. KO, and R. RADO. INTERSECTION THEOREMS FOR SYSTEMS OF FINITE SETS. *The Quarterly Journal of Mathematics*, 12(1):313–320, 01 1961.

[30] P. Erdős and M. Simonovits. Compactness results in extremal graph theory. *Combinatorica*, 2(3), 1982-9.

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

[32] P. Frankl and Z. Füredi. Exact solution of some turán-type problems. *Journal of Combinatorial Theory, Series A*, 45(2):226–262, 1987.

[33] E. Friedgut, V. Rödl, and M. Schacht. Ramsey properties of random discrete structures. *Random Structures & Algorithms*, 37(4):407–436, 2010.

[34] Z. Füredi and T. Jiang. Hypergraph turán numbers of linear cycles. *Journal of Combinatorial Theory, Series A*, 123(1):252–270, 2014.

[35] Z. Füredi and M. Simonovits. The history of degenerate (bipartite) extremal graph problems. In *Erdős Centennial*, pages 169–264. Springer, 2013.

[36] D. Gamarnik and D. A. Goldberg. Randomized greedy algorithms for independent sets and matchings in regular graphs: Exact results and finite girth corrections. *Combinatorics, Probability and Computing*, 19(1):61–85, 2010.

[37] S. Glock, D. Kühn, A. Lo, and D. Osthus. The existence of designs via iterative absorption: hypergraph f-designs for arbitrary f. *arXiv preprint arXiv:1611.06827*, 2016.

[38] E. Győri and N. Lemons. Hypergraphs with no odd cycle of given length. *Electronic Notes in Discrete Mathematics*, 34:359–362, 2009.

[39] E. Győri and N. Lemons. 3-uniform hypergraphs avoiding a given odd cycle. *Combinatorica*, 32(2):187–203, 2012.

[40] O. Janzer. The rainbow turán number of even cycles. In *Extended Abstracts EuroComb 2021*, pages 599–604. Springer, 2021.

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

[42] P. Keevash, D. Mubayi, B. Sudakov, and J. Verstraëte. Rainbow turán problems. *Combinatorics, Probability and Computing*, 16(1):109–126, 2007.

[43] J. H. Kim. The ramsey number r(3, t) has order of magnitude t2/log t. *Random Structures & Algorithms*, 7(3):173–207, 1995.

[44] Y. Kohayakawa, B. Kreuter, and A. Steger. An extremal problem for random graphs and the number of graphs with large even-girth. *Combinatorica*, 18(1):101–120, 1998.

[45] Y. Kohayakawa, T. Łuczak, and V. Rödl. On k 4-free subgraphs of random graphs. *Combinatorica*, 17(2):173–213, 1997.

[46] J. Komlós, J. Pintz, and E. Szemerédi. On heilbronn's triangle problem. *Journal of the London Mathematical Society*, 2(3):385–396, 1981.

[47] A. Kostochka, D. Mubayi, and J. Verstraete. Hypergraph ramsey numbers: Triangles versus cliques. *Journal of Combinatorial Theory, Series A*, 120(7):1491–1507, 2013.

[48] A. Kostochka, D. Mubayi, and J. Verstraëte. Turán problems and shadows i: Paths and cycles. *Journal of Combinatorial Theory, Series A*, 129:57–79, 2015.

[49] A. V. Kostochka. A class of constructions for turán's (3, 4)-problem. *Combinatorica*, 2(2):187–192, 1982.

[50] J. Lauer and N. Wormald. Large independent sets in regular graphs of large girth. *Journal of Combinatorial Theory, Series B*, 97(6):999–1009, 2007.

[51] F. Lazebnik, V. A. Ustimenko, and A. J. Woldar. Upper bounds on the order of cages. *The Electronic Journal of Combinatorics*, 4(2), 1997.

[52] F. Lazebnik and J. Verstraete. On hypergraphs of girth five. *Electron J Combin*, 10, 11 2003.

[53] H. v. Maldeghem. *Generalized polygons /*. Springer Basel,, Basel ; New York :, reprint of the 1998 edition edition, 2012.

[54] W. Mantel. Problem 28 (solution by h. gouwentak, w. mantel, j. teixeira de mattes, f. schuh and w. a. wythoff). 10:60–61, 1907.

[55] B. McKay. Independent sets in regular graphs of high girth. *Ars Combinatoria*, 23:179–185, 1987.

[56] A. Méroueh. The ramsey number of loose cycles versus cliques. *Journal of Graph Theory*, 90(2):172–188, 2019.

[57] R. Morris and D. Saxton. The number of c_{2l}-free graphs. *Advances in Mathematics*, 298:534–580, 2016.

[58] D. Mubayi and J. Verstraete. A note on pseudorandom ramsey graphs, 2019.

[59] D. Mubayi and L. Yepremyan. Random tur\'an theorem for hypergraph cycles. *arXiv preprint arXiv:2007.10320*, 2020.

[60] A. Méroueh. The ramsey number of loose cycles versus cliques. *Journal of Graph Theory*, 90(2):172–188, 2019.

[61] G. Perarnau and B. Reed. Existence of spanning $\mathscr{F}$-free subgraphs with large minimum degree. *Combinatorics, Probability and Computing*, 26(3):448–467, 2017.

[62] G. F. Pontiveros, S. Griffiths, and R. Morris. *The triangle-free process and the Ramsey number R (3, k)*. 2020.

[63] V. Rödl and L. Thoma. Asymptotic packing and the random greedy algorithm. *Random Structures & Algorithms*, 8(3):161–177, 1996.

[64] K. F. Roth. On certain sets of integers. *Journal of the London Mathematical Society*, 1(1):104–109, 1953.

[65] W. Rudin. *Principles of mathematical analysis*, volume 3. McGraw-hill New York, 1976.

[66] I. Ruzsa and E. Szemer'edi. Triple systems with no six points carrying three triangles. *Combinatorica*, 18, 01 1976.

[67] G. Sárközy. Cycles in hypergraphs. 2008.

[68] D. Saxton and A. Thomason. Hypergraph containers. *Inventiones mathematicae*, 201(3):925–992, 2015.

[69] M. Schacht. Extremal results for random discrete structures. *Annals of Mathematics*, 184(2):333–365, 2016.

[70] J. B. Shearer. A note on the independence number of triangle-free graphs. *Discrete Mathematics*, 46(1):83–87, 1983.

[71] J. B. Shearer. A note on the independence number of triangle-free graphs, ii. *Journal of Combinatorial Theory, Series B*, 53(2):300–307, 1991.

[72] J. Spencer. Asymptotic lower bounds for ramsey functions. *Discrete Mathematics*, 20:69–76, 1977.

[73] B. Sudakov. A note on odd cycle-complete graph ramsey numbers. *The electronic journal of combinatorics.*, 9(1), 2002.

[74] P. Turán. On an extremal problem in graph theory. *Matematikai és Fizikai Lapok*, 48:436–452, 1941.

[75] J. Verstraëte. *Extremal problems for cycles in graphs*, pages 83–116. Springer International Publishing, Cham, 2016.

[76] V. G. Vizing. On an estimate of the chromatic class of a p-graph. *Discret Analiz*, 3:25–30, 1964.

[77] V. Wei. A lower bound on the stability number of a simple graph. Technical report, Bell Laboratories Technical Memorandum 81-11217-9, Murray Hill, NJ, 1981.