EXTREMAL PROBLEMS ON HAMMING GRAPHS

A Thesis

Presented to

The Faculty of the Department of Mathematics

California State University, Los Angeles

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

in

Mathematics

By

Jonathan Joseph Davidson

December 2025

© 2025

Jonathan Joseph Davidson

ALL RIGHTS RESERVED

The thesis of Jonathan Joseph Davidson is approved.

Dr. Jason O'Neill, Committee Chair

Dr. Mike Krebs

Dr. Justin Troyka

Dr. Daphne Liu, Department Chair

California State University, Los Angeles

December 2025

ABSTRACT

Extremal Problems on Hamming Graphs

By

Jonathan Joseph Davidson

This thesis studies forbidden subgraph problems in Hamming graphs $H(n, q)$. Classical extremal graph theory focuses on the dense setting where the host is the complete graph, but many natural networks are sparse and possess geometric or product-like structure. Understanding extremal thresholds in these settings requires methods that reflect the combinatorial geometry of the host. Our first main result is a Turán-type theorem for $H(n, q)$ in the regime where n is fixed and $q \to \infty$. We show that cliques K_r are dimensionally constrained, meaning every copy of K_r lies inside a single coordinate fiber. Using this structure and applying Zykov symmetrization within each fiber, we prove

$$\mathrm{ex}(H(n, q), K_r) = nq^{n-1}\,\mathrm{ex}(K_q, K_r),$$

and the extremal graphs are obtained by placing a Turán graph $T(q, r-1)$ on every fiber. We also prove an Erdős–Stone type theorem for Hamming graphs. If F is dimensionally constrained and has chromatic number $\chi(F) \geq 3$, then

$$\frac{\mathrm{ex}(H(n, q), F)}{|E(H(n, q))|} = 1 - \frac{1}{\chi(F) - 1} + o_q(1).$$

Thus, in the dense regime $q \to \infty$, the extremal density in $H(n, q)$ matches the classical Erdős–Stone density. We conclude by outlining structural questions about dimensionally constrained graphs and directions for extending these methods beyond the one-dimensional fiber setting.

ACKNOWLEDGMENTS

Acknowledgements

I would like to express my deepest gratitude to my thesis advisor, Jason O'Neill, for his guidance, encouragement, and generosity throughout this project. His support has shaped both the direction of this thesis and my development as a mathematician, and I am very grateful for everything he has done for me. I also thank my committee members, Mike Krebs and Justin Troyka, for their helpful feedback and for the many conversations that contributed to the ideas in this work. Their insight and willingness to brainstorm with me greatly improved the clarity and direction of this thesis. I am grateful to the Cal State LA Mathematics Department for providing a supportive and welcoming environment in which to learn and do research. I especially thank the department chair, Daphne Liu, for fostering a community that has made my time here meaningful and rewarding. Finally, I thank my wife, Cindy, and my parents for their love, patience, and emotional support. Their encouragement has sustained me throughout my studies, and I could not have completed this thesis without them.

TABLE OF CONTENTS

Abstract iv

Acknowledgments v

List of Figures viii

Chapter

1. Introduction 1

2. Preliminaries 7

2.1. Asymptotic Notation 7

2.2. Graph Theory Background 9

2.2.1. Neighborhood and Degree 9

2.2.2. Subgraphs and Induced Subgraphs 11

2.2.3. Walks, Paths, and Cycles 12

2.2.4. Chromatic Number 13

2.3. Special Graph Classes 15

2.3.1. Turán Graphs 15

2.3.2. Product Graphs 17

2.4. Graph Homomorphisms 18

2.4.1. Injective Graph Homomorphisms 19

2.4.2. Graph Homomorphisms Detect Graph Properties 21

2.4.3. Homomorphism Counts 23

3. Classical Extremal Graph Theory 26

3.1. Forbidden Subgraphs and Extremal Numbers 26

3.2. Mantel's Theorem and Early Extremal Theory 27

3.3. Turán's Theorem and Symmetrization Methods 30
3.3.1. Zykov Symmetrization 31
3.4. Erdős–Stone–Simonovits Theory 34
3.5. Bipartite Extremal Problems 37
3.6. Open Problems and the Rational Exponents Conjecture 41
4. Extremal Graph Theory on Hamming Graphs 44
4.1. Even Cycles in the Hypercube 44
4.2. Forbidden Subgraphs in Hamming Graphs 46
4.3. Dimensionally Constrained Subgraphs of Hamming Graphs 47
5. Conclusion and Further Directions 55
5.1. Even Cycles and Chung-Type Results on Hamming Graphs 55
5.2. Extremal Problems in the High-Dimensional, Fixed-q Regime 56
5.3. Characterizing Dimensionally Constrained Graphs 58
5.4. Partially Dimensionally Constrained Graphs 58
References 60

LIST OF FIGURES

Figure

2.1. The complete graph K_5 and the empty graph $\overline{K_5}$. 10

2.2. Subgraphs vs. Induced Subgraphs 12

2.3. The graphs P_4 and C_6. 13

2.4. Odd cycle from vertices with different distance parities. 15

2.5. Example of a bipartite graph and a 3-partite graph. 15

2.6. The Turán graph $T(10, 4)$. 16

2.7. Vertex removal process on $T(8, 3)$. 17

2.8. The product graph $K_3 \square P_4$ 17

2.9. The hypercube Q_3 and the Hamming graph $H(2, 3)$ 18

2.10. Inclusion map example. 19

2.11. Induced homomorphism example. Map is specified by labeling. 20

2.12. Graph isomorphism example. Map is specified by labeling. 21

3.1. Zykov symmetrization of a pair of nonadjacent vertices u and v. 31

CHAPTER 1

Introduction

The widespread adoption of large language models and neural networks over the past decade has heightened the need for mathematical tools to analyze large and complex networks. One rigorous mathematical framework for analyzing such networks is property testing, which seeks to infer global structure from limited local information. For instance, the detection of network motifs—small induced subgraphs occurring with higher frequency within a network—has been proposed as a means of decomposing neural and biological networks into fundamental building blocks [14]. Understanding the conditions under which particular motifs arise naturally leads to questions of an extremal nature, connecting motif analysis to the forbidden subgraph problem in combinatorics. Given a host graph G and a forbidden graph F, the forbidden subgraph problem asks for the extremal number $\mathrm{ex}(G, F)$, the maximum number of edges in a subgraph G that contains no subgraph isomorphic to F.

Sparse host graphs such as certain Cayley graphs provide natural models for networks whose connectivity reflects similarity, locality, or symmetry rather than randomness. These highly regular structures capture geometric and algebraic constraints that frequently arise in neural, biological, and information networks. Recent studies in biological and neural networks have shown that geometric constraints—for example, embedding in low-dimensional spaces—significantly shape network connectivity and motif frequencies [15, 18]. At the same time, algebraic and symmetry constraints in neural network architectures and graph-based models reflect group and invariant structures in data [22, 23]. Extremal questions on such hosts quantify how dense a

system with built-in structure can become before a prescribed substructure must appear. These thresholds offer a combinatorial analogue of phase transitions in network behavior. Moreover, the symmetry and product-like geometry of these graphs make them testbeds for studying how local constraints influence global structure.

While this thesis focuses on studying the forbidden subgraph problem in the sparse host graph setting, the study of extremal combinatorics stems from the dense host graph setting. Before turning to the sparse case, it is instructive to recall the central results that characterize the dense regime. Classical extremal combinatorics provides a theoretical foundation for studying the extremal number when the host graph is the complete graph. For instance, Turán's Theorem [2] bounds the number of edges of a graph on n vertices that is K_k-free. The bound is attained by the Turán graph, thereby determining the extremal number $\mathrm{ex}(n, K_k)$. More generally, the celebrated Erdős–Stone–Simonovits Theorem [3, 6], determines asymptotically the extremal number $\mathrm{ex}(n, F)$ when the chromatic number $\chi(F) > 2$:

$$\mathrm{ex}(n, F) = \frac{\chi(F) - 2}{\chi(F) - 1}\binom{n}{2} + o(n^2).$$

When F is bipartite, the Erdős–Stone–Simonovits theorem only implies that $\mathrm{ex}(n, F) = o(n^2)$, and determining the correct growth rate remains one of the central challenges of extremal graph theory. Substantial progress has been made for certain families of bipartite graphs. In particular, determining the asymptotic behavior of $\mathrm{ex}(n, K_{s,t})$ is known as the Zarankiewicz problem. The classical Kővári–Sós–Turán theorem [5] provides the upper bound

$$\mathrm{ex}(n, K_{s,t}) = O(n^{2-1/s}),$$

valid for all integers $2 \leq s \leq t$. Matching lower bounds of order $\Omega(n^{2-1/s})$ are known for many parameter ranges, based on algebraic constructions due to Kollár, Rónyai, and Szabó [12] and extended by Alon, Rónyai, and Szabó [13]. A constant-factor gap still remains in several cases, and determining the precise order of magnitude of $\mathrm{ex}(n, K_{s,t})$ continues to be a major open problem.

Progress on the Zarankiewicz problem also yields tighter upper bounds on the extremal number of general bipartite graphs. In particular, a classical reduction due to Erdős and Simonovits [6] and refined by Füredi and Simonovits [19] shows that for every bipartite graph F with maximum degree Δ, there exists a constant C_F such that

$$\mathrm{ex}(n, F) \leq C_F \ \mathrm{ex}(n, K_{s,t}),$$

for some integer t depending only on F and some integer $s \leq \Delta$. However, this bound can be too weak to capture the precise growth rate of some bipartite graphs. For example, the classical theorem of Bondy and Simonovits [7] shows that, for each fixed $k \geq 2$,

$$\mathrm{ex}(n, C_{2k}) = \mathcal{O}(n^{1+1/k}),$$

a substantial improvement over what the reduction theorem implies. Moreover, this bound has been shown to be tight up to constant factors for several values of k. In particular, explicit constructions based on finite geometry [11] and algebraic methods [9] yield matching lower bounds for C_4, C_6, and C_{10}. For a modern exposition on the extremal number of even cycles, see [21].

The results discussed so far concern the case in which the host graph is com-

plete. However, for bipartite forbidden graphs F, the choice between the complete graph K_n and the complete bipartite graph $K_{n,n}$ as the host does not affect the asymptotic behavior of the extremal number. Indeed, by a classical lemma of Erdős, every n-vertex graph contains a bipartite subgraph with at least half of its edges [19, Section 2.1]. Consequently, any F-free graph on n vertices has an F-free bipartite subgraph with a comparable number of edges, and conversely, an extremal F-free subgraph of $K_{n,n}$ can be embedded into K_{2n}. Hence, $\mathrm{ex}(K_{n,n}, F)$ and $\mathrm{ex}(n, F)$ have the same asymptotic growth rate. This equivalence indicates that, in the dense setting, the complete and complete bipartite hosts are essentially interchangeable for extremal problems involving bipartite forbidden subgraphs.

In contrast, when the host graph is sparse or possesses additional geometric or algebraic structure, the situation changes dramatically. Sparse hosts can exhibit rigid local constraints and highly symmetric global organization, so the classical reduction arguments no longer apply. The extremal number $\mathrm{ex}(G, F)$ now depends sensitively on the structure of G itself, not just on the forbidden subgraph F. The underlying geometry of the host begins to influence the threshold at which a given configuration must appear. One of the earliest and most influential results in this direction is due to Chung [10], who obtained upper bounds for $\mathrm{ex}(Q_n, C_{2k})$, the largest number of edges in a C_{2k}-free subgraph of the n-dimensional hypercube Q_n. In particular, Chung showed, for $k = 2, 3$,

$$\lim_{n\to\infty} \frac{\mathrm{ex}(Q_n, C_{2k})}{e(Q_n)} > 0$$

where $e(G)$ is the number of edges of the graph G. Whereas the Erdős–Stone–

Simonovits theorem implies that even cycle free graphs can have at most a vanishing proportion of the edges in the dense host setting, Chung's result shows that some C_4-free (and C_6-free) subgraphs of Q_n can have an asymptotically nonzero proportion of the edges of the host graph. This contrast highlights how the geometry and regularity of the host graph can fundamentally alter extremal thresholds.

In this thesis we initiate a systematic study of forbidden subgraph problems in Hamming graphs $H(n, q)$, focusing on the dense regime where the dimension n is fixed and the alphabet size q tends to infinity. Our first main result is a Turán-type theorem for cliques in Hamming graphs. We show that the geometry of $H(n, q)$ forces every copy of K_r to lie entirely within a single one-dimensional fiber. This rigidity allows us to apply classical Zykov symmetrization within each fiber and prove that the extremal K_r-free subgraphs of $H(n, q)$ are obtained by placing a Turán graph $T(q, r-1)$ on every fiber. Consequently,

$$\mathrm{ex}(H(n, q), K_r) = nq^{n-1}\ \mathrm{ex}(K_q, K_r),$$

so the extremal density in $H(n, q)$ agrees with the classical Turán density $1 - \frac{1}{r-1}$.

We extend this result to a broad family of forbidden graphs. Our second main theorem is an Erdős–Stone-type result for Hamming graphs: if F is a fixed graph with chromatic number $r \geq 3$ and F is dimensionally constrained (i.e. every embedding of F into $H(n, q)$ uses only one coordinate), then

$$\mathrm{ex}(H(n, q), F) = \left(1 - \frac{1}{r-1} + o_q(1)\right) |E(H(n, q))|.$$

In particular, the classical Erdős–Stone density reemerges in the Hamming graph setting precisely when the forbidden configuration is forced to appear within a fiber.

To support these results, we identify and study a class of graphs that are dimensionally constrained in Hamming graphs. We show that cliques, almost-cliques, and sufficiently large Turán graphs each possess this property, and we provide a framework for reducing global forbidden subgraph problems on $H(n, q)$ to fiber-wise extremal problems on K_q. This reduction mechanism forms the core of our method and suggests new directions for approaching more intricate extremal problems in structured product graphs.

CHAPTER 2

Preliminaries

This chapter introduces the notation and definitions used throughout the thesis. We begin with basic terminology from graph theory. Then, we cover asymptotic notation and reformulate some graph properties in the language of graph homomorphisms. Lastly, we define Hamming graphs and prove some important properties of Hamming graphs.

Let $\mathbb{N}$ denote the set of natural numbers excluding 0. For a natural number n, we define $[n] \coloneqq \{1, 2, \ldots, n\}$. For a real number x we denote by $\lfloor x \rfloor$ the greatest integer less than x, and we denote $\lceil x \rceil$ to be the least integer greater than x. For a set S, the power set 2^S will denote the set of all subsets of S, and $\binom{S}{k}$ will denote the set of all subsets of S with cardinality k. A set containing k elements will be referred to as a k-set, and a subset with k elements will be called a k-subset. A set containing only one element is called a **singleton** set. A set $\{A_1, \ldots, A_m\} \subset 2^A$ is called a **partition** of A if all of the sets in $\{A_1, \ldots, A_m\}$ are pairwise disjoint and

$$\bigcup_{i=1}^{m} A_i = A$$

2.1 Asymptotic Notation

Many extremal questions in this thesis involve families of host graphs whose size grows with parameters such as dimension or alphabet size. Asymptotic notation provides a concise language for describing the behavior of extremal quantities in these limits. We will use standard asymptotic notation throughout. All functions are assumed to be nonnegative for sufficiently large arguments.

Suppose we are comparing functions $f, g : \mathbb{N} \to \mathbb{R}$. We will use the following Bachman-Landau asymptotic notation:

- We write $f(n) = O(g(n))$ if there exist constants $C > 0$ and n_0 such that

$$f(n) \leq Cg(n) \quad \text{for all } n \geq n_0.$$

- We write $f(n) = \Omega(g(n))$ if there exist constants $c > 0$ and n_0 such that

$$f(n) \geq cg(n) \quad \text{for all } n \geq n_0.$$

- We write $f(n) = \Theta(g(n))$ if both $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$.

- We write $f(n) = o(g(n))$ if

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0.$$

In the case of multivariate asymptotics, we will only vary one parameter at a time. We will adapt the notation above to specify which parameter is varying. Formally, if $f, g : \mathbb{N} \times \mathbb{N} \to \mathbb{R}$, we write $f(n, q) = O_n(g(n, q))$ if there exist constants $C > 0$ and n_0 such that

$$f(n, q) \leq Cg(n, q) \quad \text{for all } n \geq n_0$$

where q is fixed. Similar adaptations hold for Big Omega, Big Theta, and little-o notation.

Remark 2.1 (Asymptotically negligible terms)**.** We use $o(1)$ to denote any quantity tending to 0 in the relevant limit. For example, $o_d(1)$ denotes a term tending to 0 as $d \to \infty$ with all other parameters fixed.

2.2 Graph Theory Background

This section establishes the notation and terminology from graph theory that will be used throughout the thesis. The presentation here is intended to be concise and self-contained; readers seeking additional background may consult the standard reference by Bondy and Murty [17] or by Diestel [20].

A **graph** is an ordered tuple (V, E) where V is a set of **vertices**, and E is a multiset of **edges**, 2-subsets of V. We will write uv to denote an edge $\{u, v\} \in E$. Occasionally, we may allow E to contain singleton sets, called **loops**. A graph is said to be **simple**, if it does not contain any loops and if all edges have multiplicity one. For a graph G, we write $V(G)$ for its vertex set and $E(G)$ for its edge set. When the graph G is certain from the context, we will simply write V and E. The number of vertices in a graph G is called the **order**, written as $|V(G)|$. We also define the **size** $|E(G)|$ to be the number of edges in the graph. The graph G is said to be **finite**, if the order is finite. Unless otherwise noted, all graphs are assumed to be finite and simple.

2.2.1 Neighborhood and Degree

An edge e is said to be **incident** to a vertex v if $v \in e$. Two vertices are **adjacent** (or neighbors) if there exists an edge incident to both vertices. In this case, we call the vertices the **endpoints** of the edge. If a subset $U \subset V$ of k vertices are all pairwise adjacent, we call U a k-**clique**. On the other hand, if two vertices are not adjacent, we say that the vertices are **independent**. Moreover, a subset $U \subset V$ is called an **independent set** if all of the vertices are pairwise independent. The

number of vertices in the largest clique in a graph G is denoted by $\omega(G)$. The number of vertices in the largest independent set in a graph G is denoted by $\alpha(G)$.

The **complete graph** on n vertices K_n is the graph where all vertices are adjacent. The **empty graph** on n vertices $\overline{K_n}$ is the graph where no pair of vertices is adjacent. The graphs K_5 and $\overline{K_5}$ are depicted in Figure 2.1. More generally, $\overline{G}$ denote the **complement** of the graph G. The complement has the same vertex set as V, but $e \in E(\overline{G})$ if and only if $e \notin E(G)$.

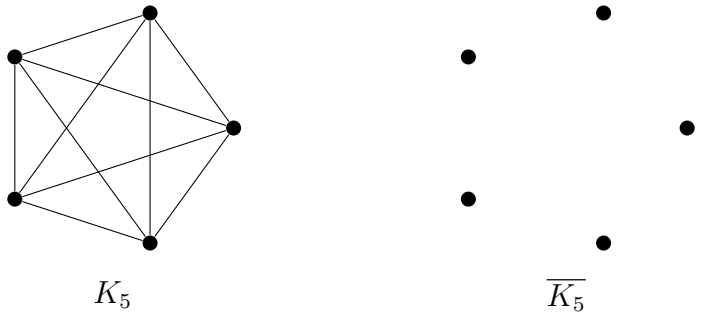


Figure 2.1: The complete graph K_5 and the empty graph $\overline{K_5}$.

For a fixed vertex $v \in V(G)$, the **neighborhood** of v is defined to be the set

$$N_G(v) \coloneqq \{u \in V(G) : uv \in E(G)\}$$

More generally, if U is any subset of V, the neighborhood $N(U)$ is the set of all vertices in $V \setminus U$ who are adjacent to at least one vertex in U. The **degree** of a vertex v is denoted $\deg_G(v) \coloneqq |N_G(v)|$. When the graph G is clear from context, we may omit the subscript and simply write $N(v)$ and $\deg(v)$. Furthermore, we denote the minimum and maximum degree of a graph G by $\delta(G)$ and $\Delta(G)$, respectively. The quantity

$$\overline{\deg(G)} = \frac{1}{|G|} \sum_{u \in V} \deg(u)$$

is called the **average degree** of the graph. We observe that the degree of a vertex also counts the number of edges incident to a vertex. Since all edges are incident to exactly two vertices, we count every edge twice when we sum up all of the degrees. Consequently, we establish the following lemma:

Lemma 2.2 (Handshaking Lemma)**.** *Let G be a graph. Then,*

$$2|E| = \sum_{v \in V} \deg(v)$$

It is immediate from the definition of the average degree that

$$\overline{\deg(G)} = \frac{2|E|}{|V(G)|}$$

A graph is said to be **regular** of degree k if every vertex has degree k. If G is k-regular, then the Handshaking Lemma can be used to established that

$$|E| = \frac{1}{2}k|V(G)|$$

In particular, the complete graph K_n is $n-1$ regular; consequently,

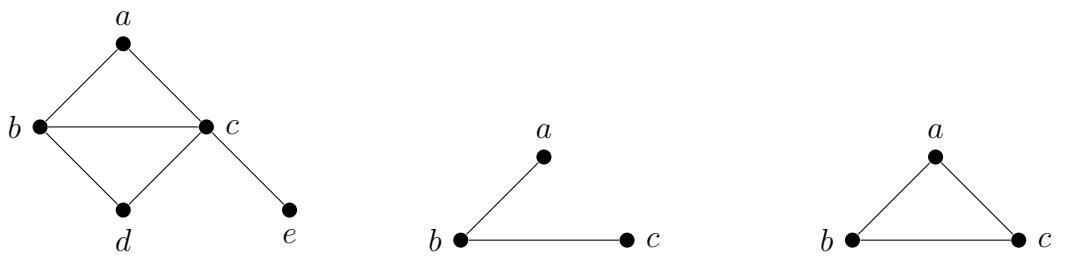
$$|E(K_n)| = \frac{n(n-1)}{2} = \binom{n}{2}$$

2.2.2 Subgraphs and Induced Subgraphs

Let $G = (V, E)$. If $H = (V', E')$ is a graph such that $V' \subset V$ and $E' \subset E$, we say that H is a **subgraph** of G, denoted by $H \subset G$. Subgraphs arise naturally when we delete edges or restrict our attention to a local structure within a larger graph. If $U \subseteq V(G)$ is any subset of vertices, the **induced subgraph** $G[U]$ is the graph that retains all adjacencies of G among vertices of U and no others. Formally speaking, $G[U] = (U, E')$ where

$$E' = \{uv \in E(G) : u, v \in U\}$$

A subgraph $H = (V', E')$ of G is said to be **induced** if $H = G[V']$. Figure 2.2 shows two subgraphs where one graph is induced and the other is not induced.



(a) Graph (b) Subgraph (c) Induced Subgraph

Figure 2.2: Subgraphs vs. Induced Subgraphs

For a fixed graph $G = (V, E)$, let $U \subset V$. We define $G - U \coloneqq G[V \setminus U]$. In other words, $G - U$ is the graph obtained by deleting the vertices in U from G and removing all edges incident to the vertices in U. If $U = \{v\}$ is singleton, the notation will be abbreviated as $G - v$. If $F \subset \binom{V}{2}$, we write $G - F \coloneqq (V, E \setminus F)$ and $G + F \coloneqq (V, E \cup F)$. Similar to the case above, $G - \{e\}$ and $G + \{e\}$ are abbreviated to $G - e$ and $G + e$, respectively.

2.2.3 Walks, Paths, and Cycles

A walk serves as the most general form of traversal in a graph, and stricter forms such as paths, and cycles emerge by prohibiting certain repetitions. Let G be a graph. A **walk** in G is a sequence of vertices $v_0, v_1, \ldots, v_k$ such that $v_{i-1}v_i \in E(G)$ for each $i = 1, \ldots, k$. The integer k is the **length** of the walk. Walks may repeat vertices. A walk that contains no repeated vertices is called a **path**. A walk is **closed** if its initial and terminal vertices coincide, that is, if $v_0 = v_k$. A closed path is called a **cycle**. The **girth** of a graph G, denoted girth(G), is the length of a shortest cycle

in G. If G contains no cycles, we adopt the convention that $\text{girth}(G) = \infty$.

The **path graph** P_k is the graph with vertex set $V = \{v_1, \ldots, v_k\}$ and edge set $E = \{v_1v_2, \ldots, v_{k-1}v_k\}$. The **cycle graph** C_k is the graph with vertex set $V = \{v_1, \ldots, v_k\}$ and edge set $E = \{v_1v_2, \ldots, v_{k-1}v_k, v_kv_1\}$. Examples of these graphs are depicted in Figure 2.3.

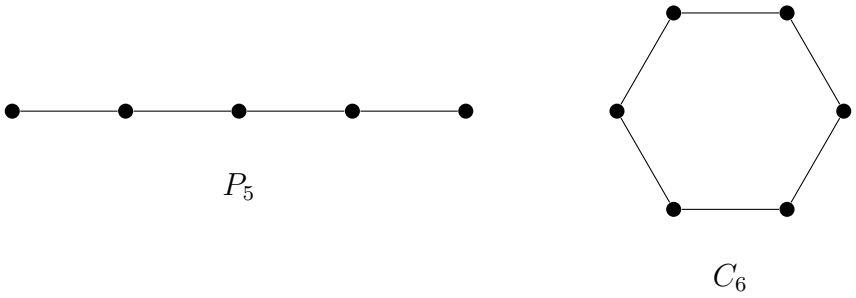


Figure 2.3: The graphs P_4 and C_6.

2.2.4 Chromatic Number

Many extremal properties of graphs depend on how their vertices can be partitioned into independent sets. A partition of the vertex set into k independent sets is called a **proper k-coloring**. If there exists a proper k-coloring of a graph G, we say that G is **k-colorable**. In this language, proper k-colorings correspond to assignments of colors to vertices in which each independent set forms a color class. Hence, a proper k-coloring is equivalent to a function $\chi : [k] \to V(G)$, such that each image $\chi(i)$ is an independent set. The **chromatic number** of a graph G, denoted $\chi(G)$, is the smallest integer k such that G is k-colorable. The chromatic number of several families of graphs can be computed directly:

- $\chi\big(\overline{K_n}\big) = 1$: the empty graph has no edges, so it is a single independent set.

- $\chi(K_n) = n$: every vertex of the complete graph K_n is adjacent to every other, so no two vertices may share a color
- $\chi(P_k) = 2$: color vertices by alternating between two colors, so this means that P_k is two colorable. However, any graph with at least one edge must have at least two colors. Therefore, $\chi(K_{m,n}) \geq 2$

The chromatic number of the cycle graph C_k depends on the parity of k: even cycles satisfy $\chi(C_{2k}) = 2$, while odd cycles satisfy $\chi(C_{2k+1}) = 3$. The behavior of cycle graphs reflects a fundamental structural characterization of bipartite graphs:

Lemma 2.3. *A graph G is bipartite if and only if it contains no odd cycle.*

Proof. If G is bipartite with parts A and B, then every cycle alternates between these parts, so every cycle has even length. Hence a bipartite graph contains no odd cycle.

Conversely, assume that G contains no odd cycle. Choose any uncolored vertex x and assign it one of two colors. Then, for every uncolored vertex, assign its color according to the parity of the length of a shortest path to x: vertices at even distance receive the same color as x, and vertices at odd distance receive the other color. If a vertex is encountered that is not reachable from any vertex already colored, select it as a new starting point and repeat the same procedure. Since G is finite, this process eventually colors every vertex. Two adjacent vertices cannot receive the same color: if they did, their distances from the chosen starting point would have the same parity, which would create an odd cycle. Thus the coloring is proper, and G is bipartite. □

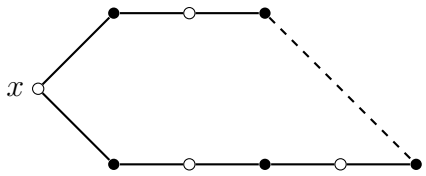


Figure 2.4: Odd cycle from vertices with different distance parities.

2.3 Special Graph Classes

2.3.1 Turán Graphs

Let $r \geq 2$ be an integer. A graph $G = (V, E)$ is called r-**partite** if V admits a partition into r subsets such that every edge has endpoints in different classes. When $r = 2$ we say bipartite instead of 2-partite. The **star graph** $S_r \coloneqq K_{1,r}$ is a special type of bipartite graph that plays a large role in extremal combinatorics.



Bipartite graph 3-partite graph

Figure 2.5: Example of a bipartite graph and a 3-partite graph.

The complete r-partite graph, denoted $K_{n_1,\ldots,n_r}$, is an r-partite graph with with parts of size $n_1, \ldots, n_r$ and the maximum number of edges. A partition $V = V_1 \sqcup \cdots V_m$ is called **equitable** if $\big||V_i| - V_j|\big| \leq 1$ for all $i, j \in [m]$. The **Turán graph** $T(n, r)$ is a complete r-partite graph with an equitable partition. Figure 2.6 illustrates the Turán graph $T(10, 4)$.

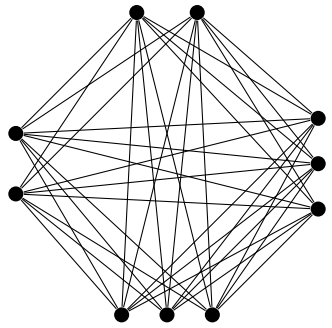


Figure 2.6: The Turán graph $T(10, 4)$.

Lemma 2.4. *Let $n, r \in \mathbb{N}$ such that $r \leq n$. Then,*

$$\left|E\big(T(n,r)\big)\right| \geq \Big(1 - \frac{1}{r}\Big)\frac{n^2}{2} - \frac{r}{8}$$

Proof. Each part of a Turán graph $T(n, r)$ has size $\lfloor n/r \rfloor$ or $\lfloor n/r \rfloor + 1$. To count the edges, we write $n = qr + s$ for some integers q and s such that $0 \leq s < r$. The Turán graph $T(n, r)$ has s subsets of cardinality $q+1$. We can iteratively remove one vertex from each part of size $q + 1$ until we have the Turán graph $T(qr, r)$. The kth vertex removed takes away $n - q - k$ edges. This removal process is depicted in Figure 2.7. By the end of this process,

$$\sum_{k=1}^{s}(n - q - k) = s(n - q) - \binom{s+1}{2}$$

edges have been removed. The remaining graph $T(qr, r)$ will be a regular graph with each vertex having degree $n - q - s$. By the Handshaking Lemma,

$$\Big|E\big(T(n,r)\big)\Big| = \frac{1}{2}qr(n - q - s) + s(n - q) - \binom{s+1}{2}$$

After some algebra,

$$\Big|E\big(T(n,r)\big)\Big| = \frac{n^2 - q^2r - 2sq - s}{2} = \frac{(r-1)n^2 - s(r-s)}{2r}$$

The quantity $-s(r-s)$ is maximized when $s = r/2$. Substituting this value for s into the expression above completes the proof. □

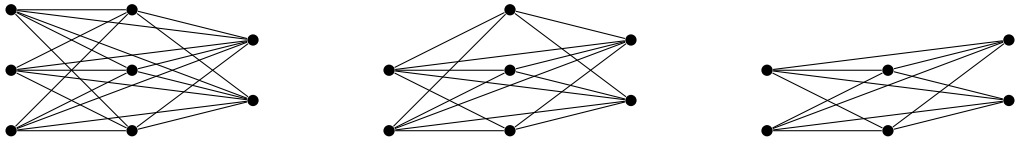


Figure 2.7: Vertex removal process on $T(8, 3)$.

2.3.2 Product Graphs

Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be two graphs. The **product graph** $G_1 \square G_2$, is the graph with vertex set $V_1 \times V_2$ and edge set $E(G_1 \square G_2) = E_1' \cup E_2'$ where

$$E_1' = \{(u, x)(v, x) : uv \in E_1 \text{ and } x \in V_2\}$$

$$E_2' = \{(u, x)(u, y) : xy \in E_2 \text{ and } u \in V_1\}$$

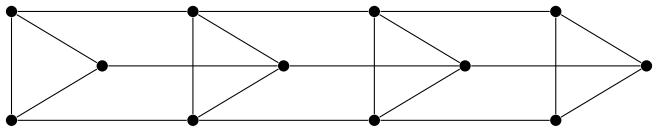


Figure 2.8: The product graph $K_3 \square P_4$.

For a fixed graph G, the **n-fold product graph** G^n is defined recursively:

$$G^n = G^{n-1} \square G, \quad G^1 = G$$

We define the **hypercube graph** $Q_n \coloneqq K_2^n$. More generally, we define the **Hamming graph** $H(n, q) \coloneqq K_q^n$. If we label the vertices of K_q by $[q]$, the Hamming graph $H(n, q)$ is the graph with vertex set $\mathbb{Z}_q^n$ where two vectors are connected if and only if they differ in one coordinate.

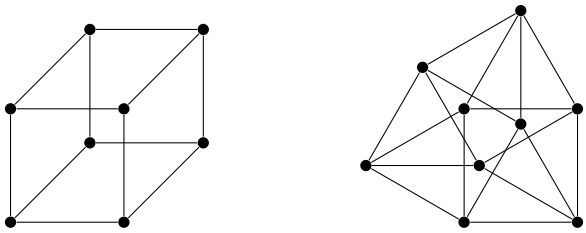


Figure 2.9: The hypercube Q_3 and the Hamming graph $H(2,3)$

Lemma 2.5. *Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be two graphs. Then,*

$$\big|E(G_1 \square G_2)\big| = |V_1|\big|E(G_2)\big| + |V_2|\big|E(G_1)\big|$$

Proof. By definition of the product, $\forall uv \in E_1$ and $\forall x \in V_2$, $(u,x)(v,x)$ will be an edge in $G_1 \square G_2$. This counts $|V_2|\big|E(G_1)\big|$ edges. Similarly, $\forall xy \in E_2$ and $\forall u \in V_1$, $(u,x)(u,y)$ will be an edge in $G_1 \square G_2$. This counts $|V_1|\big|E(G_2)\big|$ edges. This counts every edge in $G_1 \square G_2$. □

Corollary 2.6. *Let $G = (V, E)$ be a graph. Then,*

$$\big|E(G^n)\big| = n|V|^{n-1}|E(G)|$$

From Corollary 2.6, we can determine the number of edges in the Hamming graph

$$|E(H(n,q))| = nq^{n-1}\binom{q}{2}$$

2.4 Graph Homomorphisms

Graph homomorphisms provide a natural framework for several fundamental structures in graph theory. We begin with the definition and then record several important special cases.

Definition 2.7 (Graph Homomorphism)**.** *Let H and G be graphs. A graph homomorphism from H to G is a function $\varphi : V(H) \to V(G)$ such that whenever $uv \in E(H)$,*

the image satisfies $\varphi(u)\varphi(v) \in E(G)$.

Thus a graph homomorphism preserves all adjacencies present in H. However, a graph homomorphism imposes no restrictions on non-adjacent vertices: non-adjacent vertices may map to adjacent vertices in G, and distinct vertices may map to the same vertex. A graph homomorphism may additionally satisfy three independent properties: injectivity, surjectivity, and non-edge preservation.

2.4.1 Injective Graph Homomorphisms

Among homomorphisms, injective graph homomorphisms arise naturally in many contexts. A homomorphism $\varphi : H \to G$ is said to be injective if it is injective as a map on vertices. We will write $H \hookrightarrow G$ to indicate an injective graph homomorphism. We record three important examples.

(1) Inclusion maps. If H is a subgraph of G, then the inclusion $\iota : V(H) \hookrightarrow V(G)$ is an injective graph homomorphism: edges of H remain edges of G, and distinct vertices remain distinct. When H is not an induced subgraph, the inclusion map does not necessarily preserve non-adjacencies.

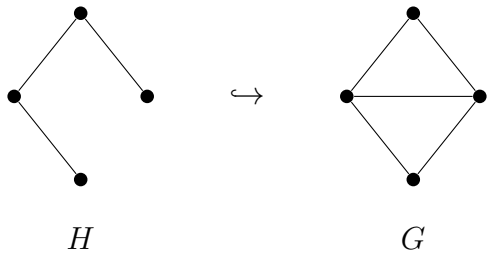


Figure 2.10: Inclusion map example.

(2) Induced Homomorphism. An injective graph homomorphism $\varphi : H \hookrightarrow G$ is called an **induced homomorphism** if it also preserves non-adjacencies; that is, $uv \notin E(H)$ implies $\varphi(u)\varphi(v) \notin E(G)$. Equivalently, $\varphi(H)$ is an induced subgraph of G.

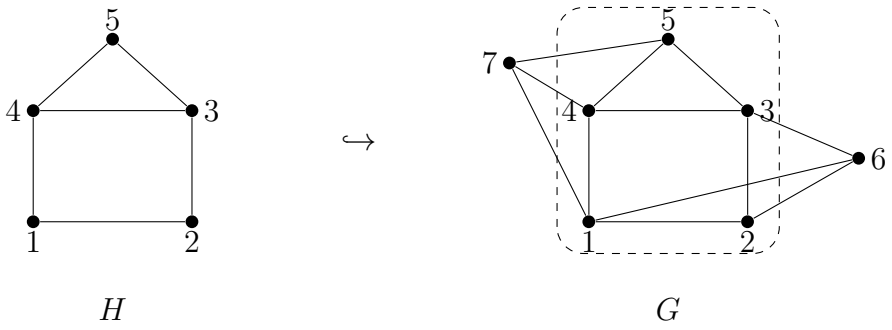


Figure 2.11: Induced homomorphism example. Map is specified by labeling.

(3) Graph isomorphisms. An induced homomorphism $\varphi : H \hookrightarrow G$ is called an **isomorphism** if it is also surjective. Isomorphisms are bijective graph homomorphisms preserving both adjacencies and non-adjacencies, and we write $H \cong G$ to indicate that there exists an isomorphism between H and G. If a graph G contains a subgraph isomorphic to H, we say that G contains a copy of H. Furthermore, a graph isomorphism $G \to G$ will be called an automorphism. The set of all automorphisms is denoted $\mathrm{Aut}(G)$.

Injective graph homomorphisms provide a natural way to express subgraph containment. If H and G are graphs, then H appears as a (not necessarily induced) subgraph of G precisely when there exists an injective graph homomorphism $H \hookrightarrow G$. Injectivity ensures that distinct vertices of H remain distinct in the image, while the graph homomorphism condition ensures that all adjacencies of H are preserved.

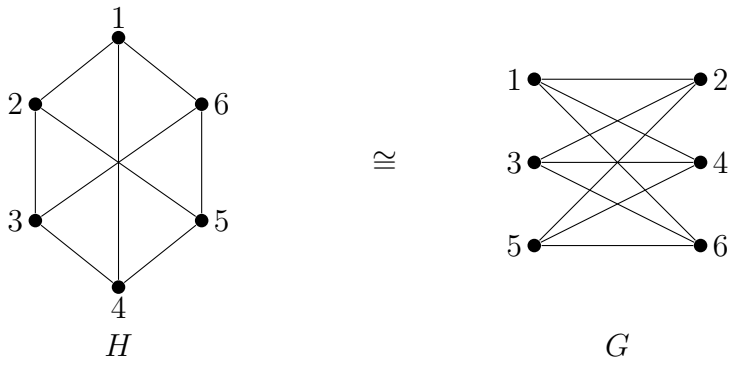


Figure 2.12: Graph isomorphism example. Map is specified by labeling.

If, in addition, the image of H in G contains no extra edges beyond those of H, then the map realizes H as an induced subgraph. Thus induced copies of H correspond exactly to induced homomorphisms that preserve both adjacencies and non-adjacencies. This distinction between general injective graph homomorphisms and induced homomorphisms will later be reflected in the corresponding notions of homomorphism density and induced homomorphism density.

Injective homomorphisms capture the presence of subgraphs, and induced injective homomorphisms capture the presence of induced subgraphs. This observation suggests that many other graph properties can likewise be characterized by the existence of appropriate graph homomorphisms. We now turn to several examples.

2.4.2 Graph Homomorphisms Detect Graph Properties

Many classical graph parameters and structural properties can be expressed naturally in terms of graph homomorphisms. The direction of the homomorphism encodes different types of information: maps of the form $G \to H$ typically impose global constraints on G, while maps of the form $H \to G$ reveal the presence of specific

local patterns in G.

Chromatic number. A proper coloring of a graph G with r colors is exactly a homomorphism $G \longrightarrow K_r$. Each color class is mapped to a distinct vertex of K_r, and the adjacency structure of K_r ensures that adjacent vertices receive different colors. Consequently,

$$\chi(G) = \min\{r : G \to K_r \text{ exists}\}.$$

Bipartite graphs. A graph G is bipartite if and only if it admits a homomorphism $G \longrightarrow K_2$. The two parts of the bipartition are simply the preimages of the two vertices of K_2.

Girth. The girth of G is the length of its shortest cycle. Homomorphisms from cycle graphs provide a direct way to detect this quantity:

$$\text{girth}(G) = \min\{k : C_k \to G \text{ exists}\}$$

A homomorphism $C_k \to G$ corresponds to a closed walk of length k in G; taking the smallest such k recovers the girth.

Cliques. The clique number $\omega(G)$ is the size of the largest complete subgraph of G, and can be expressed through homomorphisms:

$$\omega(G) = \max\{r : K_r \to G \text{ exists}\}$$

Observe that any homomorphism $K_r \to G$ that is not injective will force G to have a loop. Since we only consider simple graphs in this thesis, all homomorphism $K_r \to G$ must be injective.

Vertices of high degree. A vertex of degree at least d exists in G if and only if there is an injective homomorphism

$$S_d \hookrightarrow G.$$

Walks and cycles. A walk $v_0, v_1, \ldots, v_k$ in G is precisely a homomorphism $P_{k+1} \to G$, and a closed walk is a homomorphism $C_k \to G$. Thus the existence of long paths or cycles in G can be detected by the presence of homomorphisms from these small templates.

The examples above illustrate a general pattern: graph homomorphisms of the form $G \to H$, where H is a fixed small graph, typically encode global constraints on the structure of G. A graph homomorphism into K_r organizes the entire graph according to a coloring rule, and a graph homomorphism into K_2 reveals that a graph is bipartite. In each case, the small graph H serves as a structural template whose adjacency relations impose a global organizational pattern on G.

In contrast, homomorphisms of the form $H \to G$, where H is small, probe the local structure of G. Injective homomorphisms of stars identify vertices of large degree, and embeddings of paths or cycles detect the presence of specific local configurations such as branching or circular structure. Here the small graph H functions as a pattern whose appearance inside G reveals information about the local neighborhoods and substructures of G.

2.4.3 Homomorphism Counts

The existence of specific types of graph homomorphisms can reveal many interesting properties of a graph. It turns out that enumerating all the possible graph

homomorphism between two graphs can be a powerful tool in extracting more information out of a graph. In later chapters, we will prove some classical theorems in extremal graph theory via double counting certain graph homomorphisms. The **homomorphism count** $\hom(H, G)$ denotes the number of homomorphisms of the form $H \to G$.

The following examples illustrate how homomorphism counts measure increasingly refined structural information about a graph G.

- The homomorphism count $\hom(K_2, G)$ counts how many ways an edge can be mapped into G. Since each edge can be mapped to any edge in G in one of two orientations, we have

$$\hom(K_2, G) = 2|E(G)|$$

- For a star S_r, the center of S_r can be mapped to each vertex u in G and its r leaves can each be mapped to one of $\deg(u)$ choices. Hence,

$$\hom(S_r, G) = \sum_{u \in V(G)} \deg(u)^r$$

  Star counts record degree moments of the distribution of degrees in G. Not every map is injective since leaves of S_r can map onto the same neighbor of u.

- For a cycle C_k, the homomorphism count $\hom(C_k, G)$ counts the number of closed walks of length k in G. Unlike stars, non-injective homomorphisms from C_k abound: a cycle may collapse to a shorter cycle or even a single edge.

Since $K_2 \cong S_1$, the examples above show that

$$\sum_{u \in V(G)} \deg(u) = \hom(S_1, G) = \hom(K_2, G) = 2|E(G)|$$

Thus, double counting $\hom(K_2, G)$ provides an alternate proof of the Handshaking Lemma. Moreover, $\hom(C_k, G)$ counting closed walks means we can apply methods from spectral graph theory to the theory of homomorphism counts. These examples reveal the versatile nature of counting homomorphisms in graphs.

CHAPTER 3

Classical Extremal Graph Theory

3.1 Forbidden Subgraphs and Extremal Numbers

A fundamental question in extremal graph theory is to determine how dense a graph can be while avoiding a fixed subgraph. Let F be a fixed finite graph. A graph G on n vertices is said to be F-free if it contains no copy of F as a subgraph. The classical forbidden subgraph problem asks for the maximum number of edges in an F-free graph of order n. For a fixed graph F and integer $n \geq 1$, the maximum number of edges in an F-free graph is called the **extremal number** $\mathrm{ex}(n, F)$. More formally,

$$\mathrm{ex}(n, F) \coloneqq \max\{\, |E(G)| : v(G) = n,\ G \text{ is } F\text{-free} \,\}.$$

If $n < v(F)$, every n-vertex graph is automatically F-free, so

$$\mathrm{ex}(n, F) = \binom{n}{2}$$

Thus, forbidden subgraph problems are most interesting when we consider large n. In practice, it is very difficult to calculate the extremal number exactly. Instead, we consider the asymptotic behavior of the extremal number after normalization. The Turán density of a graph F is the limit

$$\pi(F) \coloneqq \lim_{n\to\infty} \frac{\mathrm{ex}(n, F)}{|E(K_n)|}$$

The value $\pi(F)$ describes the maximum asymptotic edge density of F-free graphs and is central to classical extremal graph theory. The following lemma ensures that the above limit is always defined.

Lemma 3.1. *The sequence* $\frac{\mathrm{ex}(n,F)}{|E(K_n)|}$ *is monotonically decreasing.*

Proof. Let F be a fixed graph and suppose G is a maximal F-free graph on $n+1$ vertices, i.e. $|E(G)| = \mathrm{ex}(n+1, F)$. Since the average degree of G is

$$\overline{\deg(G)} = \frac{2|E(G)|}{n+1},$$

there exists a vertex v with $\deg(v) \leq \overline{\deg(G)}$. The graph $G - v$ is an F-free graph on n vertices such that

$$\begin{aligned} \big|E(G-v)\big| &= |E(G)| - \deg(v) \\ &\geq \mathrm{ex}(n+1,F) - \frac{2\,\mathrm{ex}(n+1,F)}{n+1} \\ &= \mathrm{ex}(n+1,F) \cdot \frac{n-1}{n+1} \end{aligned}$$

By definition of the extremal number $\mathrm{ex}(n, F)$,

$$\mathrm{ex}(n,F) \geq \big|E(G-v)\big| \geq \mathrm{ex}(n+1,F) \cdot \frac{n-1}{n+1}$$

Dividing both sides by $\binom{n+1}{2}$ yields

$$\frac{\mathrm{ex}(n+1,F)}{\binom{n+1}{2}} \leq \frac{\mathrm{ex}(n,F)}{\binom{n}{2}}$$

Thus, $\frac{\mathrm{ex}(n,F)}{|E(K_n)|}$ is monotonically decreasing. □

Since $0 \leq \frac{\mathrm{ex}(n,F)}{|E(K_n)|} \leq 1$, the sequence $\left\{\frac{\mathrm{ex}(n,F)}{|E(K_n)|}\right\}$ is bounded and monotonically decreasing. By the monotone convergence theorem, the limit $\pi(F)$ exists for any F.

3.2 Mantel's Theorem and Early Extremal Theory

One of the earliest results in extremal graph theory is due to Mantel (1907), who resolved the special case of the forbidden subgraph problem when $F = K_3$. At

the time, graph theory was still largely combinatorial and not yet developed as a systematic field. Mantel studied a natural question: how many edges can a triangle-free graph on n vertices contain? His answer, now known as Mantel's theorem [1], is the first nontrivial determination of an extremal number.

Theorem 3.2 (Mantel, 1907)**.** *If a graph on n vertices contains no triangle, then it has at most*

$$\left\lfloor \frac{n^2}{4} \right\rfloor$$

edges, with equality achieved uniquely by the complete bipartite graph $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$.

In the notation of extremal numbers, Mantel's Theorem can be restated as

$$\mathrm{ex}(n, K_3) = \left\lfloor \frac{n^2}{4} \right\rfloor$$

Mantel's argument, although elementary, introduced ideas that would become fundamental to the field: partitioning the vertex set to control the number of edges, and recognizing that extremal examples often exhibit a highly regular, symmetric structure.

The result gained additional importance when Turán later generalized Mantel's argument to forbid larger cliques. In 1941, Turán proved that among all K_{r+1}-free graphs on n vertices, the unique maximizer of the number of edges is the complete r-partite graph with parts as equal as possible. Mantel's theorem is precisely the case $r = 2$ of Turán's more general result. There are many known proofs of Mantel's theorem; we will give a short proof of the extremal bound using a homomorphism-counting argument and defer the uniqueness part to our proof of Turán's theorem.

Proof. First we will set up an identity by counting $\hom(S_2, G)$ two different ways.

Recall from the section discussing graph homomorphisms, that

$$\hom(S_2, G) = \sum_{v \in V(G)} \deg(v)^2$$

Next, we will make use of a fact from analysis, Jensen's inequality. See [8] for a proof of the general statement. Jensen's inequality states that the average of the values of a convex function is at most the function applied to the average of the values. Since $f(x) = x^2$ is convex, we have the inequality

$$\frac{1}{|V(G)|} \sum_{v \in V(G)} \deg(v)^2 \geq \overline{\deg(G)}^2 = \frac{4|E(G)|^2}{|V(G)|^2}$$

On the other hand, we can count $\hom(S_2, G)$ by first picking an edge $uv \in E$ and then picking one of the neighbors of either u or v to complete the star S_2. Hence,

$$\hom(S_2, G) = \sum_{uv \in E(G)} \deg(u) + \deg(v)$$

If G is triangle free, then $N(u) \cap N(v) = \varnothing$ otherwise the mutual neighbor along with u and v would create a triangle. Since the neighborhoods of every pair of adjacent vertices are disjoint,

$$\deg(u) + \deg(v) = |N(u)| + |N(v)| \leq |V(G)|$$

Therefore,

$$\sum_{uv \in E(G)} \deg(u) + \deg(v) \leq \sum_{uv \in E(G)} |V(G)| = |E(G)| \cdot |V(G)|$$

Combining both inequalities

$$\frac{4|E(G)|^2}{|V(G)|} \leq \hom(S_2, G) \leq |E(G)| \cdot |V(G)|$$

Solving for $|E(G)|$ yields the result. □

3.3 Turán's Theorem and Symmetrization Methods

Pál Turán (1910–1976) was a central figure in the development of modern extremal combinatorics. Much of the subject was shaped by his vision of understanding how global properties of a graph are constrained by forbidden local patterns. Turán worked extensively on number theory, analysis, and combinatorics, and his influence on extremal graph theory is so profound that the area is often informally traced back to the publication of his 1941 paper on K_{r+1}-free graphs. His work introduced both the fundamental problem and the structural point of view that still guide the field today.

Turán's theorem is the natural generalization of Mantel's theorem from forbidding triangles to forbidding larger cliques. It asserts that among all graphs on n vertices with no copy of K_{r+1}, the unique maximizer of the number of edges is the complete r-partite graph with parts as equal as possible, the eponymous Turán graph $T(n, r)$. This result not only determines the extremal number $\mathrm{ex}(n, K_{r+1})$ exactly, but also identifies the extremal graph, revealing that symmetry and balance characterize optimal constructions.

There are several proofs of Turán's theorem, each illustrating a different philosophical approach in extremal graph theory. One classical proof uses induction on n, repeatedly removing a vertex of minimum degree and comparing the resulting subgraph to an extremal graph on $n-1$ vertices. Another proof, due to Zykov [4], uses a powerful transformation now known as **Zykov symmetrization**. This method alters the graph in a way that preserves the property of being K_{r+1}-free while increasing

(or at least not decreasing) the number of edges, gradually forcing the graph into an increasingly symmetric form. Ultimately, repeated symmetrization shows that any extremal graph must be a complete multipartite graph, and a final averaging argument identifies the Turán graph as the unique extremal example.

3.3.1 Zykov Symmetrization

Zykov symmetrization is defined as follows. Let G be a graph and let u and v be two nonadjacent vertices. The **symmetrization at the pair** (u, v) constructs a new vertex v' and replaces its neighborhood by that of u. Equivalently, one deletes all edges incident to v and adds edges so that $N(v)$ becomes exactly $N(u)$. This operation makes the vertices u and v "twins" (i.e., sharing the same neighborhood) and never creates a new copy of K_{r+1}. Moreover, when applied carefully, symmetrization never decreases the number of edges. The effect of this transformation is illustrated in Figure 3.1.

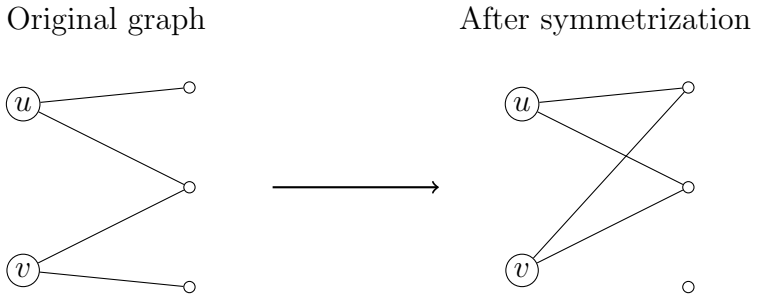


Figure 3.1: Zykov symmetrization of a pair of nonadjacent vertices u and v.

Repeatedly applying this operation to all nonadjacent pairs progressively forces the graph to become more symmetric. In the limit, one obtains a complete multipartite graph whose parts consist of sets of twin vertices. The full proof then shows

that among all such multipartite graphs with no K_{r+1}, the unique maximizer of the number of edges is the balanced r-partite Turán graph. In what follows, we present a proof of Turán's theorem based on Zykov symmetrization.

Theorem 3.3 (Turán, 1941)**.** *Let $r \geq 2$ and let G be a K_{r+1}-free graph on n vertices. Then*

$$|E(G)| \leq |E(T(n,r))|$$

Moreover, $T(n,r)$ is the unique K_{r+1}-free graph on n vertices achieving this maximum.

Proof. We use Zykov symmetrization. Let G be a K_{r+1}-free graph on n vertices with the maximum possible number of edges among all such graphs.

Step 1: Applying symmetrization. Suppose u and v are two nonadjacent vertices of G. Form a new graph G' by *symmetrizing v toward u*: delete all edges incident to v and then add edges so that

$$N_{G'}(v) = N_G(u).$$

Then u and v become twins in G'. We claim:

(i) G' is still K_{r+1}-free,

(ii) $|E(G')| \geq |E(G)|$,

(iii) If $|E(G')| > |E(G)|$, then G was not extremal.

For (i), if a copy of K_{r+1} were created in G', replacing v by u would produce a K_{r+1} in G, a contradiction. For (ii), note that v loses $\deg_G(v)$ edges and gains $\deg_G(u)$ edges. If $\deg_G(v) > \deg_G(u)$, then symmetrization would strictly increase the number of edges, contradicting maximality of G. Thus $\deg_G(u) \geq \deg_G(v)$ and hence $|E(G')| \geq |E(G)|$.

Step 2: Iterating symmetrization. Repeatedly symmetrize all nonadjacent pairs. Since each step preserves K_{r+1}-freeness and never decreases the number of edges, the final graph G^* is also extremal. By construction, every pair of nonadjacent vertices in G^* are twins. Thus the vertex set partitions into classes $V_1, \ldots, V_m$ such that

- each V_i is an independent set,
- every vertex in V_i is adjacent to every vertex outside V_i,
- all vertices in V_i have identical neighborhoods.

Hence G^* is a complete multipartite graph.

Step 4: Optimizing the number of edges. Let the part sizes of the complete multipartite graph G^* be

$$n_1, n_2, \ldots, n_m, \qquad \text{with } m \leq r \text{ and } \sum_{i=1}^{m} n_i = n.$$

Since every edge of a complete multipartite graph runs between two different parts, we have

$$|E(G^*)| = \sum_{1 \leq i < j \leq m} n_i n_j.$$

It is often more convenient to rewrite this expression using the identity

$$\sum_{1 \leq i < j \leq m} n_i n_j = \frac{1}{2} \left(\left(\sum_{i=1}^{m} n_i \right)^2 - \sum_{i=1}^{m} n_i^2 \right) = \frac{1}{2} \left(n^2 - \sum_{i=1}^{m} n_i^2 \right).$$

Thus maximizing the number of edges is *equivalent* to minimizing the sum of squares

$$\sum_{i=1}^{m} n_i^2.$$

The function x^2 is strictly convex on the nonnegative reals, and therefore the sum

$$n_1^2 + n_2^2 + \cdots + n_m^2$$

is minimized, subject to a fixed total $\sum_i n_i = n$, when the variables $n_1, \ldots, n_m$ are as equal as possible. More explicitly, if two parts differ by at least 2, say $n_i \geq n_j + 2$, then replacing the pair (n_i, n_j) by $(n_i - 1,\ n_j + 1)$ strictly decreases the sum of squares:

$$(n_i - 1)^2 + (n_j + 1)^2 = n_i^2 + n_j^2 - 2(n_i - n_j) + 2 < n_i^2 + n_j^2.$$

Hence any multipartite graph in which the part sizes are not as equal as possible cannot be extremal. Since G^* has no K_{r+1}, we have already shown that $m \leq r$. If $m < r$, we may split one part into two nonempty parts to obtain a complete $(m+1)$-partite graph with strictly more edges, contradicting maximality. Therefore $m = r$. Combining these observations:

- the extremal K_{r+1}-free graph must be r-partite,
- the edge count is maximized by an equitable partitiion

But this is exactly the structure of the Turán graph $T(n, r)$. Thus $T(n, r)$ uniquely maximizes the number of edges among all complete r-partite graphs on n vertices, and hence among all K_{r+1}-free graphs. This completes the proof. □

3.4 Erdős–Stone–Simonovits Theory

Turán's theorem provides a complete solution to the extremal problem for forbidden cliques. Turán showed that forbidding a clique K_{r+1} forces the maximum edge density to be

$$\pi(K_{r+1}) = 1 - \frac{1}{r}.$$

Equivalently, every K_{r+1}-free graph on n vertices satisfies

$$|E(G)| \leq \left(1 - \frac{1}{r} + o(1)\right) \binom{n}{2}$$

and the Turán graph $T(n, r)$ asymptotically achieves this bound. In particular, the Turán density of a clique depends only on its chromatic number, not on its size.

This observation leads naturally to a broader question: *Does the chromatic number also determine the asymptotic extremal behavior for general forbidden graphs?* A breakthrough result of Erdős and Stone shows that the answer is yes. By realizing a forbidden graph as an $\chi(F)$-partite graph, we can detect forbidden graphs by detecting Turán graphs. Therefore,

Theorem 3.4 (Erdős-Stone, 1946)**.** *Let F be any fixed graph with chromatic number $\chi(F) = r + 1 \geq 3$. then*

$$\pi(F) = 1 - \frac{1}{r}.$$

In terms of extremal numbers,

$$\mathrm{ex}(n, F) = \left(1 - \frac{1}{r} + o(1)\right)\binom{n}{2}.$$

This result is often called the "fundamental theorem of extremal graph theory": it shows that the asymptotic extremal behavior of a non bipartite forbidden graph is governed solely by its chromatic number. The precise structure of F (e.g.its size, number of edges, or complexity) is irrelevant asymptotically. In this sense, Turán's theorem is not an isolated phenomenon but the first instance of a much more general principle.

Although I will not prove the Erdős–Stone theorem here, it is useful to note the two main methodological traditions behind its proof. Erdős and Stone's original argument relies on a supersaturation principle: once a graph exceeds the Turán density for K_{r+1} by any fixed margin, it must contain many copies of K_{r+1}. A simple

counting argument then forces the appearance of a sufficiently large blow-up of K_{r+1} (a large multipartite graph mimicking the structure of K_{r+1}), and therefore any graph H with $\chi(H) = r + 1$. More recent proofs often use Szemerédi's Regularity Lemma (see [20]). In this approach, a dense graph is approximated by a bounded-complexity partition, and the reduced graph inherits edge densities that allow a direct Turán-type argument at the coarse scale. The embedding lemma then lifts this structure back to the original graph. Both perspectives explain the same phenomenon: forbidding a non-bipartite graph forces the extremal number to approach the Turán density of the appropriate complete multipartite graph, either by a quantitative supersaturation argument or by a structural approximation principle.

When $\chi(F) = 2$, i.e., F is bipartite, the Erdős–Stone theorem says that $\pi(F) = 0$, meaning that the number of edges must be subquadratic. In otherwords, the extremal graphs of F-free graphs must be sparse whenever F is bipartite. Indeed extremal problems for bipartite graphs are substantially more difficult since we do not have an easy method. This dichotomy between the non-bipartite and bipartite cases lies at the heart of extremal graph theory.

The Erdős–Stone theorem determines the asymptotic density of F-free graphs when $\chi(F) \geq 3$, but does not address their structure. A far-reaching refinement, due to Simonovits [6], shows that extremal behavior is also stable. More precisely, if a graph has nearly as many edges as the Turán bound permits, then it must be structurally close to the Turán graph.

Theorem 3.5 (Erdős–Simonovits Stability Theorem)**.** *Let F be a graph with $\chi(F) = r + 1 \geq 3$. For every $\varepsilon > 0$ there exists $\delta > 0$ such that if a graph G on n vertices*

satisfies

$$|E(G)| \geq \left(1 - \frac{1}{r} - \delta\right)\binom{n}{2} \quad \textit{and} \quad G \textit{ is } F\textit{-free},$$

then G can be transformed into the Turán graph $T(n,r)$ by deleting at most εn^2 edges.

In other words, the Turán graph is not only asymptotically extremal but also structurally rigid: any F-free graph whose edge count is close to optimal must resemble $T(n,r)$ after only a small perturbation. The Erdős–Stone and Erdős–Simonovits theorems together form the modern conceptual framework of extremal graph theory, describing both the Turán density and the approximate structure of extremal graphs whenever $\chi(F) \geq 3$.

3.5 Bipartite Extremal Problems

The Erdős–Stone and Erdős–Simonovits theorems show that for every non-bipartite forbidden graph F, the Turán density $\pi(F)$ is determined solely by the chromatic number of F, and that near-extremal graphs are structurally close to the appropriate Turán graph. In striking contrast, the bipartite case $\chi(F) = 2$ is far more delicate. Even very simple bipartite graphs admit extremal functions that are poorly understood, and exact results are known only in special cases.

A foundational result in this direction is the Kővári–Sós–Turán theorem [5], which provides a general upper bound for complete bipartite graphs $K_{s,t}$. It shows that forbidding $K_{s,t}$ forces a polynomial improvement over the trivial quadratic upper bound on the number of edges. Another important result, due to Bondy and Simonovits, gives upper bounds for the extremal numbers of even cycles C_{2k} and plays a central role in the study of bipartite extremal problems.

We begin with the Kővári–Sós–Turán theorem, which we prove using a homomorphism counting argument for stars.

Theorem 3.6 (Kővári–Sós–Turán)**.** *Let $2 \le s \le t$ be integers. Then there exists a constant $C = C(s,t)$ such that*

$$\mathrm{ex}(n, K_{s,t}) \le C\, n^{2-1/s}.$$

Proof. We first record a simple lemma of Erdős.

Claim (Erdős Bipartite Subgraph Lemma). Every graph G has a bipartite subgraph H with

$$|E(H)| \ge \frac{1}{2}|E(G)|.$$

Proof of claim. Randomly partition $V(G)$ into two parts A and B. Each edge has probability $1/2$ of having its endpoints in different parts and hence belonging to the bipartite subgraph $G[A,B]$. Thus

$$\mathbb{E}\,|E(G[A,B])| = \frac{1}{2}|E(G)|,$$

so there exists a partition with $|E(G[A,B])| \ge |E(G)|/2$. Taking $H = G[A,B]$ proves the claim. △

Now let G be a $K_{s,t}$-free graph on n vertices with $|E(G)| = m$. Apply the claim to obtain a bipartite subgraph H of G with

$$|E(H)| \ge \frac{1}{2}m.$$

Since H is a subgraph of G, it is also $K_{s,t}$-free. Thus it suffices to bound $|E(H)|$; the resulting bound for m will differ by at most a factor of 2, which can be absorbed into

the constant C. Write the bipartition of H as $X \cup Y$ with $|X| = a$, $|Y| = b$, and let $d(x)$ denote the degree of $x \in X$. Set $m' = |E(H)|$.

Counting $\hom(S_s, H)$ in two ways.

(1) Center-first. Each homomorphism $S_s \to H$ with center in X corresponds to a choice of s neighbors in Y:

$$\hom(S_s, H) = \sum_{x \in X} \deg(x)^s \geq |X| \cdot \overline{\deg(X)}^s = \frac{(m')^s}{a^{s-1}}$$

where the last inequality follows from Jensen's Inequality. *(2) Leaf-first.* Alternatively, choose an s-subset $S \subseteq Y$ (the leaves) and then a common neighbor in X (the center):

$$\hom(S_s, H) = \sum_{\substack{S \subseteq Y \\ |S| = s}} |N(S)|$$

Because H is $K_{s,t}$-free, any s vertices in Y have at most $t-1$ common neighbors in X, so

$$|N(S)| \leq t - 1 \quad \text{for all } S \subseteq Y,\ |S| = s$$

Hence,

$$\hom(S_s, H) \leq (t-1)\binom{b}{s} \leq (t-1)\frac{b^s}{s!}$$

Combining both bounds, we obtain

$$\frac{(m')^s}{a^{s-1}} \leq (t-1)\frac{b^s}{s!}$$

Since $a + b = n$, we have $a^{\,s-1} b^s \leq C_s\, n^{\,2s-1}$ for some constant C_s depending only on s. Taking s-th roots,

$$m' \leq C'(s,t)\, n^{\,2-1/s}.$$

Finally, $m \le 2m'$ by construction, so

$$m = |E(G)| \le 2C'(s,t)\, n^{\,2-1/s}.$$

Renaming $2C'(s,t)$ as $C(s,t)$ completes the proof. □

The Kővári–Sós–Turán theorem plays a foundational role in bipartite extremal graph theory. A key reason is that it provides an upper bound not only for forbidding a specific complete bipartite graph $K_{s,t}$, but in fact for *any* fixed bipartite graph F. Indeed, if F has bipartition classes of sizes s and t, then F embeds into $K_{s,t}$, and therefore

$$\mathrm{ex}(n,F) \le \mathrm{ex}(n,K_{s,t}) \le C(s,t)\, n^{2-1/s}.$$

Thus Kővári–Sós–Turán yields the general estimate

$$\mathrm{ex}(n,F) = O(n^{\,2-1/s})$$

where s is the size of the smaller color class of F. This shows that forbidding *any* bipartite graph forces a polynomial improvement below the trivial quadratic bound. In contrast with the non-bipartite case of Erdős–Stone–Simonovits, the best universal result known for bipartite forbidden graphs is not tight. A striking example is provided by even cycles. If $F = C_{2k}$, then C_{2k} has bipartition classes of size k, so KST gives

$$\mathrm{ex}(n,C_{2k}) = O\big(n^{\,2-1/k}\big)$$

Bondy and Simonovits proved a significantly stronger bound [7]:

Theorem 3.7 (Bondy–Simonovits)**.** *For each $k \ge 2$, there exists a constant $C(k)$ such that*

$$\mathrm{ex}(n,C_{2k}) \le C(k)\, n^{1+1/k}$$

Comparing the exponents above with the Bondy–Simonovits bound highlights how crude the Kővári–Sós–Turán estimate is for cycles. For example:

$$\text{KST for } C_4: \quad n^{2-1/2} = n^{3/2},$$

$$\text{Bondy–Simonovits:} \quad n^{1+1/2} = n^{3/2} \quad \text{(sharp)},$$

$$\text{KST for } C_6: \quad n^{2-1/3} = n^{5/3},$$

$$\text{Bondy–Simonovits:} \quad n^{1+1/3} = n^{4/3},$$

and the gap widens as k increases:

$$2 - \frac{1}{k} \; - \; \left(1 + \frac{1}{k}\right) = 1 - \frac{2}{k},$$

so the Kővári–Sós–Turán exponent exceeds the best-known exponent by nearly a full power of n when k is large.

This comparison illustrates a central theme in bipartite extremal graph theory: although the Kővári–Sós–Turán theorem provides a universal upper bound that applies to all bipartite graphs, these bounds are often far from optimal, and improving them typically requires much deeper combinatorial or analytic methods. The Bondy–Simonovits theorem is a landmark result demonstrating this phenomenon, especially for even cycles. Nevertheless, for many bipartite graphs—including C_{2k} with $k \geq 3$—the precise order of $\mathrm{ex}(n, F)$ remains unknown, emphasizing the difficulty and richness of bipartite extremal problems.

3.6 Open Problems and the Rational Exponents Conjecture

Despite the progress provided by Kővári–Sós–Turán and Bondy–Simonovits theorems, the general problem of determining extremal numbers for bipartite graphs

remains largely open. Unlike the non-bipartite case, where the Erdős–Stone theorem gives an essentially complete asymptotic theory, the bipartite setting features a wide range of behaviors that cannot be captured by a single parameter such as chromatic number. Several of the central open problems include:

- **The Zarankiewicz problem.** Determining the exact order of $\mathrm{ex}(n, K_{s,t})$ when $s, t \geq 2$ remains open for almost all pairs (s, t), even though KST provides a general upper bound and classical constructions give lower bounds of order $n^{2-1/s}$. The case $K_{2,t}$ is particularly well studied but still unresolved for $t \geq 3$.
- **Even cycles.** For $k \geq 3$, it is unknown whether the Bondy–Simonovits upper bound $O(n^{1+1/k})$ is tight. The correct exponent of $\mathrm{ex}(n, C_{2k})$ is unknown for all $k \geq 3$, with the notable exception of C_4, where both upper and lower bounds are of order $n^{3/2}$.
- **General bipartite graphs.** For arbitrary bipartite F, even determining whether $\mathrm{ex}(n, F) = n^{2-o(1)}$ or $\mathrm{ex}(n, F) = n^{1+o(1)}$ is often unknown. Many natural families of sparse bipartite graphs, (trees, hypercubes, expanders) exhibit extremal behavior that remains poorly understood.

A guiding question in this area, originating with Erdős and Simonovits [6], concerns the possible *exponents* that arise in extremal numbers of bipartite graphs. If one writes

$$\mathrm{ex}(n, F) = \Theta\big(n^{\alpha(F)}\big)$$

whenever such an exponent exists, what values can $\alpha(F)$ attain?

Conjecture (Rational Exponents Conjecture)**.** For every bipartite graph F, if the

limit

$$\lim_{n\to\infty} \frac{\log \mathrm{ex}(n, F)}{\log n}$$

exists, then it is a rational number.

This conjecture is consistent with all known results: the Kővári–Sós–Turán theorem gives rational exponents $2 - 1/s$, the Bondy–Simonovits theorem gives $1 + 1/k$, and all known constructions provide rational lower-bound exponents. However, no general method is known for proving rationality, and even the existence of the exponent $\alpha(F)$ is unclear for many bipartite graphs. Understanding which exponents occur—and whether irrational exponents can arise—remains one of the deepest and most intriguing problems in extremal graph theory.

These questions highlight the dramatic contrast between the non-bipartite and bipartite cases. Whereas the former is completely governed by chromatic number and leads to a uniform asymptotic theory, the latter exhibits a rich diversity of behaviors and resists a unifying principle. This makes bipartite extremal graph theory one of the most vibrant and challenging areas of modern combinatorics.

CHAPTER 4

Extremal Graph Theory on Hamming Graphs

Up to this point, we have focused on the classical Turán problem in which the host graph is the complete graph K_n. In many applications, however, one is interested in extremal problems inside a fixed sparse host graph with additional structure. In this section we begin by considering the n-dimensional hypercube Q_n, which can be viewed as the Hamming graph $H(n, 2)$, and study forbidden subgraph problems of the form

$$\mathrm{ex}(Q_n, H) \coloneqq \max\{\, |E(G)| : G \subseteq Q_n,\ G \text{ is } H\text{-free}\}$$

Since Q_n has 2^n vertices and is n-regular, it has

$$|V(Q_n)| = 2^n, \qquad |E(Q_n)| = n2^{n-1}$$

edges in total. It is therefore natural to normalize by the edge count of the host and consider the extremal density

$$\frac{\mathrm{ex}(Q_n, H)}{|E(Q_n)|}$$

in direct analogy with the Turán density in the complete-graph setting.

4.1 Even Cycles in the Hypercube

In the classical setting with host K_n, forbidding any fixed even cycle C_{2k} leads to a subquadratic extremal number: the Bondy–Simonovits theorem shows that

$$\mathrm{ex}(n, C_{2k}) \le C(k)\, n^{1+1/k},$$

so the normalized density $\mathrm{ex}(n, C_{2k})/\binom{n}{2}$ tends to 0 as $n \to \infty$. In other words, for any fixed even cycle, a typical C_{2k}-free graph on n vertices must be asymptotically

sparse compared to K_n.

Fan Chung initiated the systematic study of analogous questions in the hypercube and obtained a series of striking results on even cycles as forbidden subgraphs of Q_n [10]. She showed, among other things, that:

- There exist C_4-free subgraphs of Q_n with a positive proportion of the edges of the cube, and she proved an upper bound of the form

$$\mathrm{ex}(Q_n, C_4) \leq c_4 \, |E(Q_n)|$$

  for some constant $c_4 < 1$ (later improved to $c_4 \approx 0.6068$ by Balogh, Narayanan, and Skokan in [16]).

- She determined the bounds on the extremal density for C_6:

$$0.25 \leq \lim_{n\to\infty} \frac{\mathrm{ex}(Q_n, C_6)}{|E(Q_n)|} \leq \sqrt{2} - 1$$

  Consequently, the maximum C_6-free subgraph of the cube can also contain a fixed positive fraction of all edges.

- In sharp contrast, for longer even cycles C_{4k} with $k \geq 2$ (i.e. $C_8, C_{12}, \dots$), she proved that

$$\mathrm{ex}(Q_n, C_{4k}) = o\big(|E(Q_n)|\big)$$

  so the extremal density tends to zero as $n \to \infty$.

Thus, when the host is Q_n, the behavior of short even cycles is markedly different from the classical K_n case. The cycles C_4 and C_6 behave as exceptional configurations in the hypercube: unlike in the complete graph, where every fixed even cycle has asymptotically zero Turán density, there are subgraphs of Q_n avoiding

C_4 or C_6 that still capture a constant fraction of all edges of Q_n. Chung's work thus reveals a sharp qualitative transition in the extremal behavior of even cycles within the hypercube, and more broadly illustrates how changing the host graph from K_n to Q_n can dramatically alter the nature of forbidden subgraph problems.

4.2 Forbidden Subgraphs in Hamming Graphs

In the classical Turán problem the host graph is the complete graph K_n, and asymptotics are taken as $n \to \infty$. For the hypercube $Q_n = H(n, 2)$ we already saw that changing the host graph can dramatically alter extremal behavior. In this section we extend the forbidden subgraph viewpoint to the more general family of Hamming graphs $H(n, q) = K_q^n$. Unlike the complete graph and the hypercube, the Hamming graph comes with two natural parameters, the **dimension** n and the **alphabet size** q. When studying extremal questions of the form

$$\mathrm{ex}(H(n,q), F) \coloneqq \max\{\, |E(G)| : G \subseteq H(n,q),\ G \text{ is } F\text{-free}\}$$

one must therefore decide how these two parameters are allowed to vary. In particular, we will be interested in the normalized extremal density

$$\rho_{n,q}(F) \coloneqq \frac{\mathrm{ex}(H(n,q), F)}{|E(H(n,q))|}$$

and in its behavior in different asymptotic regimes. There are two natural ways to let the host graph grow:

- **Fixed dimension, large alphabet.** Fix n and let $q \to \infty$.
- **Fixed alphabet, large dimension.** Fix q (e.g. $q = 2$ for the hypercube) and let $n \to \infty$.

In this thesis, we will only investigate what happens in the first regime fixing n and letting $q \to \infty$. Our goal is to understand the limiting densities

$$\rho_n(F) \coloneqq \lim_{q\to\infty} \rho_{n,q}(F),$$

whenever this limit exists, and to compare them with the classical Turán and Erdős–Stone densities in the complete-graph setting. In this regime we will prove two main results:

- A *Turán-type theorem for Hamming graphs*, which identifies the asymptotic extremal density for forbidding cliques inside $H(n, q)$ as $q \to \infty$.
- An *Erdős–Stone-type theorem for Hamming graphs*, which shows that for fixed dimension n and large q, the asymptotic extremal behavior of F-free subgraphs of $H(n, q)$ is governed by the chromatic number of F, in direct analogy with the classical Erdős–Stone theorem for K_n.

These results show that, at least in the dense regime $q \to \infty$ with n fixed, the Hamming graph behaves in a Turán-theoretic sense much like the complete graph, while still retaining the additional combinatorial structure of its product decomposition.

4.3 Dimensionally Constrained Subgraphs of Hamming Graphs

In this section we establish several structural lemmas that describe how certain dense graphs, notably cliques and clique-like graphs, behave inside the Hamming graph $H(n, q)$. These lemmas will play a central role in proving our Turán-type and Erdős–Stone-type results for the regime in which the alphabet size q tends to infinity

while the dimension n is fixed.

Recall that $H(n,q)$ has vertex set $[q]^n$ and that two vertices are adjacent if and only if they differ in exactly one coordinate. For each coordinate i and choice of $(n-1)$ other coordinates $\mathbf{a} = (a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n)$, the **fiber**

$$F_{i,\mathbf{a}} = \{(a_1, \ldots, a_{i-1}, x, a_{i+1}, \ldots, a_n) : x \in [q]\}$$

induces a copy of the complete graph K_q inside $H(n,q)$.

A graph F is said to be **dimensionally constrained in** $H(n,q)$ if every embedding of F into $H(n,q)$ is contained entirely within a single fiber. A graph F is said to be dimensionally constrained if it is dimensionally constrained in $H(n,q)$ for all $n \geq 2$ and all $q \geq |V(F)|$. Our immediate goal is to demonstrate some examples of dimensionally constrained graphs in $H(n,q)$.

Lemma 4.1. *The following graphs are dimensionally constrained:*

(1) *Complete graphs K_r for $r \geq 2$.*

(2) *Almost complete graphs $K_r - e$ where e is an edge and $r \geq 4$.*

(3) *Turán graphs $T(m,r)$ where $m \geq 5$ and $r \geq 3$.*

Proof. We prove each fact separately.

(1) Complete graphs. Let $r \geq 2$, and suppose $\varphi : K_r \hookrightarrow H(n,q)$ is an embedding. Every pair of adjacent vertices differs in exactly one coordinate. Taking any three vertices u, v, w in this embedding. If u and v differ in only one coordinate and v and w differ in only one other coordinate, then u and w will differ in two coordinates. Hence, any triangle of vertices embedded in $H(n,q)$ must all be in the same fiber. Because K_r has every triple of its vertices forming a triangle, all vertices

in the image must lie in the same fiber. For $r = 2$ this is trivially true: each edge occupies exactly one fiber. Hence K_r is dimensionally constrained.

(2) Almost-complete graphs $K_r - e$. Let $r \geq 4$, and let u, v be the endpoints of the missing edge. The remaining $r-2$ vertices form a $(r-2)$-clique. By part (1), the image of this $(r-2)$-clique lies entirely in a single fiber F. Both u and v are adjacent to every vertex of this clique, so they must agree on all coordinates except possibly the coordinate indexing F; otherwise they would fail to be adjacent to the clique vertices. Thus $\varphi(u)$ and $\varphi(v)$ also lie in the same fiber F. All vertices of $K_r - e$ lie in F, so $K_r - e$ is dimensionally constrained.

(3) Turán graphs. Let $T(m, r)$ be the complete r-partite Turán graph with parts $U_1, \ldots, U_r$ and $r \geq 3$, and let $\varphi : T(m, r) \hookrightarrow H(n, q)$ be an embedding.

Choose one vertex $u_i \in U_i$ from each part. Since vertices from different parts are adjacent, these r vertices induce a K_r, which by part (1) must lie entirely in a single fiber; call it F. Thus at least one vertex of each part belongs to F.

Fix a part U_j and let v be any vertex of U_j. Choose representatives $u_i \in U_i$ for every $i \neq j$, all lying in F. Since v is adjacent to every u_i, the set

$$\{v, u_1, \ldots, u_{j-1}, u_{j+1}, \ldots, u_r\}$$

induces a copy of K_r. By part (1), every embedded K_r lies in a single fiber, and all vertices except possibly v already lie in F, so $\varphi(v)$ must also lie in F.

Since v was arbitrary in U_j, every vertex of U_j lies in F. Repeating this argument for each part shows that every vertex of $T(m, r)$ is mapped into the same fiber F. Hence $T(m, r)$ is dimensionally constrained.

□

The previous lemma shows that several graph families are forced to lie entirely within a single coordinate fiber whenever they appear inside a Hamming graph. This rigidity allows us to reduce many forbidden subgraph questions on $H(n, q)$ to one-dimensional considerations. In particular, once a graph F is known to be dimensionally constrained, the presence or absence of F in a subgraph of $H(n, q)$ is completely determined by its behavior on individual fibers. The following lemma formalizes this reduction.

Lemma 4.2. *Let F be a dimensionally constrained graph and let $G \subseteq H(n, q)$ be a subgraph of a Hamming graph. Then G is F-free if and only if for every coordinate i and every fiber*

$$F_{i,\mathbf{a}} = \{(a_1, \ldots, a_{i-1}, x, a_{i+1}, \ldots, a_n) : x \in [q]\},$$

the induced subgraph $G[F_{i,\mathbf{a}}]$ is F-free.

Proof. ($\Rightarrow$) If G is F-free, then every induced subgraph of G is also F-free. In particular, for each i and each fiber $F_{i,\mathbf{a}}$, the induced subgraph $G[F_{i,\mathbf{a}}]$ is F-free.

($\Leftarrow$) Suppose instead that G contains a copy of F. Since F is dimensionally constrained, any embedding of F into $H(n, q)$ (and hence into G) lies entirely within a single fiber $F_{i,\mathbf{a}}$. Thus $G[F_{i,\mathbf{a}}]$ contains a copy of F for some i and $\mathbf{a}$, contradicting the assumption that every fiber-induced subgraph is F-free. Hence G must be F-free. □

Lemma 4.2 shows that when F is dimensionally constrained, the problem of determining the F-free extremal subgraphs of $H(n, q)$ reduces to understanding the extremal structure inside each fiber. In the case $F = K_r$, this means that a subgraph

$G \subseteq H(n,q)$ is K_r-free if and only if every one-dimensional fiber induces a K_r-free graph. Thus the global extremal problem decomposes cleanly into nq^{n-1} independent classical Turán problems, one per fiber. The key idea in the proof is to apply classical Turán's Theorem separately within each fiber. Since the fibers are edge-disjoint and the symmetrization acts independently on each one, the extremal configuration in $H(n,q)$ is simply the product of the fiber-wise Turán graphs $T(q,r-1)$ across all coordinates. We now record this formally in the following theorem.

Theorem 4.3 (Turán's theorem on Hamming graphs)**.** *Let $n \geq 1$, $q \geq r \geq 2$, and consider the Hamming graph $H(n,q)$. Then*

$$\mathrm{ex}\big(H(n,q),K_r\big) = nq^{n-1}\,\mathrm{ex}(K_q,K_r) = nq^{n-1}\,|E(T(q,r-1))|.$$

In particular, there exists a K_r-free subgraph $G \subseteq H(n,q)$ with

$$|E(G)| = |E(T(q,r-1)^n)| = nq^{n-1}\,|E(T(q,r-1))|,$$

obtained by replacing each fiber $F_{i,\mathbf{a}} \cong K_q$ with a copy of the Turán graph $T(q,r-1)$. Moreover, if $G \subseteq H(n,q)$ is K_r-free and satisfies

$$|E(G)| = \mathrm{ex}\big(H(n,q),K_r\big),$$

then for every coordinate i and every fiber $F_{i,\mathbf{a}}$, the induced subgraph $G[F_{i,\mathbf{a}}]$ is a Turán extremal K_r-free graph on q vertices. Conversely, any subgraph of $H(n,q)$ whose restriction to each fiber is a Turán graph $T(q,r-1)$ is extremal.

Proof. Let $G \subseteq H(n,q)$ be K_r-free. By Lemma 4.1, the clique K_r is dimensionally constrained, so by Lemma 4.2, G is K_r-free if and only if each fiber $F_{i,\mathbf{a}}$ induces a K_r-free subgraph of the complete graph K_q.

Fix a fiber $F_{i,\mathbf{a}}$. Inside this fiber, the induced subgraph $G[F_{i,\mathbf{a}}]$ is a K_r-free graph on q vertices. Applying the classical Turán theorem to $G[F_{i,\mathbf{a}}]$ shows that

$$|E(G[F_{i,\mathbf{a}}])| \leq \mathrm{ex}(K_q, K_r) = |E(T(q, r-1))|,$$

with equality if and only if $G[F_{i,\mathbf{a}}]$ is a Turán extremal K_r-free graph on q vertices (i.e. a copy of $T(q, r-1)$ up to isomorphism).

Since the fibers are pairwise edge-disjoint and cover all edges of $H(n, q)$, summing over all fibers gives

$$|E(G)| = \sum_{i,\mathbf{a}} |E(G[F_{i,\mathbf{a}}])| \leq \sum_{i,\mathbf{a}} \mathrm{ex}(K_q, K_r) = \big(\#\text{fibers}\big)\ \mathrm{ex}(K_q, K_r).$$

There are nq^{n-1} fibers in $H(n, q)$ (one for each choice of coordinate and $(n-1)$ fixed coordinates), so

$$|E(G)| \leq nq^{n-1}\ \mathrm{ex}(K_q, K_r) = nq^{n-1}\,|E(T(q, r-1))|.$$

This shows

$$\mathrm{ex}\big(H(n, q), K_r\big) \leq nq^{n-1}|E(T(q, r-1))|.$$

For the matching lower bound, construct $G^* \subseteq H(n, q)$ by replacing, in each fiber $F_{i,\mathbf{a}} \cong K_q$, the complete graph with a copy of the Turán graph $T(q, r-1)$. By construction, every fiber of G^* is K_r-free, so G^* is K_r-free by Lemma 4.2. Moreover, each fiber contributes exactly $|E(T(q, r-1))|$ edges, so

$$|E(G^*)| = nq^{n-1}\,|E(T(q, r-1))|.$$

Hence

$$\mathrm{ex}\big(H(n, q), K_r\big) \geq |E(G^*)| = nq^{n-1}|E(T(q, r-1))|,$$

and combining this with the upper bound yields the desired equality.

Finally, if G is extremal with $|E(G)| = \mathrm{ex}(H(n,q), K_r)$, then every fiber must achieve the Turán bound, so $G[F_{i,\mathbf{a}}]$ is an extremal K_r-free graph on q vertices for each $i, \mathbf{a}$. Conversely, any choice of extremal K_r-free graphs on each fiber yields a K_r-free subgraph with the maximum possible number of edges. In particular, extremal graphs are precisely those obtained by placing a Turán extremal configuration on every fiber. □

The fiber-wise characterization in Lemma 4.2 extends naturally from cliques to any dimensionally constrained graph F. If $G \subseteq H(n,q)$ is a subgraph and F is dimensionally constrained, then G is F-free if and only if each one-dimensional fiber $F_{i,\mathbf{a}}$ induces an F-free subgraph. In particular, for such F, the global extremal problem on $H(n,q)$ reduces to a collection of independent extremal problems on the complete graphs K_q sitting on each fiber.

Now let F be a fixed graph with chromatic number $\chi(F) = r \geq 3$, and assume F is dimensionally constrained in $H(n,q)$. For each fiber $F_{i,\mathbf{a}} \cong K_q$, the induced subgraph $G[F_{i,\mathbf{a}}]$ is an F-free graph on q vertices. By the classical Erdős–Stone theorem in the complete-graph setting,

$$\mathrm{ex}(K_q, F) = \left(1 - \frac{1}{r-1} + o_q(1)\right)\binom{q}{2},$$

so the maximum number of edges we can retain inside a fiber while avoiding F is asymptotically $(1 - \frac{1}{r-1})\binom{q}{2}$. Since $H(n,q)$ has nq^{n-1} fibers and the fibers are edge-disjoint, the total number of edges in an F-free subgraph of $H(n,q)$ is bounded above

by

$$nq^{n-1}\,\mathrm{ex}(K_q, F) = \left(1 - \frac{1}{r-1} + o_q(1)\right) nq^{n-1}\binom{q}{2}.$$

On the other hand, this bound is asymptotically attained by choosing, independently on each fiber, an extremal F-free subgraph of K_q (for example, a Turán graph when F contains a clique of size r). Recalling that

$$|E(H(n,q))| = nq^{n-1}\binom{q}{2},$$

we see that the extremal density on $H(n,q)$ agrees with the classical Erdős–Stone density on K_n in the regime where n is fixed and $q \to \infty$. We summarize this as follows.

Theorem 4.4 (Erdős–Stone theorem on Hamming graphs)**.** *Let F be a fixed graph with chromatic number $\chi(F) = r \geq 3$, and suppose F is dimensionally constrained in $H(n,q)$. Then, for fixed n and $q \to \infty$,*

$$\mathrm{ex}\big(H(n,q), F\big) = \left(1 - \frac{1}{r-1} + o_q(1)\right) |E(H(n,q))|.$$

Equivalently, the Turán density of F in the Hamming graph $H(n,q)$ in the regime of fixed dimension and large alphabet is

$$\lim_{q\to\infty} \frac{\mathrm{ex}(H(n,q), F)}{|E(H(n,q))|} = 1 - \frac{1}{r-1}.$$

CHAPTER 5

Conclusion and Further Directions

In this thesis we studied forbidden subgraph problems in structured host graphs, with a particular focus on Hamming graphs $H(n, q)$. In the dense regime where the dimension n is fixed and the alphabet size q tends to infinity, we proved Turán-type and Erdős–Stone-type results for graphs F that are dimensionally constrained. The key observation was that, for such F, every copy of F in $H(n, q)$ is forced to lie inside a single fiber, so the global extremal problem decomposes into independent classical extremal problems on each copy of K_q. This allowed us to transfer classical results on K_n to the Hamming graph setting and to describe the extremal configurations as products of fiber-wise Turán graphs. Several natural directions remain open. In this chapter we outline a few possible avenues for further research.

5.1 Even Cycles and Chung-Type Results on Hamming Graphs

Chung's work on even cycles in the hypercube $Q_n = H(n, 2)$ shows that the choice of host graph can dramatically alter extremal behavior: in Q_n, the cycles C_4 and C_6 behave as exceptional configurations with positive Turán density, while longer even cycles C_{4k} have vanishing density. It is natural to ask whether analogous phenomena occur in Hamming graphs $H(n, q)$ when q is large.

Question. For fixed n and large q, what is the asymptotic behavior of

$$\frac{\mathrm{ex}(H(n, q), C_{2k})}{|E(H(n, q))|}$$

for even cycles C_{2k}? In particular, do there exist values of k and n for which C_{2k}

has positive extremal density in $H(n,q)$, and for which values does the normalized extremal number tend to zero as $q \to \infty$?

The methods developed in this thesis rely heavily on F being dimensionally constrained. Even cycles are not dimensionally constrained in general: copies of C_{2k} in $H(n,q)$ may use multiple coordinates in nontrivial ways. Adapting Chung's projection and direction-counting ideas from the hypercube to Hamming graphs with $q > 2$ therefore presents an interesting challenge. One would like to know whether there is a similar dichotomy between short and long cycles in $H(n,q)$, and how this depends on both n and q.

5.2 Extremal Problems in the High-Dimensional, Fixed-q Regime

The main results of this thesis focus on the situation that the dimension n is fixed and the alphabet size q tends to infinity. In this setting, embeddings of many graphs F collapse into single coordinate fibers, and the extremal problem decomposes into independent classical questions on K_q. A very different set of behaviors emerges when the alphabet size q is fixed and the dimension n tends to infinity. This is the setting closest to Chung's work on even cycles in the hypercube $Q_n = H(n,2)$, and many of the extremal phenomena observed in the binary case appear to extend to general Hamming graphs.

Question. For fixed q and $n \to \infty$, what is the asymptotic behavior of

$$\frac{\mathrm{ex}(H(n,q),F)}{|E(H(n,q))|}$$

for various forbidden graphs F? Which graphs F have positive extremal density in high dimension, and for which does the normalized extremal number tend to zero?

Chung's results show that in the hypercube Q_n the cycles C_4 and C_6 have positive extremal density, while longer cycles C_{4k} have vanishing density. For $q > 2$ the situation is more complicated. Copies of C_{2k} may use different coordinate values within the same coordinate, and embeddings can distribute themselves across coordinates in ways that do not occur in the binary case.

Question. For fixed $q \geq 3$, which even cycles C_{2k} satisfy

$$\lim_{n\to\infty} \frac{\mathrm{ex}(H(n,q), C_{2k})}{|E(H(n,q))|} > 0,$$

and for which values of k does this limit vanish? Does the answer depend monotonically on q?

A central feature of extremal problems in the hypercube is that the number of edges in each direction provides a powerful invariant. For $H(n,q)$ each coordinate contributes not only a direction but also a complete graph K_q, creating a richer set of projection behaviors.

Question. Can extremal subgraphs of $H(n,q)$ be analyzed through the distribution of their edges across coordinate directions and across the q possible values within each coordinate? Is there a projection or compression principle analogous to those used by Chung for the hypercube?

Developing such projection theories for general q could lead to new upper and lower bounds for extremal problems in high dimension and provide a unifying framework for understanding extremal problems on Hamming graphs.

5.3 Characterizing Dimensionally Constrained Graphs

Our results exploit the fact that certain graphs (such as cliques, almost-cliques, and sufficiently large Turán graphs) are dimensionally constrained: any embedding into $H(n, q)$ must lie inside a single coordinate fiber. This property is central to the fiber-wise reduction in Lemma 4.2, yet our understanding of which graphs enjoy this property is still quite limited.

Question. Can we characterize the graphs F that are dimensionally constrained in $H(n, q)$? What structural conditions on F (e.g. high minimum degree, large clique number, or some form of local density) guarantee that every copy of F in $H(n, q)$ is contained in a single fiber?

Even partial classifications would be useful. For example, one might ask whether every sufficiently dense graph F with $\chi(F) \geq 3$ is dimensionally constrained when n is fixed and q is large, or whether there exist sparse but highly structured graphs that are dimensionally constrained for combinatorial reasons unrelated to density. Understanding this phenomenon more systematically could substantially extend the range of forbidden graphs F for which our Hamming graph Turán and Erdős–Stone analogues apply.

5.4 Partially Dimensionally Constrained Graphs

A natural relaxation of our main hypothesis is to allow F to occupy a bounded number of coordinates instead of a single fiber. This naturally leads to the notion of a graph beign **k-dimensionally constrained**. A graph F is k-dimensionally constrained in $H(n, q)$ if every embedding of F into $H(n, q)$ uses at most k distinct

coordinates; equivalently, the images of the vertices of F lie in a subgraph isomorphic to $H(k, q)$. The case $k = 1$ corresponds to the dimensionally constrained graphs studied in this thesis. The next natural step is $k = 2$.

Question. Suppose F is 2-dimensionally constrained in $H(n, q)$ and $\chi(F) \geq 3$. In the dense regime of fixed n and $q \to \infty$, can we obtain Turán-type or Erdős–Stone-type results for $\mathrm{ex}(H(n, q), F)$ analogous to the one-dimensional case? In particular, does the extremal problem reduce to an extremal problem on $H(2, q)$, and can we describe the corresponding extremal configurations?

Even in the $k = 2$ case, copies of F may interact across fibers in more complicated ways, and simple fiber-wise symmetrization no longer suffices. However, one might hope to develop a mixed strategy that combines fiber-wise arguments with symmetrization or averaging on two-dimensional slices $H(2, q)$, perhaps using product decompositions or projections onto coordinate planes. Understanding how far the methods of this thesis extend to k-dimensionally constrained graphs is a promising direction for future work.

Taken together, these questions highlight the broader theme of this thesis: extremal problems in structured host graphs are governed not only by global parameters such as chromatic number, but also by the way forbidden configurations interact with the product structure of the host. The results here provide a first step in the dense Hamming-graph regime, and suggest that there is a rich landscape of phenomena yet to be explored.

Bibliography

[1] Willem Mantel. "Vraagstuk xxviii". In: *Wiskundige Opgaven met de Oplossingen* 10.2 (1907).

[2] Paul Turán. "On an extremal problem in graph theory." In: *Matematikai és Fizikai Lapok* 48 (1941), pp. 436–452.

[3] Paul Erdős and Arthur Stone. "On the number of edges in a graph". In: *Bulletin of the American Mathematical Society* 52 (1946).

[4] A. A. Zykov. "On some properties of linear complexes". In: *Matematicheskii Sbornik (N.S.)* 24(66) (1949). English translation: *American Mathematical Society Translations*, Series 2, pp. 163–188.

[5] Tamás Kővári, Vera T. Sós, and Pál Turán. "On a problem of K. Zarankiewicz". In: *Colloquium Mathematicae* 3 (1954), pp. 50–57.

[6] Paul Erdős and Miklós Simonovits. "A Limit Theorem in Graph Theory". In: *Studia Scientiarum Mathematicarum Hungarica* 1 (1966), pp. 51–57.

[7] J. A. Bondy and M. Simonovits. "Cycles of even length in graphs". In: *Journal of Combinatorial Theory, Series B* 16.2 (1974), pp. 97–105.

[8] Walter Rudin. *Real and Complex Analysis*. 3rd ed. New York: McGraw–Hill, 1987.

[9] Rephael Wenger. "Extremal graphs with no C_4's, C_6's, or C_{10}'s". In: *Journal of Combinatorial Theory, Series B* 52.1 (1991), pp. 113–116.

[10] Fan R. K. Chung. "Subgraphs of a hypercube containing no small even cycles". In: *Journal of Graph Theory* 16.3 (1992), pp. 273–286.

[11] Zoltán Füredi. "On the number of edges of quadrilateral-free graphs". In: *Journal of Combinatorial Theory, Series B* 68.1 (1996), pp. 1–6.

[12] János Kollár, Lajos Rónyai, and Tamás Szabó. "Norm-graphs and bipartite Turán numbers". In: *Combinatorica* 16.3 (1996), pp. 399–406.

[13] Noga Alon, Lajos Rónyai, and Tamás Szabó. "Norm-graphs: variations and applications". In: *Journal of Combinatorial Theory, Series B* 76.2 (1999), pp. 280–290.

[14] Ron Milo et al. "Network motifs: simple building blocks of complex networks". In: *Science* 298.5594 (2002), pp. 824–827.

[15] Shalev Itzkovitz and Uri Alon. "Subgraphs and network motifs in geometric networks". In: *Physical Review E* 71.2 (2005).

[16] József Balogh, Benny Sudakov, and Stanislav Skokan. "On the number of edges in C_4-free subgraphs of the hypercube". In: *Journal of Combinatorial Theory, Series A* 115.3 (2008), pp. 505–524.

[17] John A. Bondy and U. S. R. Murty. *Graph Theory*. Vol. 244. Graduate Texts in Mathematics. Springer, 2008.

[18] Shalev Itzkovitz et al. "Geometric constraints on neuronal connectivity facilitate a concise synaptic adhesive code". In: *Proceedings of the National Academy of Sciences of the United States of America* 105.27 (2008).

[19] Zoltán Füredi and Miklós Simonovits. "The history of degenerate (bipartite) extremal graph problems". In: *Erdős Centennial*. Vol. 25. Bolyai Society Mathematical Studies. Springer, Berlin, Heidelberg, 2013, pp. 169–264.

[20] Reinhard Diestel. *Graph Theory*. 5th. Springer Publishing Company, Incorporated, 2017. ISBN: 3662536218.

[21] David Conlon. "Extremal numbers of cycles revisited". In: *American Mathematical Monthly* 128.4 (2021), pp. 297–303. DOI: 10.1080/00029890.2021.1886845.

[22] Francesco Farina and Emma Slade. "Symmetry-driven graph neural networks". In: *arXiv preprint arXiv:2105.14058* (2021).

[23] Alejandro Parada-Mayorga and Alejandro Ribeiro. "Algebraic Neural Networks: Stability to Deformations". In: *IEEE Transactions on Signal Processing* 69.26 (2021).