Notes on Extremal Graph Theory

Ryan Martin

April 5, 2012

Contents

Chapter 1

Prologue

1.1 Apologies

What you see below are notes related to a course that I have given several times in Extremal Graph Theory. I guarantee no accuracy with respect to these notes and I certainly do not guarantee completeness or proper attribution.

This is an early draft and, with any luck and copious funding, some of this can be made into a publishable work and some will just remain as notes.

Please do not distribute this document publicly because of this lack of careful attribution.

1.2 Thanks

Thanks to a number of students who have typed notes from previous incarnations of this course:

> Nikhil Bansal, Shuchi Chawla, Abie Flaxman, Dave Kravitz, Venkatesh Natarajan, Amitabh Sinha, Giacomo Zambelli, Chad Brewbaker, Eric Hansen, Jake Manske, Olga Pryporova, Doug Ray, Tim Zick

Chapter 2

Introduction to Extremal Graph Theory

The fundamental question of any extremal problem is:

How much of something can you have, given a certain constraint?

Indeed many of the fundamental questions of science and philosophy are of this form:

How many economists can there be, given that a lightbulb cannot be changed?[1]

or

How much money can you make, given that you are a mathematician?[2]

Turán's theorem can be viewed as the most basic result of extremal graph theory.

How many edges can an n-vertex graph have, given that it has no k-clique?

Ramsey's theorem, Dirac's theorem and the theorem of Hajnal and Szemerédi are also classical examples of extremal graph theorems and can, thus, be expressed in this same general framework.

In this text, we will take a general overview of extremal graph theory, investigating common techniques and how they apply to some of the more celebrated results in the field.

---

[1] Unbounded. They sit in the dark waiting for the Invisible Hand to do it.

[2] Not nearly enough, just go into a dark room and ask an economist.

2.1 Notation and terminology

2.1.1 General mathematical notation

Let $\mathbb{N}, \mathbb{Z}, \mathbb{R}$ denote the natural numbers, integers and real numbers, respectively.

For a natural number r, the **r-subsets of S** are the subsets of S which have size r. For a set S and natural number r, let $\binom{S}{r}$ denote the family of r-subsets of S and let $\binom{S}{\leq r}$ and $\binom{S}{\geq r}$ denote the subsets of S that are of size at least r and at most r, respectively. If a set S is partitioned into sets $S_1, \ldots, S_k$ then we write $S = S_1 + \cdots + S_k$.

2.1.2 Graph theory notation

By this time, basic notation has become reasonably standardized in graph theory. Nonetheless, we formalize the basic ideas.

A **graph** G is a pair (V, E) in which V is a set and E is a multiset of subsets of V of size at most 2. The members of V are called **vertices** and the set V is called **the vertex set of** G and is denoted $V(G)$ when necessary. The members of E are called **edges** and the set E is called **the edge set of** G and is denoted $E(G)$.

A **multiple edge** is an edge which occurs more than once in the multiset E. A **loop** is an edge which has only one vertex. A graph G is **simple** if it has no multiple edges or loops. Unless stated otherwise, a graph is assumed to be simple.

We use $v(G) = |V(G)|$ and $e(G) = |E(G)|$ as shorthand. Typically, we will say that a graph G has n vertices and m edges. The vertices contained by edge e are the **endvertices** of e. If x and y are connected by edge e, we write $e = xy$ or $e = \{x, y\}$ or $x \sim y$.

If two vertices, v and w, are nonadjacent in graph G, then we write $G + vw$ to mean the graph $(V(G), E(G) \cup \{vw\})$..

A **subgraph** H of a graph $G = (V, E)$ is an injection $\varphi : V(H) \to V(G)$ such that if $h_1 \sim h_2$, then $\varphi(h_1) \sim \varphi(h_2)$. An **induced subgraph** H of a graph $G = (V, E)$ is an injection $\varphi : V(H) \to V(G)$ such that $h_1 \sim h_2$, if and only if $\varphi(h_1) \sim \varphi(h_2)$ and we say that H is a **graph induced by** the image of $V(H)$. If S is such an image than we write $G[S]$ to mean the graph induced by S. For $T \subseteq V(G)$, we write $G - T$ to mean the graph induced by $V(G) - T$.

A bipartite graph G is a graph whose vertex set is partitioned into two pieces $V(G) = X + Y$, usually denoted $G = (X, Y; E)$, so that each edge has one endvertex in X and one endvertex in Y.

The **order of a graph** G is $|V(G)|$, sometimes denoted $\|G\|$ and the **size of a graph** G is $|E(G)|$, denoted $|G|$.

A **directed graph** D (also called a **digraph**) is a pair (V, A) in which V is a set and A is a multiset of ordered pairs of V. The members of V are called **vertices** and the set V is called **the vertex set of** D and is denoted $V(D)$ when necessary. The members of A are called **arcs** and the set A is called **the**

arc set of D and is denoted $A(D)$ when necessary. If $(x, y) \in A$, then we also write $x \to y$.

A property of graphs is merely a set of graphs, but the term "property" comes from the fact that they are usually defined by a characteristic (e.g. planar graphs, triangle-free graphs, perfect graphs). A property, $\mathcal{P}$, is called **monotone increasing** if, whenever $G \in \mathcal{P}$ and $e \notin E(G)$, then $G+e \in \mathcal{P}$. It is called **monotone decreasing** if, whenever $G \in \mathcal{P}$ and $e \in E(G)$, then $G - e \in \mathcal{P}$. If it is either monotone increasing or monotone decreasing, then it is simply **monotone**. A property of graphs $\mathcal{P}$ is **hereditary** if, whenever $G \in \mathcal{P}$, then every induced subgraph of G is in $\mathcal{P}$.

For a positive integer k, we use $[k]$ to denote the set $\{1, \ldots, k\}$.

2.1.3 Some special graph families

The **path** on n vertices and $n-1$ edges is denoted P_n. It is also said to be the path of **length** $n-1$. The **cycle** on n vertices is denoted C_n.

2.1.4 Useful bounds

The binomial coefficients have a few useful bounds. For $1 \leq k \leq n$,

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \frac{n^k}{k!} \leq \left(\frac{en}{k}\right)^k .$$

A precise expression of Stirling's formula is as follows:

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \leq n! \leq e^{1/(12n)}\sqrt{2\pi n}\left(\frac{n}{e}\right)^n .$$

Chapter 3

The Basics

3.1 König-Hall theorem

The 1935 theorem due to Philip Hall is one of the cornerstones of graph theory. A **matching** is a subgraph which is a set of vertex-disjoint edges. A matching is said to **saturate** a vertex set S if each vertex in S is incident to an edge of the matching.

Theorem 3.1.1 (Hall's matching theorem [Hal35]) *Let $G = (A, B; E)$ be a bipartite graph. The graph G has a matching that saturates A if and only if*

$$|N(X)| \geq |X| \qquad \textit{for all} \quad X \subseteq A. \tag{3.1}$$

The condition (3.1) is called **Hall's Condition**.

Proof. It is clear that Hall's condition is necessary to guarantee a matching that saturates A.

Suppose Hall's condition is satisfied but there is no matching that saturates A. Let M be a maximum-sized matching in G. Let a_1 be a vertex in $A \setminus V(M)$. Since M is maximum-sized, it is maximal and all of the neighbors of a_1 must be in $V(M) \cap B$. If a_1 has no neighbors, stop. Otherwise, let one such neighbor be b_1. Let the neighbor of b_1 in M be a_2. If a_2 has a neighbor outside of $V(M)$, stop. If a_1 and a_2 have no more neighbors other than b_1, stop. Otherwise, let one such neighbor be b_2 and let its neighbor in M be a_3. Continue until the algorithm terminates.

If the algorithm terminates because $a_1, \ldots, a_k$ have no neighbors other than $\{b_1, \ldots, b_{k-1}\}$, then let $X = \{a_1, \ldots, a_k\}$. As a result, $|N(X)| = k - 1$ and Hall's condition is violated, a contradiction.

Therefore, termination must have occurred because we found some b_k which is not in $V(M)$. The vertex b_k has some neighbor a_j, $j < k$. The partner of a_j in M, b_j, has a neighbor $a_{j'}$, $j' < j$, and so on, until we conclude with a_1.

Therefore, there is a path, P, as follows

$$a_1 = a_{i_1}, b_{i_2}, a_{i_2}, b_{i_3}, \ldots, a_{i_{t-1}}, b_{i_t} = b_k$$

such that $a_{i_j} b_{i_j} \notin M$ for $j = 1, \ldots, t$ but $b_{i_j} a_{i_{j+1}} \in M$ for $j = 1, \ldots, t-1$. Because this path has vertices alternately in M and not in M, P is called an **M-alternating path**. Because the first and last edges are not in M, P is called an **M-augmenting path**. This terminology is used because we can simply switch the non-M edges with M-edges to get a larger matching. That is, let $M' = M \triangle E(P)$. The fact that M' is a matching of cardinality $|M| + 1$ gives a contradiction to the assumption that M had maximum size. □

One immediate consequence of Hall's theorem is an older theorem. A **perfect matching** is a matching that saturates every vertex. Thus, a perfect matching can only exist in bipartite graphs which have the same number of vertices in the each part of the bipartition. Frobenius [Fro17] proved Theorem 3.1.2 in 1917 and is called the **Marriage Theorem**.

Corollary 3.1.2 (Frobenius [Fro17]) *Let G be a k-regular bipartite graph with n vertices in each part. Then, G has a perfect matching.*

Proof. Let $G = (A, B; E)$ be a k-regular bipartite graph with $|A| = |B| = n$. Let $X \subseteq A$ and since G is k-regular, $e(X, N(X)) = k|X|$ because G is k-regular and, by definition, there are no edges in $(X, B \setminus N(X))$. By counting according to the members of $N(X)$,

$$k|X| = e(X, N(X)) \leq \sum_{v \in N(X)} \deg(v) \leq k|N(X)|,$$

and Hall's condition is satisfied. □

Theorem 3.1.1 is a generalization of Theorem 3.1.3 but we use the weaker Hall's theorem to prove it as a corollary.

Theorem 3.1.3 *Let $G = (A, B; E)$ be a bipartite graph and let*

$$d = \max\left\{|X| - |N(X)| : X \subseteq A\right\}.$$

Then the largest matching in G has size $|A| - d$.

Note that if Hall's condition is satisfied, then $d = 0$ by choosing $X = \emptyset$.

Proof. It is obvious that no matching in G can be larger than $|A| - d$.

Create d new dummy vertices that are adjacent to every vertex in A. The resulting bipartite graph, call it $G' = (A, B'; E')$, satisfies Hall's condition. Thus, it has a matching that saturates A. By deleting the dummy vertices, the resulting matching is of size at least $|A| - d$. □

3.1.4 Equivalent theorems

Since Dénes Kőnig proved an earlier and equivalent theorem 4 years earlier, both Theorem 3.1.1 and Theorem 3.1.5 are often jointly called *König-Hall.*[1]

Theorem 3.1.5 (König's theorem [Kön31]) *Let G be a bipartite graph. The maximum size of a matching in G is equal to the minimum size of a vertex cover in G.*

Below we prove König's theorem using Hall's. The converse is left as an exercise.

Proof. It is clear that if M is a matching and C is a cover, then $|C| \geq |M|$ because one requires at least $|M|$ vertices to cover the edges of M.

Let C be a minimum-sized cover with $C_A = C \cap A$ and $C_B = C \cap B$. There is no edge between $A \setminus C_A$ and $B \setminus C_B$ by the definition of C. Let H_1 be the graph induced by $(C_A, B \setminus C_B)$ and H_2 be the graph induced by $(C_B, A \setminus C_A)$.

Let $X \subseteq C_A$. If $|X| > |N_{H_1}(X)|$, then the set $C_B \cup (C_A \setminus X \cup N_{H_1}(X))$ is a smaller vertex cover because C_B covers every edge not in H_1 and $C_A \setminus X$ covers every edge not incident to X. Those edges are covered by $N(X)$. Hence, H_1 satisfies Hall's condition and so has a matching that saturates C_A. By a symmetric argument, H_2 also satisfies Hall's condition and has a matching that saturates C_B. The union of these matchings produces a matching M of size $|C|$. $\square$

Note that König's theorem does not extend to general graphs. A 5-cycle has maximum matching size 2 and minimum vertex cover size 3.

In addition to Hall's theorem (Theorem 3.1.1), König's theorem (Theorem 3.1.5), there are 6 additional theorems (Theorems 3.1.6, 3.1.7, 3.1.8, 3.1.9, 3.1.10 and 3.1.11) that can be viewed as a restatement of König-Hall.

Menger's theorem is a theorem about general graphs and was the first of these types to appear. In a graph G, for any two vertices, v and w, a ***vw*-separating set** $S \subseteq V(G)$ is a subset of the vertices so that there is no path from v to w in $G \setminus S$.

Theorem 3.1.6 (Menger [Men27]) *The maximum number of vertex-disjoint paths connecting two distinct non-adjacent vertices v and w is equal to the minimum number of vertices in a vw-separating set.*

Egerváry's theorem is expressed in terms of $(0,1)$-matrices. The **term rank** of a $(0,1)$-matrix is the largest number of 1s that can be chosen so that no 2 selected 1s are in the same row or column. A set S of rows and columns is a **cover** of a $(0,1)$-matrix if the matrix has no 1s not in S.

---

[1] Kőnig's name is spelled with the uniquely Hungarian double acute accent, but the theorem attributed to him is spelled with the umlaut. Since *König* is the German word for *king*, one doubts that he would be offended that his theorem has obtained this honorific. The Hungarian word for *king*, however, is *király*.

Theorem 3.1.7 (Egerváry [Ege31]) *The term rank of a* $(0,1)$*-matrix is the size of its smallest cover.*

Hall's theorem can be expressed in set theoretic notation. Let $S = \{S_1, \ldots, S_n\}$ be a family of subsets of a ground set X. Then, a **system of distinct representatives (SDR)** for S is a sequence of distinct elements $\{x_1, \ldots, x_n\}$ of X such that $x_i \in S_i$, $1 \leq i \leq n$.

Theorem 3.1.8 (P. Hall [Hal35]) *The family* S *has an SDR iff the union of any* k *members of* S *contains at least* k *elements.*

The Birkhoff-Von Neumann theorem is a statement of matrix decompositions. A matrix with real nonnegative entries is **doubly stochastic** if the sum of the entries in any row and any column equals one. A **permutation matrix** is a doubly stochastic $(0,1)$-matrix. A matrix $\mathbf{A}$ is a **convex combination** of matrices $\mathbf{A}_1, \ldots, \mathbf{A}_s$ if there exist nonnegative reals $\lambda_1, \ldots, \lambda_s$ such that $\sum_{i=1}^{s} \lambda_i = 1$ and $\mathbf{A} = \sum_{i=1}^{s} \lambda_i \mathbf{A}_i$.

Theorem 3.1.9 (Birkhoff[Bir46]-Von Neumann [vN53]) *Any doubly stochastic matrix can be written as a convex combination of permutation matrices.*

Dilworth's theorem is a theorem on posets. A **partially ordered set (poset)** is a set, together with a relation that is reflexive, antisymmetric and transitive. A **chain** is a set that is totally ordered and an **antichain** is a set of pairwise unrelated elements.

Theorem 3.1.10 (Dilworth [Dil50]) *If* $\mathcal{P}$ *is a finite poset, then the maximum size of an antichain in* $\mathcal{P}$ *equals the minimum number of chains needed to cover the elements of* $\mathcal{P}$*.*

The Max Flow-Min Cut theorem was proved in 1956 by Elias, Feinstein and Shannon and independently by Ford and Fulkerson in the same year. The statement is one of network flows and it is an excellent example of linear programming. A network is a directed graph with a **source** s and a **target** t with each edge assigned an integer called its capacity. An **edge cut** $[S, S']$ is the set of edges directed from S to S'. The **value of an edge cut** is the sum of the capacities. A **flow** is a function f on the arcs in which $f(u,v)$ is at most the capacity of (u,v) and we define $f^+(v) = \sum_u f(v,u)$ (flow out of v) and $f^-(v) = \sum_u f(u,v)$ (flow into v) with the condition that $f^+(v) = f^-(v)$ for all $v \notin \{s,t\}$. The **value of a flow** is $f^+(s) - f^-(s)$.

Theorem 3.1.11 (Max Flow-Min Cut [EFS56, FF56]) *The maximum value of a flow in a network* D *is equal to the value of a minimum cut of* D*.*

Exercises.

(1) Prove Hall's theorem (Theorem 3.1.1) from König's theorem (Theorem 3.1.5).

(2) Prove Menger's theorem (Theorem 3.1.6) from Hall's theorem.

(3) Prove that Egerváry's statement (Theorem 3.1.7) is equivalent to König's.

(4) Prove that Theorem 3.1.1 is equivalent to Theorem 3.1.8

(5) Prove the Birkhoff-Von Neumann theorem (Theorem 3.1.9) from Hall's theorem.

(6) Prove Dilworth's theorem (Theorem 3.1.10) from Hall's theorem.

(7) Prove the Max Flow-Min Cut theorem (Theorem 3.1.11) from Hall's theorem.

3.2 Tutte's theorem

Although König's theorem does not extend to general graphs, Hall's theorem does have an analogue, due to Tutte [Tut47]. A k-factor in a graph is a k-regular spanning subgraph. In particular, a 1-factor is a perfect matching. Let $o(G)$ denote the number of odd-order components of a graph G.

Theorem 3.2.1 (Tutte [Tut47]) *A graph G has a 1-factor iff*

$$o(G-S) \le |S| \qquad \text{for all} \quad S \subseteq V(G). \tag{3.2}$$

Condition 3.2 is known as **Tutte's condition**. The proof presented here is due to Lovász [Lov75].

Proof. We may assume that G is simple, as deleting loops or multiple edges does not effect the presence of a 1-factor nor does it effect Tutte's condition.

Tutte's condition is, indeed, necessary because in order for G to have a 1-factor, at least one vertex from every odd component of $G-S$ must match with some vertex in S.

Note that if G satisfies Tutte's condition (3.2), then adding an edge to G gives $o(G+e-S) \le o(G-S)$ because the number of odd components of a graph will only change if two odd components are attached to each other, reducing the number by 2. Thus, we may assume that G is maximal. That is, adding an edge to G produces a graph with a 1-factor.

By considering $S=\emptyset$, we see that n must be even.

Let U be the set of vertices adjacent to all other vertices. First, suppose $G-U$ consists of complete graphs, find a 1-factor in the even components and for the odd components, find a 1-factor that covers all but one vertex and add the last arbitrarily to a neighbor in U. The remaining vertices are in U and, since that number is even and $G[U]$ is a clique, the matching can be easily completed.

Second, suppose $G - U$ is not a disjoint union of cliques. Therefore, $G - U$ must contain an induced P_3, xyz, where $y \sim x, z$. Furthermore, $y \notin U$ means there is some $w \in G - U$ nonadjacent to y. The maximality of G gives a 1-factor in $G + xz$ and in $G + wy$. Call them M_1 and M_2, respectively.

Let $D = M_1 \triangle M_2$. Each vertex has degree 0 or 2 in D because it has degree 1 in each of M_1 and M_2. Thus, D is a family of disjoint cycles and isolated vertices. Moreover, these cycles must have even length, alternating between members of M_1 and M_2.

Let C be the cycle in D which contains xz. If C does not contain wy, then the 1-factor can be formed by $(M_1 \setminus C) \cup (M_2 \cap C)$.

So we may assume that C contains both xz and wy. As to the next step, add xy and yz to C and then delete xz. What results are two cycles, joined by a vertex. Exactly one of them contains the vertex w. Call C_1 the one that contains w and C_2 the one that doesn't contain w. Let the matching consist of $(C_1 \cap M_1) \cup (C_2 \cap M_2)$ and either xy or yz – whichever belongs to C_1. This completes the 1-factor. □

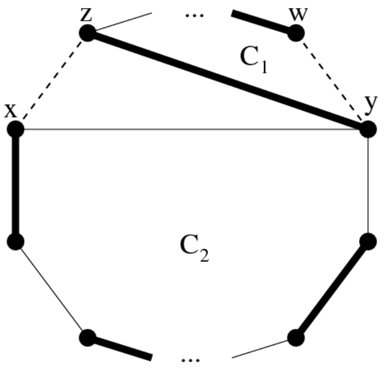

Figure 3.1: C_1 is an even cycle. Edges in bold are in the 1-factor.

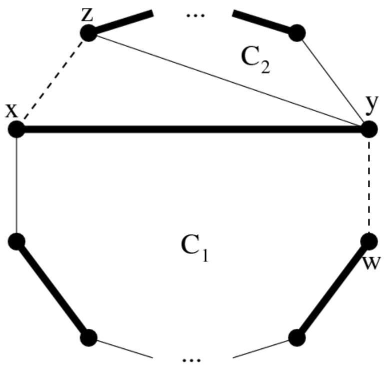

Figure 3.2: C_1 is an odd cycle. Edges in bold are in the 1-factor.

3.3 Turán's theorem

The **Turán graph,** $T_{n,r}$ is a graph on n vertices which is r-partite such that all of the parts differ in size by at most one. That is, each part has size $\lfloor n/r \rfloor$

or $\lceil n/r \rceil$. The Turán number $t_{n,r} = |T_{n,r}|$ is

$$\begin{aligned} t_{n,r} &= \binom{n}{2} - (n - r\lfloor n/r \rfloor)\binom{\lceil n/r \rceil}{2} - (r(\lfloor n/r \rfloor + 1) - n)\binom{\lfloor n/r \rfloor}{2} \\ &= \left(1 - \frac{1}{r}\right)\frac{n^2}{2} - \frac{r}{2}\left(\left\lceil \frac{n}{r} \right\rceil - \frac{n}{r}\right)\left(\frac{n}{r} - \left\lfloor \frac{n}{r} \right\rfloor\right) \\ &\geq \left(1 - \frac{1}{r}\right)\frac{n^2}{2} - \frac{r}{8} \end{aligned}$$

and, of course,

$$t_{n,r} \leq \left(1 - \frac{1}{r}\right)\frac{n^2}{2}.$$

The typical formulation of Turán's theorem [Tur41] is that the maximum number of edges in a graph with no copy of K_{r+1} is $t_{n,r}$. There's a stronger statement and proof due to Erdős [Erd70] from 1970. A sequence $a_1 \geq a_2 \geq \cdots \geq a_n$ is said to **majorize** $b_1 \geq b_2 \geq \ldots \geq b_n$ if $a_i \geq b_i$ for $i = 1, \ldots, n$.

Theorem 3.3.1 (Turán [Tur41]) *Let G be a simple graph on n vertices with no copy of K_{r+1} and degree sequence $d_1 \geq d_2 \geq \cdots \geq d_n$. There exists an r-partite graph on n vertices whose degree sequence majorizes the degree sequence of G.*

Proof. The proof proceeds by induction on r and the case $r = 1$ is trivial. Let $r \geq 1$ and G be a graph on n vertices with no copy of K_{r+1} and degree sequence $d_1 \geq d_2 \geq \cdots \geq d_n$. Let v_1 be a vertex of maximum degree, d_1. If $v_i \in N(v_n)$ has degree d_i, then let $d_i' = |N(v_i) \cap N(v_1)|$. Note that $d_i - d_i' \leq n - |N(v_1)|$.

Since $G[N(v_1)]$ has no copy of K_r, the inductive hypothesis gives that there is an $(r-1)$-partite graph G' on $|N(v_1)|$ vertices whose degree sequence majorizes that of $G[N(v_1)]$. In particular, the vertex corresponding to v_i has degree at least d_i' in G'. Construct G'' by appending $n - |N(v_1)|$ vertices to G', connecting each of them to the vertices in G'.

The vertices not in G' have degree d_1, which is the largest degree-value in G. The vertex in G' corresponding to v_i has degree, in G'', at least $d_i' + n - |N(v_1)| \geq d_i$.

Therefore, G'' is an r-partite graph whose degree sequence majorizes that of G. □

Exercises.

(1) Prove that $T_{n,r}$ is the n-vertex r-partite graph with the most number of edges.

3.4 Königsberg

The famous bridges of Königsberg[2] problem asked if one could traverse the bridges of Königsberg exactly once and return to the same point. Euler proved that it was impossible in 1741 and stated, without proof, that the necessary condition was sufficient. A history of the Königsberg problem can be found in Wilson [Wil86]. An **Eulerian circuit** of a graph G is a circuit (i.e., a closed trail) that contains all the edges of G. A graph with an Eulerian circuit is called an **Eulerian graph**. An **Eulerian trail** is a trail that contains all the edges of G. The proof of Theorem 3.4.1 is due to Hierholzer and Weiner [HW73].

Theorem 3.4.1 *A graph G has an Eulerian circuit iff all vertex degrees are even and all edges belong to a single component.*

A graph G has an Eulerian trail iff G all but 2 vertex degrees are even and all edges belong to a single component.

Proof. The even-degree condition for an Eulerian circuit is clearly necessary. So suppose G is a graph with all even vertex degrees and all edges belonging to a single component and G has at least one edge. Let T be a maximum-length trail in G. It must be a maximal trail and must be a circuit because the endvertices of T must be saturated in T and if they are not the same vertex, then their degrees are odd, a contradiction.

Let $G' = G - E(T)$. Since T is a circuit, G' has all even degrees. If $E(G')$ is nonempty, then the vertex set of each nontrivial component of G' must intersect with the vertex set of C. So, there is some $uv \in E(G')$ that is incident to a vertex $v \in V(C)$. So, $T + uv$ is a trail, which can be seen by beginning with u and then traversing the circuit T, beginning and ending at v. Thus, $T + uv$ is a longer trail than T, a contradiction. Therefore, the first part of Theorem 3.4.1 is proved.

As to the existence of an Eulerian trail, this is easy, given the previous part. Let x and y be the odd-degree vertices. The graph $G+xy$ is Eulerian. Construct its Eulerian circuit C and so $C - xy$ is the Eulerian trail. $\square$

3.5 Dirac's theorem

A **Hamilton cycle** in a graph G is a cycle that contains every vertex. A graph is **Hamiltonian** if it contains a Hamilton cycle.

Theorem 3.5.1 (Dirac [Dir52]) *If G is a simple graph on $n \geq 3$ vertices with minimum degree at least $n/2$, then G is Hamiltonian.*

Ore [Ore60] stated that if $\deg(u) + \deg(v) \geq n$ for all nonadjacent vertices u and v, then the graph is Hamiltonian iff $G + uv$ is Hamiltonian. Combining

---

[2] This is now the city of Kaliningrad, Russia, a province noncontiguous with the rest of the country.

these, we can prove the following generalization of Dirac. Note that the case of K_2, which fulfills the minimum degree condition but is not Hamiltonian is excluded by Theorem 3.5.2.

Theorem 3.5.2 (Ore [Ore60]) *If G is a simple graph on n vertices such that $\deg(u) + \deg(v) \geq n$ for all nonadjacent u and v, then G is Hamiltonian.*

Proof. The degree condition gives that G is connected. (In fact, it gives that the diameter of G is at most 2.) Let $P = v_1, \ldots, v_k$ be a maximum-sized path in G. If $v_k \sim v_1$, then $v_1, \ldots, v_k$ is a cycle and, since G is connected, if G is not Hamiltonian, then there is some vertex not on the cycle adjacent to a vertex on the cycle. This contradicts the maximality of P.

Since $v_1 \not\sim v_k$, $\deg(v_1) + \deg(v_k) \geq n$, but since P is maximal, all of the neighbors of v_1 and of v_k are in P. Let T be the set of neighbors of v_k and let S be the *predecessors* of the neighbors of v_1. That is, $v_i \in S$ iff $v_{i+1} \sim v_1$. Note that $S \subseteq \{v_1, \ldots, v_{k-2}\}$ and $T \subseteq \{v_2, \ldots, v_{k-1}\}$. Since $k \leq n$, $|S \cup T| \leq n - 1$ but $|S| + |T| \geq n$. So, $S \cap T \neq \emptyset$. Let $v_j \in S \cap T$. So, $v_1 \sim v_{j+1}$ and $v_k \sim v_j$. The vertices

$$v_j, v_{j-1}, \ldots, v_1, v_{j+1}, v_{j+2}, \ldots, v_k, v_j$$

form a cycle. Since P was maximum-sized and G is connected, $k = n$ and we have exhibited the Hamilton cycle. □

Exercises.

(1) For each n, find two examples of connected graphs with minimum degree at least $\lceil n/2 \rceil - 1$ which are not Hamiltonian.

(2) Prove that if G is a simple graph on n vertices with minimum-degree at least $n/2$, then G contains a matching (1-regular subgraph) with $\lfloor n/2 \rfloor$ edges.

(3) Prove that, if G is a simple graph with minimum degree at least $n/2$ then there exists a matching with $\lfloor n/2 \rfloor$ edges.

3.6 The Hajnal-Szemerédi theorem

An **equitable k-coloring** of a graph G is a proper coloring of G in k colors such that any two color classes differ in size by at most 1.

In 1963, Corrádi and Hajnal [CAH63] proved that, for every graph G with maximum degree $\Delta(G) \geq 2$, then G has an equitable 3-coloring. In 1964, Erdős conjectured that any graph with maximum degree $\Delta(G) \leq r$ has an equitable $(r + 1)$-coloring.

It is easy to see that this is best possible. Let G be a graph that contains an $(r + 2)$-clique. No matter what the rest of the graph is, even though it can be chosen so that $\Delta(G) = r + 1$, G cannot admit any $(r + 1)$-coloring, let alone an equitable one.

In 1970, Hajnal and Szemerédi [HS70] proved Erdős' conjecture to be correct, although their argument is rather complicated. The proof presented here is due to Kierstead and Kostochka [KK08]. Before we prove the theorem itself, we note that the complementary version is often the form used.

Theorem 3.6.1 (Hajnal-Szemerédi [HS70] – complementary form) *If G is a simple graph on n vertices with minimum degree $\delta(G) \geq \frac{k-1}{k}n$, then G contains a subgraph that consists of $\lfloor n/k \rfloor$ vertex-disjoint copies of K_k.*

The statement of Hajnal-Szemerédi is exactly that of Erdős' conjecture:

Theorem 3.6.2 (Hajnal-Szemerédi [HS70]) *If G is a simple graph on n vertices with maximum degree $\Delta(G) \leq r$, then G has an equitable $(r+1)$-coloring.*

Proof. We may assume that n is divisible by $r+1$ because if $p = n - (r+1)\left\lfloor \frac{n}{r+1} \right\rfloor$, then any equitable coloring of $G + K_p$ induces an equitable coloring of G.

So, let G be a graph on $s(r+1)$ vertices with maximum degree r. We say that G has a **nearly equitable $(r+1)$-coloring**, which is a proper coloring c such that all color classes have size s except one $V^+(c)$, of size $s+1$, and another $V^-(c)$, of size $s-1$.

The proof of the theorem is in two parts. Part I describes the important properties of a nearly equitable $(r+1)$-coloring. Part II shows that our graph G may be assumed to have such a coloring and uses Part I to complete the proof.

Part I. Properties of a nearly equitable coloring. Given a nearly-equitable coloring c, let $D = D(G, c)$ be an auxiliary digraph whose vertices are the color classes[3] of G under c and an arc (X, Y) belongs to $A(D)$ iff some vertex $x \in X$ has no neighbors in Y.

Such a vertex x is said to be **movable** to Y. If there exists a directed path in D from X to V^-, then X is called **accessible**. We also define V^- to be trivially accessible. If V^+ is accessible, it is an easy exercise to see that the proof is finished:

Lemma 3.6.3 *If G has a nearly equitable $(r+1)$-coloring c, for which $V^+(c)$ is accessible, then G has an equitable $(r+1)$-coloring.*

The family of accessible classes is denoted $\mathcal{A}(c)$, $A := \bigcup \mathcal{A}$ and the family of inaccessible classes is denoted $\mathcal{B}(c)$, $B := V(G) - A$. Let m be the number of accessible classes that are not V^-. Let q be the size of $\mathcal{B}$; i.e., q is the number of inaccessible classes. That is,

$$m := |\mathcal{A}| - 1 \qquad \text{and} \qquad q := r - m.$$

---

[3] In order to avoid confusion, we will refer to the vertices of D as "classes". The vertices of G will be lowercase Latin letters and the classes of D will be uppercase Latin letters.

Consequently,

$$|A| = ms + (s-1) \qquad \text{and} \qquad |B| = qs + 1.$$

No vertex $b \in B$ can be moved to A and so b must be adjacent to at least one vertex in every class of $\mathcal{A}$:

$$\deg_A(b) \geq m+1 \quad \Longrightarrow \quad \deg_B(b) \leq q-1 \qquad \text{for all} \quad b \in B. \tag{3.3}$$

If V^- is the only accessible set, then $m = 0$, $q = r$ and

$$e(A, B) \leq r|V^-| = r(s-1) < rs + 1 = |B|,$$

contradicting the first inequality of (3.3).

So we may assume that $2 \leq m + 1 = |\mathcal{A}|$. Call a class $Y \in \mathcal{A}$ **terminal** if V^- is reachable from every class $X \in \mathcal{A} - \{Y\}$ in the digraph $D - Y$. In other words, class Y is terminal if, for every accessible class X, there is a directed path from X to V^- that does not involve Y. Since $m \geq 1$, we can define V^- to be non-terminal.

Every non-terminal X partitions $\mathcal{A} - \{X\}$ into $\mathcal{S}_X$ and $\mathcal{T}_X \neq \emptyset$ where $\mathcal{S}_X$ is the set of classes that are reachable from V^- in $D - X$. Observe that there is no arc from $\mathcal{T}_X - \{X\}$ to $\mathcal{S}_X$.

Choose some non-terminal class U so that $\mathcal{A}' := \mathcal{T}_U \neq \emptyset$ is minimal. Then every class in $\mathcal{A}'$ is terminal, otherwise some class $U' \in \mathcal{T}_U$ could have been chosen rather than U. If that were the case, then $\mathcal{S}_U \supseteq \mathcal{S}_U \cup \{U\}$. See Figure 3.3. Set $t := |\mathcal{A}'|$ and $A' = \bigcup \mathcal{A}'$. Since there are no classes in $\mathcal{A}'$ that point to a class in $\mathcal{A} - (\mathcal{A}' \cup \{U\})$, every vertex in A' must be adjacent to at least one vertex in each color class in $\mathcal{A} - (\mathcal{A}' \cup \{U\})$. Therefore,

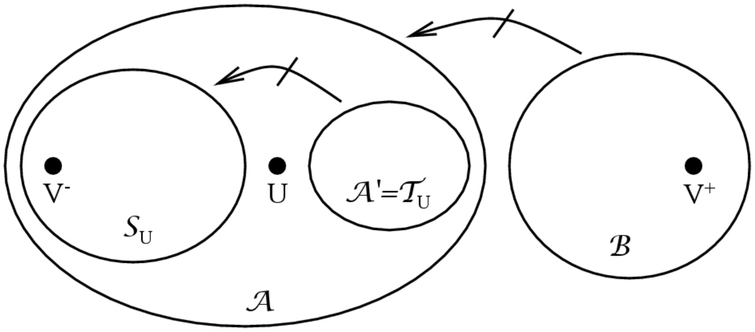

Figure 3.3: Diagram of the digraph D. The class U is a non-terminal class such that $\mathcal{T}_U$ is minimal. $|\mathcal{A}'| = t$, $|\mathcal{S}_U| = m - t$, $|\mathcal{B}| = q = r - m$.

$$\deg_A(a) \geq m - t \qquad \text{for all} \quad a \in A'. \tag{3.4}$$

Let ab be an edge with $a \in W \in \mathcal{A}'$ and $b \in B$. We call ab a **solo edge** if $N_W(b) = \{a\}$. The endvertices of solo edges are **solo vertices** and vertices linked by solo vertices are called **special neighbors** of each other. Let S_a denote the special neighbors in B of $a \in A'$ and S^b denote the set of special neighbors in A' of $b \in B$. Because b must be adjacent to at least one vertex in each member of $\mathcal{A}$, the number of color classes in $\mathcal{A}$ in which b has at least two neighbors is at most $r - (m + 1 + \deg_B(b))$. Consequently, we can lower bound the number of vertices that are special neighbors of $b \in B$:

$$|S^b| \geq t - (r - m - 1 - \deg_B(b)) = t - q + 1 + \deg_B(b). \tag{3.5}$$

Lemma 3.6.4 *If there exists $W \in \mathcal{A}'$ such that no solo vertex in W is movable to a class in $\mathcal{A} - \{W\}$, then $q + 1 \leq t$. Furthermore, every vertex $b \in B$ is solo.*

Proof. Let W_1 be the set of solo vertices in W and $W_2 := W - W_1$. Every vertex in B is adjacent to at least one vertex in W. Furthermore, every vertex in $B - \{N_B(W_1)\}$ is adjacent to at least two vertices in W. Therefore,

$$e(W, B) \geq 2|B| - |N_B(W_1)| = 2(qs + 1) - q|W_1| = qs + q|W_2| + 2.$$

No vertex in W_1 can be moved to another class in $\mathcal{A}$. So, $\deg_B(x) \leq r - m = q$ for all $x \in W_1$. By (3.4), $\deg_B(w) \leq r - (m - t) = q + t$ for all $w \in W$. So,

$$qs + q|W_2| + 2 \leq e(W, B) \leq q|W_1| + (t + q)|W_2| \leq qs + t|W_2|$$

and it follows that $t \geq q + 1$. Furthermore, according to (3.5), if $b \in B$, then

$$|S^b| \geq t - q + 1 + \deg_B(b) \geq t - (t - 1) + 1 + \deg_B(b) \geq 2.$$

Since b has at least 2 special neighbors, it must be a solo vertex. □

Lemma 3.6.5 *There exists a solo vertex $z \in W \in \mathcal{A}'$ such that either z is movable to a class in $\mathcal{A} - \{W\}$ or z has two nonadjacent special neighbors in B.*

Proof. Assume, by way of contradiction, the lemma is not true. By Lemma 3.6.4, every vertex in B is a solo vertex and S_z, the set of special neighbors of z, induces a clique for every solo $z \in W$. Let μ be defined on $E(A', B)$ as follows:

$$\mu(xy) := \begin{cases} \frac{q}{|S_x|}, & \text{if } xy \text{ is a solo edge;} \\ 0, & \text{otherwise.} \end{cases}$$

As a result, for $z \in A'$, it is the case that $\mu(z, B) := \sum_{b \in B} \mu(zb) = |S_z| \frac{q}{|S_z|} = q$ if z is a solo vertex and $\mu(z, B) = 0$ otherwise. Trivially, $\mu(A', B) := \sum_{z \in A} \mu(z, B) \leq q|A'| = qst$.

Now we count $\mu(A', B)$ by summing over the second coordinate. Let $b \in B$ and $c_b := \max\{|S_z| : z \in S^b\}$ Recall that S_z is a clique and use the bound on the right hand side of (3.3), to obtain $c_b - 1 \le \deg_B(b) \le q-1$. Therefore, $c_b \le q$ and along with (3.5), we have

$$\mu(A', b) = \sum_{z \in S^b} \frac{q}{|S_z|} \ge |S^b| \frac{q}{c_b} \ge (t - q + c_b)\frac{q}{c_b} = (t-q)\frac{q}{c_b} + q \ge (t-q) + q = t.$$

Therefore,

$$\mu(A', B) \ge t|B| = t(qs+1) > qst \ge \mu(A', B),$$

a contradiction. □

Part II. Proof of the theorem. To prove the Hajnal-Szemerédi theorem itself, we proceed by a triple induction. The first induction is on r, the second is on $e(G)$ and the third is on q, the number of non-accessible classes of a given nearly-equitable coloring.

The base case of the induction on r, $r = 0$, is trivial, the empty graph has a 1-coloring.

The second induction is on $e(G)$. The base case $e(G) = 0$ is likewise trivial, so suppose the theorem is true for $e(G) - 1$. Let xy be an edge of G. By the induction hypothesis there is an equitable $(r+1)$-coloring of $G - xy$. We are done unless there is a color class V with both x and y. The fact that $\deg(x) \le r$ gives that there is some color class W such that x is movable to W. This gives a nearly equitable coloring c of G. Let $V^- = V - \{x\}$ and $V^+ = W \cup \{x\}$.

We now proceed by a tertiary induction on $q = r - m$. The base step $q = 0$ holds because it would imply that $V^+ \in \mathcal{A}$ and Lemma 3.6.3 allows us to finish. So, by Lemma 3.6.5, there is a solo vertex $z \in W \in \mathcal{A}'$ and a $b_1 \in S_z$ such that either z is movable to an $X \in \mathcal{A} - \{W\}$ or z is not movable in $\mathcal{A}$ and there is another $b_2 \in S_z$ which is nonadjacent to b_1. See Figures 3.4 and 3.5.

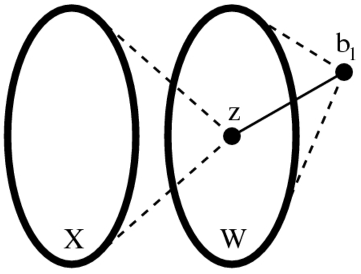

Figure 3.4: Case 1, z is movable to $X \in \mathcal{A}$.

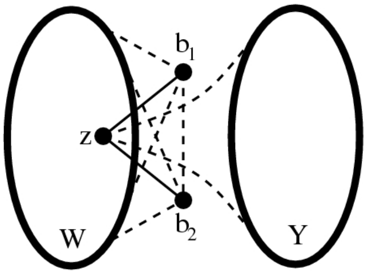

Figure 3.5: Case 2, z is not movable to any class in $\mathcal{A}$ has two nonadjacent special neighbors in B. Further, it is nonadjacent to $Y \in \mathcal{B}$.

We use the first induction hypothesis and (3.3) on $B^- := B - \{b_1\}$. Recall $|B - \{b_1\}| = qs$. Since $\Delta(G[B^-]) \leq q - 1 < r$, this gives an equitable q-coloring – denoted g – of B^-. Set $A^+ := A \cup \{b_1\}$.

Case 1: z is movable to $X \in \mathcal{A}$.
Move z to X and b_1 to $W - \{z\}$ to obtain a nearly equitable $(m+1)$-coloring, γ, of A^+. See Figure 3.4. Since $W \in \mathcal{A}'(c)$, $V^+(\gamma) = X \cup \{z\} \in \mathcal{A}(\gamma)$. But, $V^+(\gamma)$ is accessible, so Lemma 3.6.3 gives that A^+ has an equitable $(m+1)$-coloring γ'. Combine this with the equitable coloring g of B^- and we see that $\gamma' \cup g$ is an equitable $(r+1)$-coloring of G.

Case 2: z is not movable to any class in $\mathcal{A}$.
In this case, $\deg_{A^+}(z) = \deg_A(z) + 1 \geq m + 1$ and so $\deg_{B^-}(z) \leq r - (m+1) = q - 1$. Therefore, there is a color class under g, $Y \in \mathcal{B}^-(g)$ to which z can be added to get a new coloring f' of $B^* := B \cup \{z\} - \{b_1\}$. Also move b_1 to W to obtain a $(m+1)$-coloring φ of $A^* := V(G) - B^*$. See Figure 3.5.

Combine these colorings to obtain $\varphi' = \varphi \cup f'$. Not only is φ' an nearly equitable coloring of G but also W is terminal, so every class in $\mathcal{A} \setminus \{W\}$ is still accessible. Since z was not movable but W was itself accessible, the new class $W^* := W \cup \{b_1\} - \{z\}$ is accessible. Even more, b_2 is movable to W^*, so the class of φ' to which it belongs is also accessible. So, $q(\varphi') < q(c)$ and by the third induction, G has an equitable $(r+1)$-coloring. □

Exercises.

(1) Prove Lemma 3.6.3.

(2) Prove the complementary form of Hajnal-Szemerédi, using the original form.

3.7 Gems: The Hoffman-Singleton theorem

The Hoffman-Singleton theorem is one of the most elegant theorems in extremal graph theory. It is an ideal example of the application of linear algebraic methods.

The **diameter** of a graph G is the largest distance between two vertices and the **girth** is the length of the shortest cycle. Clearly, an r-regular graph with diameter d has at most $1 + r\sum_{i=0}^{d-1}(r-1)^i$ vertices.

The **girth** of a graph G is the length of the shortest cycle. Clearly an r-regular graph with girth $2d+1$ has at most $1 + r\sum_{i=0}^{d-1}(r-1)^i$ vertices.

Hoffman and Singleton [HS60] defined a Moore graph to be an r-regular graph that has diameter d and exactly $1 + r\sum_{i=0}^{d-1}(r-1)^i$ vertices. We leave as an exercise that this is equivalent to having girth $2d+1$ and exactly $1 + r\sum_{i=0}^{d-1}(r-1)^i$ vertices. Singleton [Sin68] proved later that there can be no irregular graph with diameter d and girth $2d+1$:

Theorem 3.7.1 (Singleton [Sin68]) *Let d be a positive integer and G be a graph with diameter d and girth $2d+1$. Then there is a positive integer r such that G is r-regular.*

The theorem of Hoffman and Singleton proved that, in fact, there cannot be many Moore graphs. They proved that none can exist for $d \geq 4$ and for $d = 3$, must be a 7-cycle. As for $d = 2$, they proved the following astounding result.

Theorem 3.7.2 (Hoffman-Singleton [HS60]) *Let G be an r-regular, diameter-2 graph on r^2+1 vertices. Then $r \in \{2, 3, 7, 57\}$.*

We leave as an exercise that the 5-cycle is the unique graph for $r = 2$ and the Petersen graph (see Figure 3.6) is the unique example for $r = 3$. For $r = 7$, it is proven in [HS60] that there is also a unique graph, known as the Hoffman-Singleton graph. It is, of course, 7-regular and has 50 vertices. For $r = 57$, it is not known whether such a graph exists.

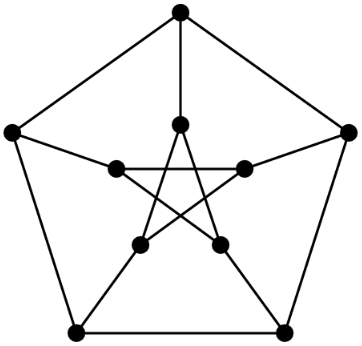

Figure 3.6: The Petersen graph. The vertices can be labeled with members of $\binom{[5]}{2}$ such that there is an edge between two sets if and only if they are disjoint.

Proof. Let G be an r-regular, diameter-2 graph on r^2+1 vertices. Let A denote the adjacency matrix of G. That is, a $(0,1)$-matrix where $a_{ij} = 1$ if and only if vertex v_i is adjacent to vertex v_j. Let $\overline{A}$ be the adjacency matrix of the complement of G. Since the diameter is 2,

$$A + \overline{A} = J - I, \tag{3.6}$$

where I is the $n \times n$ identity matrix and J is the $n \times n$ all ones matrix. The ij^{th} entry of A^2 is the number of length 2 walks between vertices v_i and v_j. Hence,

$$A^2 = \overline{A} + rI. \tag{3.7}$$

Combining (3.6) and (3.7), we obtain

$$A^2 + A - (r-1)I = J.$$

Now let us consider the eigenvalues and eigenvectors. The all-ones vector **1** is an eigenvector of A corresponding to eigenvalue r because the regularity of G. That is, $A\vec{1} = r\vec{1}$. Let $\vec{1}, v_2, \ldots, v_n$ be a set of pairwise orthogonal eigenvectors.

If v_i, $i \geq 2$ is an eigenvector of A corresponding to eigenvalue λ, then $J\vec{1} = \vec{0}$.

$$\begin{aligned} A^2\vec{v_i} + A\vec{v_i} - (r-1)I\vec{v_i} &= J\vec{v_i} \\ (\lambda^2 + \lambda - (r-1))\vec{v_i} &= \vec{0} \end{aligned}$$

Since eigenvectors cannot be $\vec{0}$, we have $\lambda = (-1 \pm \sqrt{4r-3})/2$, neither of which is r.

Let m_1 and m_2 be the multiplicities of $\lambda_1 = (-1 - \sqrt{4r-3})/2$ and $\lambda_2 = (-1 + \sqrt{4r-3})/2$, respectively. Since r has multiplicity 1,

$$1 + m_1 + m_2 = n = r^2 + 1. \tag{3.8}$$

Since the sum of the eigenvalues is equal to the trace, $\operatorname{tr}(A) = 0$ and

$$\begin{aligned} r + m_1\lambda_1 + m_2\lambda_2 &= 0 \\ r + m_1\left(-\frac{1}{2} - \frac{\sqrt{4r-3}}{2}\right) + m_2\left(-\frac{1}{2} + \frac{\sqrt{4r-3}}{2}\right) &= 0 \\ 2r - (m_1 + m_2) + (m_2 - m_1)\sqrt{4r-3} &= 0 \\ 2r - r^2 + \sqrt{4r-3}(m_2 - m_1) &= 0, \end{aligned} \tag{3.9}$$

using (3.8).

Let $s = \sqrt{4r-3}$. Observe that since s is the square root of an integer, it is either irrational or a positive integer. In the case where s is irrational, $m_1 - m_2 = 0$ and (3.9) implies $r = 2$: the 5-cycle.

So we may assume that s is an integer. Substitute $4r - 3 = s^2$ into (3.9):

$$\begin{aligned} 2\left(\frac{s^2+3}{4}\right) - \left(\frac{s^2+3}{4}\right)^2 + s(m_2 - m_1) &= 0 \\ s^4 - 2s^2 + 16(m_1 - m_2)s - 15 &= 0. \end{aligned} \tag{3.10}$$

The only integer solutions of (3.10) must divide 15 and, since s is positive, we have that $s \in \{1, 3, 5, 15\}$, implying $r \in \{1, 3, 7, 57\}$. The case $r = 1$ is a matching and cannot be a radius 2 graph. Combining this with the case where s is irrational, the only possible values of r are in $\{2, 3, 7, 57\}$. $\square$

Babai and Frankl [BF92] refer to Theorem 3.7.2 in the section appropriately entitled "Beauty is Rare" and our proof is similar to theirs, although the original proof of Hoffman and Singleton is not substantially different.

Going back through the Hoffman-Singleton proof, we can derive that the spectrum of A for the missing Moore graph is $57^1(-8)^{1520}7^{1729}$ (multiplicities are in the exponent). Higman (see [Cam99]) proved that the missing Moore graph cannot be vertex-transitive. Mačaj and Širáň [MŠ] proved further that the automorphism group of the missing Moore graph must have order at most

375. Compare this to the Hoffman-Singleton graph, whose automorphism group has order $252,000$ [Haf03].

Exercises.

(1) Prove that the following are equivalent:
   (a) G is r-regular with diameter d and has exactly $1 + r\sum_{i=0}^{d-1}(r-1)^i$ vertices.
   (b) G is r-regular with girth $2d+1$ and has exactly $1 + r\sum_{i=0}^{d-1}(r-1)^i$ vertices.

(2) Prove that the 5-cycle is the unique Moore graph of degree 2 and that the Petersen graph is the unique Moore graph of degree 3.

Chapter 4

Ramsey Theory

4.1 Basic Ramsey theory

The basic graph version of Ramsey's theorem is that, given positive integers k and l, there is an $n = R(k,l)$ such that if $E(K_n)$ are colored red and blue, there is other a red K_k or a blue K_l.

Definition 4.1.1 *The family of r-sets of a set S is denoted $\binom{S}{r}$. A q-**coloring of** $\binom{S}{r}$ is a function $f : \binom{S}{r} \to [q]$. A **homogeneous set** T is a subset $T \subseteq S$ for which each set in $\binom{T}{r}$ is the same color. We say T is i-homogeneous if all of its sets receive color i.*

The notation $n \to (s_1, \dots, s_q)^r$ means that, for every q-coloring of $\binom{[n]}{r}$, there exists an $i \in [q]$ such that there is an i-homogeneous set of size s_i.

The notation $R(k,l)$ is the least number n such that $n \to (k,l)^2$.

Theorem 4.1.2 (Ramsey [Ram30]) *For positive integers q, r and $s_1, \dots, s_q$, there exists an integer n such that $n \to (s_1, \dots, s_q)^r$.*

The proof proceeds by a double induction on r and then $\sum_i s_i$. We neglect the general proof in favor of concentrating on the graph Ramsey bound.

Theorem 4.1.3 *If $k, l \geq 2$ and $n \geq \binom{k+l-2}{k-1}$, then $n \to (k,l)^2$.*

That is, $R(k,l) \leq \binom{k+l-2}{k-1}$.

Proof. We proceed by induction on $k + l$. It is easy to see that $R(k,2) = k$ and $R(2,l) = l$, which suffices for a base case.

Suppose the statement of the theorem is true for $k + l - 1$.

Let $n = R(k,l) - 1$ and color the edges of K_n such that there is no blue K_k or red K_l. Choose any vertex v. The red neighborhood of v has size at most $R(k-1,l) - 1$, otherwise either this neighborhood has a blue l-clique or the red neighborhood has a red $(k-1)$-clique, a contradiction. Similarly, the blue neighborhood of v has size at most $R(k,l-1) - 1$.

Hence, $R(k,l) - 2 = n - 1 \leq (R(k-1,l) - 1) + (R(k,l-1) - 1)$ which simplifies to

$$R(k,l) \leq R(k-1,l) + R(k,l-1) \tag{4.1}$$

and the expression $R(k,l) = \binom{k+l-2}{k-1}$ satisfies both (4.1) as well as the base cases. □

See the dynamic survey by Radziszowski [Rad94] for a summary of known results on Ramsey numbers. Much of the focus on interest in Ramsey theory is on the so-called *diagonal Ramsey numbers*. That is, the numbers $R(k,k)$. According to Stirling's formula (see [Wei]), the binomial coefficient bound has the following asymptotic:

$$R(k,k) \leq \binom{2k-2}{k-1} = (1+o(1)) \frac{\sqrt{2\pi(2k-2)}\left(\frac{2k-2}{e}\right)^{2k-2}}{\left(\sqrt{2\pi(k-1)}\left(\frac{k-1}{e}\right)^{k-1}\right)^2} = \frac{1+o(1)}{4\pi k}\,4^k \leq \frac{C}{\sqrt{k}}\,4^k,$$

for some constant C.

As to the lower bound on the diagonal Ramsey numbers, we leave that for the next chapter.

Exercises.

(1) Prove the general form of Ramsey's theorem.

4.1.4 Infinite Ramsey theory

The infinite version of Ramsey's theorem is as follows:

Theorem 4.1.5 *Given integers r, q and a coloring $c : \binom{\mathbb{N}}{r} \to [q]$, there is an infinite set $M \subseteq \mathbb{N}$ such that $\binom{M}{r}$ is monochromatic.*

Proof. It is sufficient to prove Theorem 4.1.5 in the case where $q = 2$, we leave this as an exercise. Note also that in the statement of the theorem, $\mathbb{N}$ can be replaced by any countably infinite set.

The proof proceeds by induction on r. If $r = 1$, then the statement says that if $c : \mathbb{N} \to [2]$, then there is an infinite monochromatic set, which is obvious.

Suppose, by way of induction, that for $r \geq 1$, M' an infinite set and any coloring $c : \binom{M'}{r} \to [2]$, there is an infinite set M'' such that $\binom{M''}{r}$ is monochromatic. Let c be a coloring of the $(r+1)$-sets of $\mathbb{N}$. Set $Y_0 = \mathbb{N}$.

Choose $x_0 = 1$. The coloring c of the $(r+1)$-sets of Y_0 induces a coloring of the r-sets of Y_0 by giving the set S the color $c(\{x_0\} \cup S)$. By the inductive hypothesis, there is a set $Y_1 \supseteq Y_0$ such that the r-sets of Y_1 all get the same color by this derived coloring. Let $x_1 \in Y_1$.

Repeat this process to get $x_0, x_1, x_2, \ldots$ with the property that if $\{x_{i_1}, \ldots, x_{i_{r+1}}\}$ is an $(r+1)$-set with $i_1 < i_2 < \cdots < i_{r+1}$, then the color of this set only depends on i_1. Since there are only 2 colors, one color occurs for an infinite subsequence of $x_0, x_1, x_2, \dots$ and so that subsequence is M. $\square$

Theorem 4.1.2 follows from Theorem 4.1.5. Theorem 4.1.5. Bollobás [Bol98] notes that this consequence is a special case of Tychonov's theorem:
Proof. Finite Ramsey from infinite Ramsey ($s_1 = \cdots = s_q = l$).
We are given Theorem 4.1.2 and we will suppose, by way of contradiction, that Theorem 4.1.5 is false. Then there exists r, l, q such that for every integer n, there exists a $c : \binom{[n]}{r} \to [q]$ such that there is no monochromatic set of size l.

For every integer n, let $\mathcal{C}_n$ be a nonempty set of k-colorings of $\binom{[n]}{r}$ such that if $n < m$ and $c_m \in \mathcal{C}_m$, then restricting of c_m to $\binom{[n]}{r}$ (denote it to be $c_m|_{[n]}$) is in $\mathcal{C}_n$.

For $m > n$, let $\mathcal{C}_{n,m} \subseteq \mathcal{C}_{n,m}$ be the k-colorings that are restrictions of colorings of $\mathcal{C}_m$. Thus, $\mathcal{C}_n \supset \mathcal{C}_{n,m} \supset \mathcal{C}_{n,m+1}$. For every n, we can define

$$\tilde{\mathcal{C}}_n \stackrel{\text{def}}{=\!=} \bigcap_{m=n+1}^{\infty} \mathcal{C}_{n,m} \neq \emptyset.$$

This is nonempty because each $\mathcal{C}_{n,m}$ is finite.

Thus, we can find $c_r \in \tilde{\mathcal{C}}_r$ and choose $c_{r+1} \in \tilde{\mathcal{C}}_{r+1}, c_{r+2} \in \tilde{\mathcal{C}}_{r+2}, \dots$ with the property that each is the restriction of the previous one. I.e., $c_n = c_{n+1}|_{[n]}$.

This now allows us to define a coloring $c : \binom{\mathbb{N}}{r} \to [k]$ where, for any $S \in \binom{\mathbb{N}}{r}$, S receives the color

$$c(S) = c_n(S) = c_{n+1}(S) = \cdots$$

where n is the maximum element of S. $\square$

Exercises.

(1) Prove that Theorem 4.1.5 holds for an arbitrary q if it holds for $q = 2$.

4.2 Canonical Ramsey theory

Erdős and Rado [ER52] introduced the notion of **canonical Ramsey theory**. In the Ramsey theorem, the number of colors is fixed and if the system is sufficiently large, there is a monochromatic clique (in either the graph or hypergraph sense). The canonical version of Ramsey does not restrict the number of colors but still demonstrates that certain substructures are unavoidable.

Definition 4.2.1 *For any pair of infinite sets N_1, N_2, positive integer r, and colorings $c_1 : \binom{N_1}{r} \to C_1$ and $c_2 : \binom{N_2}{r} \to C_2$, we say c_1 and c_2 are* ***equivalent*** *if there is a one-to-one map $\varphi : N_1 \to N_2$ such that for $e, e' \in \binom{N_1}{r}$, it is the case that $c_1(e) = c_1(e')$ iff $c_2(\varphi(e)) = c_2(\varphi(e'))$.*

For any infinite set N, positive integer r, and coloring $c : \binom{N}{r} \to C$, we say c is ***irreducible*** *if for every infinite subset N_1 of N, the restriction of c to $\binom{N_1}{r}$ is equivalent to c.*

A set $\mathcal{C}$ *of colorings* $\binom{\mathbb{N}}{r} \to \mathbb{N}$ *is* ***unavoidable*** *if for every coloring* $c : \binom{\mathbb{N}}{r} \to \mathbb{N}$*, there is an infinite set* $M \subset \mathbb{N}$ *such that the restriction of* c *to* $\binom{M}{r}$ *is equivalent to a member of* $\mathcal{C}$*.*

It is easy to construct some irreducible colorings of $\binom{\mathbb{N}}{r}$. The monochromatic and rainbow colorings are irreducible. For any $S \in [r]$, define the S**-canonical coloring** $c_S : \binom{N}{r} \to \binom{N}{|S|}$ such that $c_S(e) = e_S$, where $e_S = \{e_i : i \in S\}$.

Observe that $c_\emptyset$ is the monochromatic coloring and $c_{[r]}$ is the rainbow coloring. Further, observe that there are four canonical colorings when $r = 2$. In that case, we can denote $c_{\{1\}}(ij) = i$ and $c_{\{2\}}(ij) = j$.

4.2.2 Infinite version

Let N be an infinite set and recall that, for a positive integer r, $\binom{N}{r}$ denotes the set of r-subsets of N. The general canonical Ramsey theorem is stated as follows:

Theorem 4.2.3 (Erdős-Rado, [ER52]) *Let* r *be a positive integer and* $c : \binom{\mathbb{N}}{r} \to \mathbb{N}$ *be a coloring. Then there is an infinite set* $M \subseteq \mathbb{N}$ *such that the restriction of* c *to* $\binom{M}{r}$ *is canonical.*

A very nice proof of this theorem is given in Bollobás' *Modern Graph Theory* [Bol98]. We will show the proof for $r = 2$, the general proof is not much more difficult.

Proof. $r = 2$.
Let c be a coloring of $\binom{\mathbb{N}}{2}$. There are only a finite number of possible patterns of $\binom{[4]}{2}$, so by the infinite version of Ramsey's theorem, there is an infinite set M such that all 4-sets of M receive the same pattern.

We shall prove that the restriction of c to $\binom{M}{2}$ is canonical. If $c \neq c_{[2]}$, then there are two edges with the same color. Suppose $c(m_i m_j) = c(m_k m_l)$ with $m_i \notin \{m_j, m_k\}$. We cannot assume $i < j$ or $i > j$. Then $c(m_{2i} m_{2j}) = c(m_{2k} m_{2l})$ and $c(m_{2k} m_{2l}) = c(m_{2i+1} m_{2j})$. Thus, $\binom{M}{2}$ has two adjacent edges of the same color.

Case 1. $c(m_i m_j) = c(m_i m_k)$ for some $i < j < k$.
By considering $\{m_i, m_j, m_k, m_{k+1}\}$, the edges 12 and 13 get the same color and so any pair of edges sharing their first vertices get the same color. Thus, there is a coloring $d : M \to \mathbb{N}$ such that if $r < s$, then $c(m_r m_s) = d(m_r)$.

Case 2. $c(m_i m_k) = c(m_j m_k)$ for some $i < j < k$.
Similarly to Case 1, there is a coloring $d : M \to \mathbb{N}$ such that if $r < s$, then $c(m_r m_s) = d(m_s)$.

Case 3. $c(m_i m_j) = c(m_j m_k)$ for some $i < j < k$.
So, $c(m_1 m_3) = c(m_3 m_5)$ and $c(m_2 m_3) = c(m_3 m_4)$. Hence, there are edges of the same color sharing their first vertices and edges of the same color sharing their second vertices. Therefore, there are maps $d_1 : M \to \mathbb{N}$ and $d_2 : M \to \mathbb{N}$ such that if $i < j$ then $c(m_i m_j) = d_1(m_i) = d_2(m_j)$. Hence, any two edges have the same color and $c = c_\emptyset$.

If Case 3 holds, of course $c = c_\emptyset$. If Case 1 holds and Case 3 does not, then $c = c_{\{1\}}$. If Case 2 holds and Case 3 does not, then $c = c_{\{2\}}$. □

4.2.4 Canonical Ramsey numbers

The finite version of canonical Ramsey can be expressed as follows: Given integers r and l, what is the minimum value of n such that any coloring $\binom{[n]}{r}$ has a canonical coloring on some subset of l integers? This number is denoted $\mathrm{ER}(r;l)$, the **Erdős-Rado** numbers.

For graphs, there are only three canonical colorings as described above. They are monochromatic (the coloring $c_\emptyset$), **rainbow** (the coloring $c_{\{1,2\}}$ which colors each edge distinctly) and **lexicographic** (either of the colorings $c_{\{1\}}$ or $c_{\{2\}}$). Simply put, a lexicographic coloring of $E(K_\ell)$ is one such that there is an order on the vertices $v_1, \ldots, v_\ell$ and the color of edge v_iv_j is $\min\{i,j\}$.

Let $\mathrm{tow}_k(l)$ denote a tower of $k-1$ twos and an l in the last exponent. I.e., $2^{2^{\cdot^{\cdot^{2^l}}}}$.

Theorem 4.2.5 (Lefmann-Rödl, [LR95]) *Let r be a positive integer. Then there exist positive constants c_r, C_r such that for all positive integers l with $l \geq l_0(r)$ the following holds:*

$$\begin{aligned} 2^{c_2 \cdot l^2} &\leq \mathrm{ER}(2;l) \leq && 2^{C_2 \cdot l^2 \cdot \log l} \\ \mathrm{tow}_r(c_r \cdot l^2) &\leq \mathrm{ER}(r;l) \leq && \mathrm{tow}_{r+1}\left(C_r \frac{l^{2r-1}}{\log l}\right) \end{aligned}$$

The proof of the upper bound is not easy, but the lower bound is due to a previous paper of Lefmann and Rödl [LR93].

In order to show this lower bound for $r = 2$, we need some results on Ramsey numbers. Denote $n = R_q(r;l)$ to mean that $n \to (\underbrace{l, \ldots, l}_{q})^r$. That is, any coloring of the r-sets of n with q colors yields a monochromatic set of order l. The following theorem is due to several papers: Erdős-Rado [ER52], Erdős-Hajnal-Rado [EHR65] and Erdős-Hajnal [EH89].

Theorem 4.2.6 *Let r, q be positive integers with $r \geq 3$ and $q \geq 2$. Then there exist positive constants $c_{r,q}, C_{r,q}$ such that if $l \geq l_0(r)$,*

$$\begin{aligned} R_q(r;l) &\leq \mathrm{tow}_r(C_{r,q} \cdot l) \\ R_q(r;l) &\geq \begin{cases} \mathrm{tow}_r(c_{r,q} \cdot l), & \text{if } q \geq 4; \\ \mathrm{tow}_{r-1}(c_{r,3} \cdot l^2 \cdot \log l), & \text{if } q = 3; \\ \mathrm{tow}_{r-1}(c_{r,2} \cdot l^2), & \text{if } q = 2. \end{cases} \end{aligned}$$

The following theorem is due to several papers: Erdős, [Erd47], Erdős-Szemerédi, [ES72] and Lefmann, [Lef87].

Theorem 4.2.7 *There exist constants* c, C *such that*

$$2^{c \cdot l \cdot q} \leq R_q(2; l) \leq 2^{C \cdot l \cdot q \cdot \log q}$$

for all $l \geq 3$.

To prove the lower bound of Theorem 4.2.5, we need the following claim:

Claim 4.2.8 $\mathrm{ER}(r; l) \geq R_{l-r}(r; l)$.

Proof. First some notation: If $S = \{s_1, \ldots, s_r\}$ is an ordered set ($s_1 < s_2 < \cdots < s_r$)and $I \subseteq [r]$, then $S|_I = \{s_i : i \in I\}$.

Let $n = R_{l-r}(r; l) - 1$ and $c : \binom{[n]}{r} \to [l - r]$ be a coloring that has no monochromatic l-subset of $[n]$. Suppose there is an l-subset $X \subseteq [n]$ which has a canonical coloring of $\binom{X}{r}$. In order for this to happen, there must be a set $I \subseteq [r]$, $I \neq \emptyset$ such that $c(S) = c(T)$ iff $S|_I = T|_I$. Since c only has $l - r$ colors, no such canonical coloring can exist because it would require at least $l - r + 1$ colors.

All that remains is the monochromatic coloring, which is impossible because c has no monochromatic l-subset of $[n]$. □

Using the fact that $R_q(2; l) \geq 2^{c_3 \cdot l \cdot q}$ for all positive integers $l \geq 3$ where c_3 is a positive constant. Claim 4.2.8 gives that $\mathrm{ER}(2; l) \geq R_{l-2}(2; l) \geq 2^{c' \cdot l^2}$ for some constant c'.

Exercises.

(1) For any $S \in [r]$, prove that the S-**canonical coloring** is an irreducible coloring.

(2) Prove that $\mathrm{ER}(1; l) = (l - 1)^2 + 1$. Lefmann and Rödl say that this result is folklore.

Chapter 5

The Power of Probability

Some of the most useful techniques in extremal graph theory are probabilistic in nature. Sometimes the use of probability is explicit, sometimes it is implicit. An important resource on this methodology is the text of Alon and Spencer [AS00].

We will try to rely on an intuitive understanding of probability, but we will present the formal definitions because readers who are familiar with analysis will find them enlightening and would provide some additional context.

5.1 Probability spaces

5.1.1 Formal definitions

In the most general setting, a **probability space**, Ω, is a measure space with measure Pr, such that $\Pr(\Omega) = 1$. An **outcome** in the probability space Ω is some $\omega \in \Omega$. An **event** in Ω is a subset $A \subseteq \Omega$. We say that a sequence of events $A_1, A_2, \ldots$ occurs with high probability (**whp**) if $\lim_{n\to\infty} \Pr(A_n) = 1$. A **random variable** X is a bounded, measurable function on Ω and, unless otherwise stated, is real-valued. That is, $X : \Omega \to \mathbb{R}$. The **distribution function** of real-valued random variable X, is $F(x) = \Pr(X \leq x)$. Two random variables are said to have the **same distribution** if they have the same distribution function.

A **probability mass function** of a random variable X p_i is a function of the real numbers with a countable domain D such that $\sum_{i\in D} p_i = 1$. A **probability density function** $p(x)$ over the reals is a nonnegative-valued function such that $\int_{-\infty}^{\infty} p(y)\,dy = 1$. Note that $F(x) = \int_{-\infty}^{x} p(y)\,dy$ is a distribution function. In this chapter, we will use something similar to the usual set theory notation for probability. For example, if A_1 and A_2 are events, then the common event is denoted $A_1 \cap A_2$. It is usual in probability to denote this as $A_1 \wedge A_2$.

5.1.2 Probability in a discrete setting

The formal setting is not necessary for our understanding. We often use a **discrete random variable**, which can be viewed as a countable (usually finite) number of values in a range space, say $\{x_i\}_{i\geq 1}$. Each value has a probability associated with it, denoted $\Pr(x_i)$. For this to be a probability space, it is required that $\sum_i \Pr(x_i) = 1$. The function $p(i) = \Pr(x_i)$ is a probability mass function. An event is simply a possible outcome. or a subset of the possible values. The probability of event A is simply the sum of the probabilities of the individual possible outcomes. For example, if $A = \{X \in \{x_2, x_3, x_5\}\}$, then $\Pr(A) = \Pr(X = x_2) + \Pr(X = x_3) + \Pr(X = x_5)$.

For our purposes, almost every random variable will be a discrete random variable, although the careful and diligent reader can prove all of the theorems using measure theory.

5.1.3 Mean and variance

The **expectation** or **mean** of real-valued random variable X is defined to be $\mathbb{E}[X] = \int_\Omega x\, d\Pr$, if the integral is finite. If X is a discrete random variable that takes on values x_i, $i \geq 1$, then

$$\mathbb{E}[X] = \sum_i x_i \Pr(X = x_i).$$

If we say that "X is a random variable with mean μ," this will imply that μ exists and is finite. We can compute $\mathbb{E}[f(X)]$ for any f that makes the following summation finite:

$$\mathbb{E}[f(X)] = \sum_i f(x_i) \Pr(X = x_i).$$

This kind of notation will be useful in random graph theory because the random variables themselves are graph-valued, but f will be a real-valued function.

If A is an event, we denote $\mathbf{1}_A$ to be the **indicator variable** of A. That is $\mathbf{1}_A = 1$ if the outcome of the random experiment is in A and is zero otherwise. So, $\mathbb{E}\left[\mathbf{1}_A\right] = \Pr(A)$.

For any random variable X with mean μ, the **variance of** X, $\mathrm{Var}(X) := \mathbb{E}\left[(X-\mu)^2\right]$, provided such an expression is finite. Otherwise, $\mathrm{Var}(X) = \infty$. The proof of Proposition 5.1.4 is left as an exercise.

Proposition 5.1.4 *Let X be a random variable with finite variance and a and b be real numbers. Then,*

$$\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X).$$

5.1.5 Independence

Finally, we say that the events $A_1, \ldots, A_n$ are **(mutually) independent** if

$$\Pr\left(\bigwedge_{i\in S} A_i\right) = \prod_{i\in S}^{n} \Pr(A_i) \qquad \forall S \subseteq [n].$$

A weaker condition is that the events $A_1, \ldots, A_n$ are **pairwise independent** if, for all distinct $i, j \in [n]$,

$$\Pr\left(A_i \wedge A_j\right) = \Pr(A_i)\Pr(A_j).$$

A set of variables $X_1, \ldots, X_n$ are **independent** if the events $\{X_i \in S_i\}_{i\in I}$ are independent for all subsets S_i and all $I \subseteq [n]$. A set of variables $X_1, \ldots, X_n$ are **pairwise independent** if the events $\{X_i \in S_i\}_{i\in I}$ are pairwise independent for all subsets S_i and all $I \subseteq [n]$.

Exercises.

(1) Prove Proposition 5.1.4.

(2) Prove that, for a random variable X, $\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2$.

(3) Prove that, for discrete independent random variables X and Y, that $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. (**Note:** This is true even if the random variables are not discrete.)

(4) Use the previous exercise to prove that if $X_1, \ldots, X_n$ are pairwise independent random variables, then

$$\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{Var}(X_i).$$

5.1.6 Expectation

The most basic fact of expectation is that there is an outcome that is at least as large as $\mathbb{E}[X]$ and an outcome that is at most as small as $\mathbb{E}[X]$.

Proposition 5.1.7 *If X be a real-valued random variable on a probability space and $\mathbb{E}[X] \geq m$, then $\Pr(X \geq m) > 0$. In the case of a discrete random variable, that means that there is an $x \geq m$ such that $\Pr(X = x) > 0$.*

Symmetrically, if $\mathbb{E}[X] \leq m$, then $\Pr(X \leq m) > 0$.

Proof. Suppose, by way of contradiction, that $\Pr(X \geq m) = 0$. There exists an $\epsilon > 0$ such that $\Pr(X \leq m - \epsilon) \geq 1/2$. Then,

$$\begin{aligned} m &\leq \mathbb{E}[X] \leq (m-\epsilon)\Pr(X \leq m-\epsilon) + m\Pr(m-\epsilon < X < m) \\ &\leq m - \epsilon/2, \end{aligned}$$

a contradiction. □

5.2 Linearity of expectation

There is surprising power in using the linearity of the expectation function. We leave the proof of Proposition 5.2.1 as an exercise.

Proposition 5.2.1 *Let* $a_1, \dots, a_n$ *be real numbers and* $X_1, \dots, X_n$ *be random variables on the same probability space. Then,*

$$\mathbb{E}[a_1X_1 + \cdots + a_nX_n] = a_1\mathbb{E}[X_1] + \cdots + a_n\mathbb{E}[X_n].$$

In particular, if $X = \sum_i \mathbf{1}_{A_i}$, then $\mathbb{E}[X] = \sum_i \Pr(A_i)$.

5.2.2 A lower bound for diagonal Ramsey numbers

In the previous chapter, we establish that an upper bound on the Ramsey number $R(k,k)$ is $\frac{C}{\sqrt{k}}4^k$, for some constant C. The lower bound, first due to Erdős [Erd47], is remarkably simple and is often cited as an early example of the probabilistic method.

Theorem 5.2.3 (Erdős [Erd47])

$$R(k,k) > (1-o(1))\frac{k}{e\sqrt{2}}\left(\sqrt{2}\right)^k.$$

Proof. For each edge in K_n, color it red or blue, independently, with probability 1/2. Let X be the random variable representing the number of homogeneous sets in this random coloring. I.e., for every $S \in \binom{[n]}{k}$, $X = \sum_S \mathbf{1}_{\{S \text{ is monochromatic}\}}$.

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}\left[\sum_{S\in\binom{[n]}{k}} \mathbf{1}_{\{S \text{ is monochromatic}\}}\right] \\
&= \sum_{S\in\binom{[n]}{k}} \mathbb{E}\left[\mathbf{1}_{\{S \text{ is monochromatic}\}}\right] = \sum_{S\in\binom{[n]}{k}} \Pr\left(S \text{ is monochromatic}\right) \\
&\le \binom{n}{k}2^{1-\binom{k}{2}} \le \left(\frac{en}{k}\right)^k 2^{1-\binom{k}{2}}. \qquad (5.1)
\end{aligned}$$

If this expectation is less than 1, then there is a coloring of the edges, red and blue, with no homogeneous set of size k and, thus, $R(k,k) > n$.

For any fixed $\epsilon > 0$, plugging $n = (1-\epsilon)\frac{k}{e\sqrt{2}}2^{k/2}$ will result in the expression in (5.1) beingless than 1, for k large enough. ☐

There have been small improvements on the general bounds on $R(k,k)$, the so-called diagonal Ramsey numbers, but there has not been movement since 1947 on the following:

$$\sqrt{2} \le \liminf_{k\to\infty} R(k,k)^{1/k} \le \limsup_{k\to\infty} R(k,k)^{1/k} \le 4.$$

5.2.4 Finding a dense bipartite subgraph

Before we begin, let us consider the following theorem:

Theorem 5.2.5 *Let G be a simple graph with n vertices and e edges. There is a partition of the vertex set of G, say $V(G) = A + B$ such that at least half of the edges of G have one endpoint in A and one endpoint in B.*

Proof. (#1)
A typical proof of this theorem is algorithmic in nature. Let G be a simple graph and $V(G) = A_0 + B_0$ be a bipartition. Let $e(A_0, B_0)$ denote the number of edges with one endpoint in A_0 and one endpoint in B_0. If there exists a $v \in A_0$ such that $|N(v) \cap A_0| > |N(v) \cap B_0|$, then create a new bipartition where $A_1 = A_0 \setminus \{v\}$ and $B_1 = B_0 \cup \{v\}$. The number of edges between A_1 and B_1 is

$$e(A_1, B_1) = e(A_0, B_0) - |N(v) \cap B_0| + |N(v) \cap A_0|.$$

As long as a partition (A_i, B_i) has a vertex whose neighborhood in its own part is larger than its neighborhood in the other part, then there is a partition (A_{i+1}, B_{i+1}) such that $e(A_{i+1}, B_{i+1}) > e(A_i, B_i)$. Since $e(A_i, B_i) \leq e(G)$ for all i, this algorithm will terminate. Upon termination, there is a partition (A, B) such that either $e(A, B) = e(G)$ or, for every vertex $a \in A$, $|N(a) \cap B| \geq |N(a)|/2$ and for every vertex $b \in B$, $|N(b) \cap A| \geq |N(b)|/2$. Therefore,

$$2e(A, B) = \sum_{a \in A} |N(a) \cap B| + \sum_{b \in B} |N(b) \cap A| \geq \frac{1}{2} \sum_{a \in A} |N(a)| + \frac{1}{2} \sum_{b \in B} |N(b)| = e(G),$$

and the result follows. □

Proof. (#2)
A probabilistic proof is quite a bit shorter. Independently color each vertex $v \in V(G)$ blue with probability 1/2 and red with probability 1/2. The probability that any edge is monochromatic is 1/2. Let $\mathbf{1}_e$ be the indicator of the event that edge e is bichromatic. That is, $\mathbf{1}_e = 1$ if e is bichromatic and $\mathbf{1}_e = 0$ otherwise. The expected number of bichromatic edges is

$$\mathbb{E}\left[\sum_{e \in E(G)} \mathbf{1}_e\right] = \sum_{e \in E(G)} \mathbb{E}\left[\mathbf{1}_e\right] = \sum_{e \in E(G)} \Pr\left(e \text{ is bichromatic}\right) = e(G)/2.$$

Therefore, there exists a bicoloring (i.e., a bipartition) of the vertex set so that the number of bichromatic edges is at least $e(G)/2$. □

While Proof 2 can be generalized to multipartitions, so can Proof 1. In fact, the approach from Proof 1 can be used to solve the exercises.

5.2.6 Dominating sets

A **dominating set**, S, in a graph G is a set $S \subseteq V(G)$ such that for every $v \in V(G)$, either $v \in S$ or there exists an $s \in S$ for which $v \sim s$. The size of the smallest dominating set in a graph G is called the **domination number** of G and is sometimes denoted $D(G)$.

We leave the proof of the following as an exercise.

Proposition 5.2.7 *Let G be a graph on n vertices with no isolated vertices. Then G has a dominating set of size at most $\lfloor n/2 \rfloor$.*

Alon and Spencer [AS00] prove the following theorem.

Theorem 5.2.8 (Alon-Spencer [AS00]) *Let G be a graph on n vertices with a minimum degree at least $\delta > 1$. There exists a dominating set of size at most $n\frac{1+\ln(\delta+1)}{\delta+1}$.*

Note that the theorem is true for $\delta = 1$ also, but is unnecessary because of Proposition 5.2.7.

Proof. Choose the members of the set S at random, each with probability p and independently. Let $T = V(G) \setminus (S \cup N(S))$. Clearly $S \cup T$ is a dominating set in G. It is also easy to see that $\mathbb{E}[|S|] = np$. By linearity of expectation,

$$\mathbb{E}[|T|] = \sum_{v \in V(G)} \mathbb{E}\left[\mathbf{1}_{v \in T}\right] = \sum_{v \in V(G)} (1-p)^{\deg(v)+1} \leq n(1-p)^{\delta+1}.$$

By another application of linearity of expectation,

$$\mathbb{E}[|S| + |T|] = \mathbb{E}[|S|] + \mathbb{E}[|T|] = np + n(1-p)^{\delta+1} \leq n\left(p + e^{-p(\delta+1)}\right).$$

The right-hand side is minimized at $p = \frac{\ln(\delta+1)}{\delta+1}$ and so there is a graph for which $|S \cup T| = |S| + |T|$ is at most $n\frac{1+\ln(\delta+1)}{\delta+1}$. $\square$

Exercises.

(1) Let G be a 3-regular graph. Prove that there exists a bipartition of $V(G) = A + B$ such that at least two-thirds of the edges of G have on endvertex in A and one endvertex in B.

(2) Let G be a graph on n vertices with degree sequence $d_1 \leq d_2 \leq \cdots \leq d_n$. Prove that there is a partition of $V(G) = A_1, \ldots, A_k$ such that at least

$$\frac{1}{2}\sum_{i=1}^{n}\left\lceil \frac{k-1}{k} d_i \right\rceil = e(G) - \frac{1}{2}\sum_{i=1}^{n}\left\lfloor \frac{d_i}{k} \right\rfloor$$

edges have endpoints in distinct parts A_i and A_j.

(3) Prove Proposition 5.2.1.

(4) Prove Proposition 5.2.7.

(5) For any positive integer n and any δ, $1 \leq \delta \leq n-1$, construct a simple graph on n vertices with minimum degree n and no dominating set smaller than $\left\lfloor \frac{n}{\delta+1} \right\rfloor$.

(6) For any positive integer n and any even δ, $1 \leq \delta \leq n-1$, construct a simple graph on n vertices with minimum degree δ and no dominating set smaller than $2\left\lfloor \frac{n}{\delta+2} \right\rfloor$.

5.3 Useful bounds

Boole's inequality is among the simplest and arises from inclusion-exclusion.

Proposition 5.3.1 (Boole's Inequality) *Let $A_1, \ldots, A_n$ be events in a probability space.*

$$\Pr\left(\bigvee_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \Pr(A_i).$$

Proof. By induction it is easy to see that it is sufficient to prove the statement for $n = 2$. It is sufficient to use this as a partition of the measure space.

$$\begin{aligned}\Pr(A_1 \wedge A_2) &= \Pr(A_1 \setminus A_2) + \Pr(A_2 \setminus A_1) + \Pr(A_1 \wedge A_2)\\ &= \Pr(A_1) + \Pr(A_2) - \Pr(A_1 \wedge A_2) \leq \Pr(A_1) + \Pr(A_2).\end{aligned}$$

□

Theorem 5.3.2 (Markov's Inequality) *Let Z be a random variable such that $\Pr(Z < 0) = 0$ and $a > 0$, then*

$$\Pr(Z \geq a) \leq \frac{\mathbb{E}[Z]}{a}.$$

Proof. We prove this in the case that Z is a discrete random variable (with finite mean). The measure theory case is parallel.

$$\mathbb{E}[Z] = \sum_{x_i < a} x_i \Pr(X = x_i) + \sum_{x_i \geq a} x_i \Pr(X = x_i) \geq 0 + a \Pr(Z \geq x_i).$$

Dividing by a gives the result. □

Theorem 5.3.3 (Chebyshev's Inequality) *Let X be a random variable with expectation μ and variance $\sigma^2 < \infty$. For any $b > 0$,*

$$\Pr\left(|X-\mu| \geq \sigma b\right) \leq \frac{1}{b^2}.$$

Proof. This is a direct result of Markov's inequality.

$$\Pr\left(|X-\mu| \geq \sigma b\right) = \Pr\left((X-\mu)^2 \geq \sigma^2 b^2\right) \leq \frac{\mathbb{E}\left[(X-\mu)^2\right]}{\sigma^2 b^2} = \frac{1}{b^2}.$$

□

Exercises.

(1) Prove Proposition 5.4.2. In the first part, you don't need to know the distribution of the random variables. In the second case, assume the random variables are discrete.

5.4 Chernoff-Hoeffding bounds

Chernoff bounds or Chernoff-Hoeffding bounds are very powerful, but require the notion of independence.

5.4.1 Independence

Recall that events $A_1, \ldots, A_n$ are (mutually) independent if $\Pr\left(\bigwedge_{i\in S} A_i\right) = \prod_{i\in S} \Pr(A_i)$ for all $S \subseteq [n]$ and real-valued random variables $X_1, \ldots, X_n$ are (mutually) independent if, for $A_i \subseteq \mathbb{R}$, $i = 1, \ldots, n$, then

$$\Pr\left(\bigwedge_{i=1}^{n} \{X_i \in A_i\}\right) = \prod_{i=1}^{n} \Pr(X_i \in A_i).$$

We leave the following as an exercise:

Proposition 5.4.2 *If $X_1, \ldots, X_n$ are pairwise independent random variables, each with finite mean, then*

$$\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{Var}(X_i).$$

If $X_1, \ldots, X_n$ are mutually independent random variables, then

$$\mathbb{E}\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} \mathbb{E}[X_i].$$

5.4.3 A general Chernoff bound

The following proof of a general Chernoff bound (sometimes called a Chernoff-Hoeffding bound) is due originally to a lecture from Van Vu [Vu10], transcribed by Kirill Levchenko.

Theorem 5.4.4 (Chernoff bound) *Let* $X_1, \dots, X_n$ *be discrete (finite domain), independent real-valued random variables such that* $\mathbb{E}[X_i] = 0$ *and* $|X_i| \leq 1$ *for all* i. *Let* $X = \sum_{i=1}^{n} X_i$ *and* $\sigma^2 = \mathrm{Var}(X)$. *If* $0 \leq \lambda \leq 2\sigma$, *then*

$$\Pr(|X| \geq \lambda\sigma) \leq 2e^{-\lambda^2/4}.$$

Proof. A main reason this proof works is the following lemma:

Lemma 5.4.5 *Let* Z *be a discrete (finite domain), real-valued random variable such that* $-1 \leq Z \leq 1$ *and* $\mathbb{E}[Z] = 0$. *Then, for* $0 \leq t \leq 1$,

$$\mathbb{E}\left[e^{tZ}\right] < 1 + t^2\mathrm{Var}(Z) \leq \exp\{t^2\mathrm{Var}(Z)\}.$$

Proof of Lemma 5.4.5. Let Z take on the values $a_1, \dots, a_m$ and $p_j = \Pr(Z = a_j)$ for $j = 1, \dots, m$.

$$\begin{aligned}
\mathbb{E}\left[e^{tZ}\right] &= \sum_{j=1}^{m} p_j e^{ta_j} \\
&= \sum_{j=1}^{m} p_j \left(\sum_{i=0}^{\infty} \frac{1}{i!}(ta_j)^i \right) \\
&= \sum_{j=1}^{m} p_j + \sum_{j=1}^{m} p_j(ta_j) + \sum_{j=1}^{m} p_j \left(\sum_{i=2}^{\infty} \frac{1}{i!}(ta_j)^i \right) \\
&= \sum_{j=1}^{m} p_j + t\sum_{j=1}^{m} p_j a_j + \sum_{j=1}^{m} p_j t^2 a_j^2 \left(\sum_{i=2}^{\infty} \frac{1}{i!}(ta_j)^{i-2} \right) \\
&\leq \sum_{j=1}^{m} p_j + t\sum_{j=1}^{m} p_j a_j + t^2\sum_{j=1}^{m} p_j a_j^2 \left(\sum_{i=2}^{\infty} \frac{1}{i!} \right) \\
&= 1 + t\mathbb{E}[Z] + t^2(e-2)\sum_{j=1}^{m} p_j a_j^2 \\
&< 1 + t^2\mathrm{Var}(Z) \\
&\leq \exp\{t^2\mathrm{Var}(Z)\},
\end{aligned}$$

where this last comes from the fact that $1 + x \leq e^x$. □

Now to the proof of the theorem: By symmetry, it is sufficient to prove that $\Pr(X \geq \lambda\sigma) \leq e^{-\lambda^2/4}$. We use Lemma 5.4.5 and Markov's inequality.

$$\begin{aligned} \Pr(X \geq \lambda\sigma) &= \Pr\left(e^{tX} \geq e^{\lambda\sigma}\right) \\ &\leq \frac{\mathbb{E}\left[e^{tX}\right]}{e^{t\lambda\sigma}}. \end{aligned}$$

As to the numerator, we use the independence of the X_i. Let $0 \leq t \leq 1$.

$$\begin{aligned} \mathbb{E}\left[e^{tX}\right] &= \mathbb{E}\left[e^{t(X_1+\cdots+X_n)}\right] = \mathbb{E}\left[\prod_{i=1}^{n} e^{tX_i}\right] = \prod_{i=1}^{n} \mathbb{E}\left[e^{tX_i}\right] \\ &< \prod_{i=1}^{n}\left(1+t^2\mathrm{Var}(X_i)\right) \leq \prod_{i=1}^{n}\exp\left\{t^2\mathrm{Var}(X_i)\right\} \\ &= \exp\left\{t^2\sum_{i=1}^{n}\mathrm{Var}(X_i)\right\} = \exp\left\{t^2\sigma^2\right\}. \end{aligned}$$

Therefore,

$$\Pr(X \geq \lambda\sigma) \leq \frac{\mathbb{E}\left[e^{tX}\right]}{e^{t\lambda\sigma}} < \exp\left\{t^2\sigma^2 - t\lambda\sigma\right\}.$$

The optimal choice for t is $t = \lambda/(2\sigma)$, which is in the acceptable interval for t and so, $\Pr(X \geq \lambda\sigma) \leq \exp\{-\lambda^2/4\}$, exactly what we wanted to prove. □

Note that the previous proof even gives a bound in the case that $\lambda > 2\sigma$ by choosing $t = 1$. That is, for X described as in Theorem 5.4.4,

$$\Pr(X \geq \lambda\sigma) \leq \begin{cases} e^{-\lambda^2/4}, & \text{if } 0 \leq \lambda \leq 2\sigma\text{; and} \\ e^{-\sigma(\lambda-\sigma)}, & \text{if } \lambda > 2\sigma. \end{cases}$$

5.4.6 Binomial random variables

A random variable Y is **Bernoulli with parameter** p (denoted $Y \sim \mathrm{Ber}(p)$) if $\Pr(Y = 1) = p$ and $\Pr(Y = 0) = 1 - p$. It is easy to compute $\mathbb{E}[Y] = p$ and $\mathrm{Var}(Y) = p(1-p)$. A Bernoulli random variable is often referred to as a **biased coin flip**. A random variable X is **binomial with parameters n and p (denoted $X \sim \mathrm{bin}(n,p)$)** if $X \sum_{i=1}^{n} X_i$ where $\{X_i\}$ is a set of independent $\mathrm{Ber}(p)$ random variables. Again, it is easy to see that $\mathbb{E}[X] = np$ and $\mathrm{Var}(X) = np(1-p)$.

Corollary 5.4.7 follows directly from Theorem 5.4.4.

Corollary 5.4.7 *Let* $X \sim \mathrm{bin}(n,p)$*,*

$$\Pr\left(|X - np| \geq \lambda\sqrt{np(1-p)}\right) \leq 2\exp\{-\lambda^2/4\}$$

for all λ*,* $0 \leq \lambda \leq 2\sqrt{np(1-p)}$*.*

Exercises.

(1) Let X be a binomial random variable with parameters n and p. That is, a coin is flipped n times and is heads with probability p and tails with probability $1-p$.

- Show that the mean of X is np and the variance is $np(1-p)$.
- Use the Chernoff bound to show that, for all λ, $0 \leq \lambda \leq 2\sqrt{np(1-p)}$.

$$\Pr(|X - np| \geq \lambda\sqrt{np(1-p)}) \leq 2e^{-\lambda^2/4}.$$

In particular, for fixed p and any $\lambda = \lambda(n) \to \infty$, $|X - np| \leq \lambda\sqrt{n}$ **whp**.

5.5 The random graph

The traditional Erdős-Rényi model of a random graph [ER60] is as follows: Let $V(G) = [n]$ and each edge vw is present, independently with probability p. The resulting *graph-valued* random variable is denoted $G(n,p)$ (sometimes $G_{n,p}$). Note that this is a labeled graph and the probability that any particular graph G, on n vertices, is chosen is $p^{e(G)}(1-p)^{\binom{n}{2}-e(G)}$.

We have seen the random graph before. In the proof of Theorem 5.2.3, both the red and blue graphs are distributed according to $G(n,p)$. In fact, if $G \sim G(n,p)$, then the complement $\overline{G}$ is distributed according to $G(n, 1-p)$.

The random graph has a number of interesting properties.

Theorem 5.5.1 *Fix p, $0 < p < 1$ and let $f = f(n)$ have the property that $f \to \infty$ as $n \to \infty$. Let $G \sim G(n,p)$.* ***Whp,***

(1) $np - f(n)\sqrt{n \ln n} < \deg(v) < np + f(n)\sqrt{n \ln n}$, for all $v \in V(G)$.

(2) $\alpha(G) \leq \frac{2 \ln n}{-\ln(1-p)}$ and $\omega(G) \leq \frac{2\ln n}{-\ln p}$.

Proof.

(1) Note that each $\deg(v)$ is a $\mathrm{bin}(n-1,p)$ random variable. Boole's inequality (Proposition 5.3.1), the Chernoff bound (Theorem 5.4.4) and the fact that $p = o(\sqrt{n \ln n})$ gives the result.

$$\begin{aligned} \Pr\left(\bigvee_{v \in V(G)} \left\{|\deg(v) - (n-1)p| \geq f(n)\sqrt{\ln n}\sqrt{np(1-p)}\right\}\right) & \\ \leq \quad n\left(2e^{-f^2(n)\ln n/4}\right) = o(1). & \end{aligned}$$

(2) Let $\alpha_0 = 1 + 2\ln n/(-\ln(1-p))$.

$$\begin{aligned}\Pr(\alpha(G) \geq \alpha_0) &\leq \binom{n}{\alpha_0}(1-p)^{\binom{\alpha_0}{2}} \leq \left(\frac{en}{\alpha_0}\right)^{\alpha_0}(1-p)^{\alpha_0(\alpha_0-1)/2} \\ &\leq \left(\frac{en}{\alpha_0}(1-p)^{(\alpha_0-1)/2}\right)^{\alpha_0} = \left(\frac{e}{\alpha_0}\right)^{\alpha_0} = o(1).\end{aligned}$$

So, $\alpha_0 \leq 2\ln n/(-\ln(1-p))$ **whp**.

□

5.6 Alteration method

The classical application of the so-called alteration method is a famous theorem of Erdős [Erd59]. The **chromatic number** of a graph G, denoted $\chi(G)$, is the smallest number of independent sets into which $V(G)$ can be partitioned. The **girth** of G, denoted $\mathrm{girth}(G)$ is the length of its smallest cycle.

It is easy to see that, for any n-vertex graph G, if $\omega(G)$ and $\alpha(G)$ denote the clique number and independence number of G, respectively, then $\omega(G) \leq \chi(G)$ and $n/\alpha(G) \leq \chi(G)$.

Theorem 5.6.1 (Erdős [Erd59]) *For all integers k, g, there exists a graph G with $\chi(G) \geq k$ and $\mathrm{girth}(G) \geq g$.*

Proof. Construct $G \sim G(n,p)$, where $p := n^{\theta-1}$ for some fixed $\theta < 1/g$. Let X be the random variable that counts the number of cycles of girth at most g.

$$E[X] = \sum_{i=3}^{g} \frac{(i-1)!}{2}\binom{n}{i}p^i \leq \sum_{i=3}^{g}\frac{(np)^i}{2i} \leq \frac{g-2}{6}n^{\theta g} \leq gn^{\theta g}.$$

Using Markov's inequality,

$$\Pr(X \geq n/2) \leq \frac{\mathbb{E}[X]}{n/2} = 2gn^{\theta g - 1} = o(1).$$

So, for n large enough, there is an outcome that gives fewer than $n/2$ small cycles.

Now we look at the large independent sets

$$\Pr(\alpha(G) \geq a) \leq \binom{n}{a}(1-p)^{\binom{a}{2}} < n^a e^{-p\binom{a}{2}} = \left[ne^{-p(a-1)/2}\right]^a.$$

Set $a := \lceil (4\ln n)/p \rceil + 1$ and so

$$\Pr(\alpha(G) \geq a) \leq n^{-a} = o(1).$$

Choose an n large enough so that both $\Pr(X \geq n/2) < 1/2$ and $\Pr(\alpha(G) \geq a) < 1/2$. So there is a graph on n vertices with at most $n/2$ small cliques and no independent set of size a. Remove one vertex from each small cycle to ensure that the resulting graph G' has no small cycles, but $\alpha(G') < a = \lceil (4 \ln n) n^{1-\theta} \rceil + 1$. Hence,

$$\chi(G') \geq \frac{n/2}{\alpha(G')} \geq \frac{n/2}{\lceil (4\ln n) n^{1-\theta} \rceil + 1} \geq \frac{n}{(10 \ln n) n^{1-\theta}} = \frac{n^\theta}{6 \ln n},$$

which approaches infinity, so choose n large enough to ensure that the right hand side is at least k. □

5.7 Second moment method

5.7.1 Threshold functions

The notation $f \ll g$ is often used instead of the cumbersome $f = o(g)$.

Definition 5.7.2 *Let $\mathcal{P}$ be a property of graphs. The function $r(n)$ is a **threshold function** for $\mathcal{P}$ if*

- *Whenever $p = p(n) \ll r(n)$, $\lim_{n\to\infty} \Pr(G(n,p) \in \mathcal{P}) = 0$.*
- *Whenever $p = p(n) \gg r(n)$, $\lim_{n\to\infty} \Pr(G(n,p) \in \mathcal{P}) = 1$.*

Erdős and Rényi[ER60] made the observation that if $\mathcal{P}$ is a monotone property of graphs, then there is a threshold function $r(n)$.

5.7.3 Threshold for the emergence of a triangle

A very direct application of Chebyshev's inequality (Theorem 5.3.3) is striking. In the notation used below, X is a random variable which is a function of some n which is approaching infinity. Strictly speaking, we should be discussing a sequence of random variables, but we choose this notation for intuitive convenience.

We also use the notation $X \sim \mu$ to identify the fact that for all $\epsilon > 0$, $X \in ((1-\epsilon)\mu, (1+\epsilon)\mu)$.

Theorem 5.7.4 *Let $X = X(n)$ be a random variable with $\mathbb{E}[X] \to \infty$ such that $\mathrm{Var}(X) = o\left(\mathbb{E}[X]^2\right)$, then $X > 0$ **whp** and $X \sim \mathbb{E}[X]$ **whp**.*

Proof. Let $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathrm{Var}(X)$. If $\sigma^2 = 0$, then $X = \mathbb{E}[X]$ with probability 1 and the theorem results. Otherwise, by Chebyshev, $\Pr(|X - \mu| \geq \lambda\sigma) \leq \lambda^{-2}$, as long as $\lambda \leq 2\sigma$.

If $\lambda = \mu/\sigma$, then

$$\Pr\left(|X - \mu| \geq \mu = \frac{\mu}{\sigma}\sigma\right) \leq \frac{\sigma^2}{\mu^2} \to 0.$$

Fix $\epsilon > 0$. If $\lambda = \epsilon\mu/\sigma$, then

$$\Pr\left(|X-\mu| \geq \epsilon\mu = \frac{\epsilon\mu}{\sigma}\sigma\right) \leq \frac{\sigma^2}{\epsilon^2\mu^2} \to 0.$$

As a result, for any fixed $\epsilon > 0$, $X \in ((1-\epsilon)\mu, (1+\epsilon)\mu)$, **whp**. ☐

As an application of this, we require the following corollary:

Corollary 5.7.5 *Let $A_1, \ldots, A_n$ be events in a probability space and construct an auxiliary graph on $[n]$ for which $i \sim j$ iff $i \neq j$ and the events A_i, A_j are not independent. Let*

$$\Delta = \sum_{i \sim j} \Pr(A_i \cap A_j).$$

Let $X = \sum_i \mathbf{1}_{A_i}$. If $\mathbb{E}[X] \to \infty$ and $\Delta = o\left(\mathbb{E}[X]^2\right)$, then $X > 0$ **whp** *and $X \sim \mathbb{E}[X]$* **whp**.

Proof. By Theorem 5.7.4, it is sufficient to show that $\mathrm{Var}(X) = o\left(\mathbb{E}[X]^2\right)$. Let $\mu = \mathbb{E}[X]$, $\sigma^2 = \mathrm{Var}(X)$ and $p_i = \Pr(A_i)$ for $i = 1, \ldots, n$. Let $X_i = \mathbf{1}_{A_i}$ so that $\mathbb{E}[X_i] = p_i$. Note also that $X_i^2 = X_i$ because it is an indicator variable. By linearity of expectation,

$$\begin{aligned}
\sigma^2 &= \mathbb{E}\left[(X-\mu)^2\right] = \mathbb{E}\left[\left(\sum_i (X_i - p_i)\right)^2\right] \\
&= \mathbb{E}\left[\sum_{i,j}(X_i - p_i)(X_j - p_j)\right] = \sum_{i,j}\mathbb{E}\left[(X_i - p_i)(X_j - p_j)\right] \\
&= \sum_{i,j}\mathbb{E}\left[X_iX_j - p_iX_j - p_jX_i + p_ip_j\right] = \sum_{i,j}\left(\mathbb{E}[X_iX_j] - p_ip_j\right).
\end{aligned}$$

Since $\mathbb{E}[X_iX_j] = \mathbb{E}[X_i]\mathbb{E}[X_j] = p_ip_j$ for any distinct i, j with $i \not\sim j$,

$$\begin{aligned}
\sigma^2 &= \sum_{i \sim j \text{ or } i=j}\left(\mathbb{E}[X_iX_j] - p_ip_j\right) \\
&\leq \sum_{i \sim j \text{ or } i=j}\mathbb{E}[X_iX_j] = \sum_i \Pr(A_i) + \sum_{i \sim j}\Pr(A_i \cap A_j) \\
&= \Delta + \mu.
\end{aligned}$$

Thus, $\Delta = o(\mu^2)$ iff $\sigma^2 = o(\mu^2)$ and since $\Delta = o(\mu^2)$ and $\mu = o(\mu^2)$, the corollary follows. ☐

Theorem 5.7.6 (Erdős-Rényi [ER60]) *The property of $\omega(G) \geq 3$ has threshold function n^{-1}.*

Proof. For a subset S of three vertices, let A_S be the event that S induces a triangle. Let X be the number of triangles in $G(n,p)$.

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_S \mathbf{1}_{A_S}\right] = \sum_S \Pr(A_S) = \binom{n}{3}p^3 \sim \frac{n^3p^3}{6}.$$

If $p \ll n^{-1}$, then $\mathbb{E}[X] \to 0$. By Markov's inequality, since X is integral, $\Pr(X > 0) \leq \mathbb{E}[X] \to 0$ and so $X = 0$, **whp**.

For distinct S and T, $S \sim T$ iff $|S \cap T| = 2$. As such, it is easy to see that

$$\Delta = \frac{1}{2}\binom{n}{3} \cdot 3(n-3)p^5 \sim \frac{n^4p^5}{4}$$

.

If $p \gg n^{-1}$, then $\Delta = o(\mathbb{E}[X]^2)$ and Corollary 5.7.5 gives that $X > 0$; i.e., there will be a triangle **whp**. ☐

Exercises.

(1) Prove that the threshold function for the property $\omega(G) \geq k$ is $n^{-2/(k-1)}$.

5.8 Conditional probability

The conditional probability $\Pr(A \mid B)$ is, informally, the probability that event A occurs, given that event B occurs.

5.8.1 The famous Monty Hall Problem:

An ideal example of conditional probability is that of the so-called Monty Hall problem [HHPa, vS97, Mos04]:

Question 5.8.2 *You are playing a game of Let's Make A Deal [HHPb]. The host, Monty Hall, shows you three doors. Behind one of those doors is a fabulous prize. Behind the other two is a goat[1]. Monty Hall asks which door you choose. You choose a door and then Monty opens a door you didn't choose, revealing a goat. He then asks, "Would you like to switch to the other door, or stay with your choice?"*

Should you switch?

Let W be the event that you win. Let S be the event that you switch. If you don't switch, then the probability that you win is the probability that you chose the right door to begin with. Assuming no knowledge about the location of the fabulous prize, this gives

$$\Pr(W \mid S^c) = 1/3.$$

---

[1] The goat is not a fabulous prize.

If you do switch, then the probability that you win is the probability that you chose a wrong door to begin with. In that circumstance, when you switch, you must select the fabulous prize. So, this gives,

$$\Pr(W \mid S) = 2/3.$$

Therefore, you should switch.[2]

5.8.3 Formal definition of conditional probability

Definition 5.8.4 *The* ***conditional probability*** *of event A, given event B, is denoted* $\Pr(A \mid B)$ *and obeys the rule*

$$\Pr(A \mid B)\Pr(B) = \Pr(A \wedge B).$$

Generalizing:

Proposition 5.8.5 *Let $E_1, \ldots, E_n$ be events in a probability space, then*

$$\begin{aligned} \Pr\left(\bigwedge_{i=1}^{n} E_i\right) &= \prod_{i=1}^{n} \Pr\left(E_i \,\middle|\, \bigwedge_{j=1}^{i-1} E_j\right) \\ &= \Pr(E_1) \cdot \Pr(E_2 \mid E_1) \cdot \Pr(E_3 \mid E_1 \wedge E_2) \cdot \ldots \cdot \Pr\left(E_n \,\middle|\, \bigwedge_{j=1}^{n-1} E_j\right) \end{aligned}$$

Clearly, if A and B are independent events, then $\Pr(A \mid B) = \Pr(A)$.

One way to think of $\Pr(A \mid B)$ is to view event B as the new "universe" and $\Pr(A \mid B)$ is the proportion of A in the new universe B. Clearly, $\Pr(\cdot \mid B)$ is a probability space and we can compute all of the same statistics. In the case of a discrete random variable X, and a real-valued function f,

$$\begin{aligned} \mathbb{E}[X \mid B] &= \sum_i x_i \Pr(X = x_i \mid B) \\ \mathbb{E}[f(X) \mid B] &= \sum_i f(x_i) \Pr(X = x_i \mid B) \\ \mathrm{Var}(X \mid B) &= \mathbb{E}\left[(X - \mathbb{E}[X \mid B])^2 \mid B\right] \end{aligned}$$

and so on.

---

[2] Monty Hall's real method of operation was more complex than this simplified version, however.

5.8.6 Thresholds in random graphs

We are going to obtain a more useful form of 5.7.5:

Corollary 5.8.7 *Let $A_1, \dots, A_n$ be events in a probability space and construct an auxiliary graph on $[n]$ for which $i \sim j$ iff $i \neq j$ and the events A_i, A_j are not independent. Let*

$$\Delta^* = \max \left\{ \sum_{j \sim i} \Pr(A_j \mid A_i) : i = 1, \dots, n \right\}.$$

Let $X = \sum_i \mathbf{1}_{A_i}$. If $\mathbb{E}[X] \to \infty$ and $\Delta^* = o(\mathbb{E}[X])$, then both $X > 0$ and $X \sim \mathbb{E}[X]$ **whp**.

Proof. By definition,

$$\begin{aligned} 2\Delta &= 2\sum_{i \sim j} \Pr(A_i \cap A_j) = \sum_i \sum_{j \sim i} \Pr(A_j \mid A_i) \Pr(A_i) \\ &= \sum_i \Pr(A_i) \sum_{j \sim i} \Pr(A_j \mid A_i) \\ &\leq \sum_i \Pr(A_i) \Delta^* = \Delta^* \mathbb{E}[X]. \end{aligned}$$

So $\Delta = o(\mathbb{E}[X]^2)$ iff $\Delta^* = o(\mathbb{E}[X])$. □

We will use this to find the threshold for the appearance for a variety of graphs. But first we need a definition.

Definition 5.8.8 *For a graph H with v vertices and e edges, we define the* ***density of*** *H to be $\rho(H) = e/v$. Note that the average degree of H is exactly $2\rho(H)$. We call H* ***balanced*** *if, for every subgraph H' of H, then $\rho(H') \leq \rho(H)$. We call H* ***strictly balanced*** *if, for every proper subgraph H' of H, then $\rho(H') < \rho(H)$.*

Theorem 5.8.9 (Erdős-Rényi [ER60]) *Let H be a balanced graph with v vertices and e edges. Let $A(G)$ be the event that H is a subgraph of G. Then $p = n^{-v/e}$ is the threshold function for A.*

Proof. For any set S of v vertices, let A_S be the event that $G|_S$ has H as a subgraph. Then, $p^e \leq \Pr(A_S) \leq v! p^e$. By linearity of expectation,

$$\mathbb{E}[X] = \sum_{|S|=v} \mathbf{1}_{A_S} = \binom{n}{v} \Pr(A_S) = \theta(n^v p^e).$$

If $p \ll n^{-v/e}$, then $\mathbb{E}[X] = o(1)$ and $X = 0$ **whp**.

So, assume $p \gg n^{-v/e}$ and so $\mathbb{E}[X] \to \infty$. Computing Δ^*,

$$\Delta^* = \sum_{T \sim S} \Pr(A_T \mid A_S) = \sum_{i=2}^{v-1} \sum_{|T \cap S| = i} \Pr(A_T \mid A_S).$$

For each i, there are at most $\binom{v}{i}\binom{n-v}{v-i} = O(n^{v-i})$ choices of T. There are at most $v! = O(1)$ copies of H in T. Since H is balanced, there are at most ie/v edges with both of the endvertices in S. Hence there are at least $e - ie/v$ edges with at least one endvertex in T and so, $\Pr(A_T \mid A_S) = O(p^{e-ie/v})$. Computing Δ^*:

$$\begin{aligned}\Delta^* &= \sum_{i=2}^{v-1} O\left(n^{v-i} p^{e-ie/v}\right) = \sum_{i=2}^{v-1} O\left((n^v p^e)^{1-i/v}\right)\\ &= \sum_{i=2}^{v-1} o(n^v p^e) = o(\mathbb{E}[X])\end{aligned}$$

because $n^v p^e \to \infty$. So, we can apply Corollary 5.8.7 and conclude that $n^{-v/e}$ is a threshold function. □

Similar methods, but additional details give some more results. We neglect to prove them here.

Theorem 5.8.10 (Erdős-Rényi [ER60])

(1) Let H have the property that $\rho^*(H) = \max_{H' \leq H}\{\rho(H')\}$. Then $p = n^{-1/\rho^*(H)}$ is the threshold function for the event that H is a subgraph of G.

(2) Let H be a graph and for any $H' \leq H$, let the number of copies of H' in $G(n,p)$ be $X_{H'}$. If $\mathbb{E}[X_{H'}] \to \infty$ for every H'. Then, $X_H \sim \mathbb{E}[X_H]$.

Corollary 5.8.11 (Erdős-Rényi [ER60])

(1) The function $p = n^{-2/(v-1)}$ is the threshold function for the event that K_v is a subgraph of G.

(2) Let H be a strictly balanced graph with v vertices and e edges and a automorphisms. Let X be the number of copies of H in $G(n,p)$. If $p \gg n^{-v/e}$, then $X \sim \frac{n^v p^e}{a}$ **whp**.

Exercises.

(1) Prove Proposition 5.8.5.

(2) Prove Corollary 5.8.11 from Theorem 5.8.10 or directly.

(3) Prove that the property that $G \sim G(n,p)$ has no isolated vertices has threshold function $\ln n/n$.

5.9 Lovász' Local Lemma

The Lovász local lemma (which, despite the alliteration, is due to both Erdős and Lovász) is a way of measuring the probability of the intersection of events among whom there are very few dependencies.

In order to apply the local lemma, we need to define a dependency digraph.

Definition 5.9.1 *Let $A_1, \ldots, A_n$ be events in a probability space. We say that a directed graph $D = (V, A)$ is a* ***dependency digraph*** *if $V = [n]$ and A is a set of arcs such that for every $i \in [n]$, the event A_i is independent of every event of the form $\left\{\bigwedge_{j \in J} \overline{A_j} : J \subseteq \{j : i \not\to j\}\right\}$.*

The most common statement of the Lovász local lemma is as follows:

Theorem 5.9.2 (Erdős-Lovász [EL75]; Local Lemma, Symmetric Form) *Let $A_1, \ldots, A_n$ be events in a probability space with dependency digraph D. If d represents the maximum out-degree of D and that $\Pr(A_i) \le p$ for all $1 \le i \le n$. If $ep(d+1) \le 1$, then*

$$\Pr\left(\bigwedge_{i=1}^{n} \overline{A_i}\right) > 0.$$

Here, $e \approx 2.7128$ is Euler's constant.

The proof of the symmetric form requires a more general statement, proved by conditional probabilities. We will do the proof later, but first we will discuss a simple result.

5.9.3 Property B

Bernstein [Ber08] asked, given a collection of subsets of V, is there a partition into V_1 and V_2 so that no subset is either in V_1 or V_2? In other words, can the vertices of a given hypergraph be 2-colored so that no hyperedge is monochromatic? Erdős popularized this question and said that a hypergraph whose vertices can be 2-colored so that there is no monochromatic hyperedge has **property B** – in honor of Bernstein.

A hypergraph is r-uniform if every hyperedge has exactly r vertices. We leave the proof of Theorem 5.9.4 as an exercise as well as the corollary that an r-uniform hypergraph with m vertices has property B if $m < 2^{r-1}$.

Theorem 5.9.4 (Erdős-Lovász [EL75]) *Let $\mathcal{H}$ be a hypergraph in which every edge has size r and suppose that each edge of $\mathcal{H}$ has nonempty intersection with at most d other edges. If $e(d+1) \le 2^{r-1}$, then $\mathcal{H}$ has property B.*

Theorem 5.9.4 was the problem that motivated the development of the Lovász local lemma. It has subsequently been used on a great many problems since.

5.9.5 Lovász local lemma – asymmetric form

The symmetric form of the Lovász local lemma is derived from the asymmetric form:

Theorem 5.9.6 (Erdős-Lovász [EL75]; Local Lemma, General Form) *Let* $A_1, \dots, A_n$ *be events in an arbitrary probability space and let* D *be the dependency digraph. Suppose there are real numbers* $x_1, \dots, x_n$ *such that* $0 \le x_i < 1$ *and* $\Pr(A_i) \le x_i \prod_{j:i\to j}(1-x_j)$, *for* $i = 1, \dots, n$. *Then* $\Pr\left(\bigwedge_{i=1}^{n} \overline{A_i}\right) \ge \prod_{i=1}^{n}(1-x_i) > 0$. *I.e., with positive probability, no* A_i *holds.*

Proof. The following Claim is the main part of the proof:

Claim 5.9.7 *For any* $S \subset \{1, \dots, n\}$, $|S| = s < n$ *and any* $i \notin S$,

$$\Pr\left(A_i \,\middle|\, \bigwedge_{j\in S} \overline{A_j}\right) \le x_i. \tag{5.2}$$

Proof of Claim 5.9.7. The statement is true trivially for $s = 0$. Assume, by a strong inductive hypothesis, that the statement is true for all $s' < s$. Let $S_1 = \{j \in S : i \to j\}$ and $S_2 = S \setminus S_1$. Then,

$$\begin{aligned}
\Pr\left(A_i \,\middle|\, \bigwedge_{j\in S} \overline{A_j}\right) &= \Pr\left(A_i \,\middle|\, \bigwedge_{j\in S_1} \overline{A_j} \wedge \bigwedge_{j\in S_1} \overline{A_j}\right) \\
&= \frac{\Pr\left(A_i \wedge \left(\bigcap_{j\in S_1} \overline{A_j}\right) | \bigwedge_{\ell\in S_2} \overline{A_\ell}\right)}{\Pr\left(\bigwedge_{j\in S_1} \overline{A_j} \,|\bigwedge_{\ell\in S_2} \overline{A_\ell}\right)}.
\end{aligned}$$

Bounding the numerator,

$$\begin{aligned}
\Pr\left(A_i \cap \left(\bigwedge_{j\in S_1} \overline{A_j}\right) \,\middle|\, \bigwedge_{\ell\in S_2} \overline{A_\ell}\right) &\le \Pr\left(A_i \,\middle|\, \bigwedge_{\ell\in S_2} \overline{A_\ell}\right) \\
&= \Pr(A_i) \le x_i \prod_{j:i\to j}(1-x_j)
\end{aligned}$$

To bound the denominator, we use the inductive hypothesis. Let $S_1 = \{j_1, \dots, j_r\}$. (If $r = 0$, then the denominator is 1, and (5.2) is satisfied.) Otherwise, by Proposition 5.8.5

$$\begin{aligned}
&\Pr\left(\overline{A_{j_1}} \wedge \overline{A_{j_2}} \wedge \cdots \overline{A_{j_r}} \,\middle|\, \bigwedge_{\ell\in S_2} \overline{A_\ell}\right) \\
&\quad = \Pr\left(\overline{A_{j_1}} \,\middle|\, \bigwedge_{\ell\in S_2} \overline{A_\ell}\right) \cdot \ldots \cdot \Pr\left(\overline{A_{j_r}} \,\middle|\, \overline{A_{j_{r-1}}} \wedge \cdots \wedge \overline{A_{j_1}} \cap \bigwedge_{\ell\in S_2} \overline{A_\ell}\right) \\
&\quad = \left(1 - \Pr\left(A_{j_1} \,\middle|\, \bigwedge_{\ell\in S_2} \overline{A_\ell}\right)\right) \cdot \ldots \cdot \left(1 - \Pr\left(A_{j_r} \,\middle|\, \overline{A_{j_{r-1}}} \wedge \cdots \wedge \overline{A_{j_1}} \cap \bigwedge_{\ell\in S_2} \overline{A_\ell}\right)\right) \\
&\quad \geq (1-x_{j_1})(1-x_{j_2}) \cdot \ldots \cdot (1-x_{j_r}) \geq \prod_{j: i\to j} (1-x_j).
\end{aligned}$$

So, the induction is finished. □

The statement of the lemma follows because Proposition 5.8.5 gives

$$\begin{aligned}
\Pr\left(\bigwedge_{i=1}^{n} \overline{A_i}\right) &= (1-\Pr(A_1)) \cdot \left(1 - \Pr\left(A_2 \,\middle|\, \overline{A_1}\right)\right) \cdot \ldots \\
&\quad \ldots \cdot \left(1 - \Pr\left(A_n \,\middle|\, \bigwedge_{i=1}^{n-1} \overline{A_i}\right)\right) \geq \prod_{i=1}^{n} (1-x_i)
\end{aligned}$$

□

It is now easy to prove the symmetric form of the local lemma: **Proof of Proof of Theorem 5.9.2..** The result is trivial for $d = 0$. If $d \geq 1$, then apply Theorem 5.9.6 with $x_i = 1/(d+1)$. for all i and use the fact that $\left(1 - \frac{1}{d+1}\right)^d > e^{-1}$. □

5.9.8 Application: $R(3,k)$

An application of the asymmetric version of the Lovász Local Lemma is the Ramsey number $R(3,k)$. That is, if $n \geq R(3,k)$, then any coloring of $E(K_n)$ yields either a red K_3 or a blue K_k. This bound was proven by Erdős [Erd61] earlier but as an application of the local lemma the proof is due to Spencer [Spe78]:

Theorem 5.9.9 (Erdős [Erd61]; Spencer [Spe78]) *There exists a constant c such that $R(k,3) > ck^2/\log^2 k$.*

Proof. We will start with K_n and color each edge independently at random, red with probability p and blue with probability $1-p$.

For any subset T of 3 vertices, we denote by A_T the event that the edges induced by T are all red. For any subset S of k vertices, we denote by B_S the event that the edges induced by S are all blue. So, $\Pr(A_T) = p^3$ for all T and $\Pr(B_S) = (1-p)^{\binom{k}{2}}$ for all S.

The dependency digraph, D will have one vertex for every A_T and every B_S. Two nodes in D are adjacent (in each direction) if the sets they represent share at least two vertices.

Each A_T node is adjacent to $3(n-3) < 3n$ nodes $A_{T'}$ and at most $\binom{n}{k}$ nodes $B_{S'}$. Each B_S node is adjacent to $\binom{k}{2}(n-k) + \binom{k}{3} < k^2n/2$ nodes $A_{T'}$ and at most $\binom{n}{k}$ nodes $B_{S'}$.

Therefore, we want to find a $p \in (0,1)$ and real numbers $x \in [0,1)$ and $y \in [0,1)$ for which

$$\begin{aligned} p^3 &\leq x(1-x)^{3n}(1-y)^{\binom{n}{k}} && (5.3) \\ (1-p)^{\binom{k}{2}} &\leq y(1-x)^{k^2n/2}(1-y)^{\binom{n}{k}}. && (5.4) \end{aligned}$$

If we can do so, then this random coloring has neither a red K_3 nor a blue K_k and so $R(k,3) > n$.

It is not so easy to determine the values of p, k, x and y that make k minimal for fixed n. It is, however relatively easy to verify that there are constants c_1, c_2, c_3, c_4 such that $p = c_1 n^{-1/2}$, $k = c_2 n^{1/2} \ln n$, $x = c_3 n^{-3/2}$ and $y = c_4 e^{-n^{1/2} \ln^2 n}$ satisfy both (5.4) and (5.4). Applying Theorem 5.2, we see that for these values of n and k, there is a coloring of the edges of K_n so that there is neither a red triangle nor a blue k-clique.

Since $k = c_2 n^{1/2} \ln n$, there exists a c such that $n \geq ck^2/\log^2 k$ and so $R(3,k) > n$. □

Erdős [Erd61] established this lower bound and it was improved to $R(k,3) > ck^2/\log k$ by Kim [Kim95]. Graver and Yackel [GY68] proved that an upper bound was $Ck^2 \log\log k/\log k$. Ajtai, Komlós and Szemerédi [AKS80, AKS81] removed the $\log\log$ term in the numerator and Shearer [She83] (see also [She91]) improved the constant to obtain an upper bound of $R(k,3) < Ck^2/\log k$.

As a side note, we proved in Theorem 5.7.6 that the threshold for having a triangle in $G(n,p)$ is $p \sim n^{-1}$. We showed further that the expected number of triangles is large for $p \sim n^{-1/2}$. Yet the proof above shows that with positive probability $G(n,p)$, with $p \sim n^{-1/2}$ has no triangle (and simultaneously, no independent set of size k).

Exercises.

(1) Prove Theorem 5.9.4.

(2) Let $\mathcal{H}$ be an r-uniform hypergraph with m hyperedges. Prove that $\mathcal{H}$ has Property B if $m < 2^{r-1}$.

(3) Prove that $\left(1 - \frac{1}{d+1}\right)^d > e^{-1}$.

(4) Verify that the choices for p, k, x, y satisfy inequalities (5.3) and (5.4). **Hint:** You may want to use the inequality $1 - z \geq \exp\{z - z^2\}$, for all $|z| \leq 1/2$, and the inequality $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$, for $1 \leq k \leq n$.

(5) Verify that if $k = c_2 n^{1/2} \ln n$, there exists a c such that $n \geq ck^2/\log^2 k$.

(6) Use the symmetric version of the Lovász Local Lemma to prove that $R(k,k) > \frac{\sqrt{2}}{e}(1 - o(1))k2^{k/2}$. This is an improvement of a multiplicative factor of 2 over the naïve probabilistic bound in Theorem 5.2.3.

(7) An arithmetic progression of length k is a sequence $a, a+d, a+2d, \ldots, a+(k-1)d$. The van der Waerden number $W(k)$ is the least n so that if $\{1, \ldots, n\}$ is two-colored, then it has a monochromatic arithmetic progression of length k. Use the Lovász Local Lemma to prove that $W(k) > (1 - o(1))\frac{2^{k-1}}{ek}$.

5.10 Martingales

What may be counterintuitive, if X and Y are random variables, then $\mathbb{E}[X \mid Y]$ is a random variable that is a function of Y. That is, for a discrete random variable X, and a function f,

$$\begin{aligned}
\mathbb{E}[X \mid Y = y] &= \sum_i x_i \Pr(X = x_i \mid Y = y) \\
\mathbb{E}[f(X) \mid Y = y] &= \sum_i f(x_i) \Pr(X = x_i \mid Y = y)
\end{aligned}$$

A quick aside on an important fact about conditional expectations:

Theorem 5.10.1 *Let X and Y be discrete random variables with $\mathbb{E}[|X|] < \infty$. Then, $\mathbb{E}\left[\mathbb{E}[X \mid Y]\right] = \mathbb{E}[X]$.*

We note that in $\mathbb{E}\left[\mathbb{E}[X \mid Y]\right]$, the second expectation is with respect to random variable Y and so it is appropriate to write $\mathbb{E}_Y\left[\mathbb{E}[X \mid Y]\right]$.
Proof. As with many proofs, this is true in the more general measure-theoretical setting, but it is useful to see it for discrete random variables.

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}[X \mid Y]\right] &= \sum_y \mathbb{E}[X \mid Y = y] \Pr(Y = y) \\
&= \sum_y \sum_x x \Pr(X = x \mid Y = y) \Pr(Y = y) \\
&= \sum_y \sum_x x \Pr\left(\{X = x\} \wedge \{Y = y\}\right) \\
&= \sum_x x \sum_y \Pr\left(\{X = x\} \wedge \{Y = y\}\right) \\
&= \sum_x x \Pr(X = x) = \mathbb{E}[X],
\end{aligned}$$

which exists and is finite. □

Definition 5.10.2 *A* ***martingale***[3] *is a sequence of random variables* $X_0, X_1, \ldots, X_m$ *such that,*

- $\mathbb{E}[|X_i|] < \infty$ *for* $i = 0, \ldots, m$ *and*
- $\mathbb{E}[X_{i+1} \mid X_i, X_{i-1}, \ldots, X_0] = X_i$, *for* $i = 0, \ldots, m-1$.

Again, $\mathbb{E}[X_{i+1} \mid X_i, X_{i-1}, \ldots, X_0]$ is a function of the random variables $X_i, X_{i-1}, \ldots, X_0$ and it takes on the value of X_i.

Imagine playing a fair game m times. A fair game is one in which the expected winnings is zero. Let X_i be the amount of winnings at game i and set $X_0 = 0$. If it is a fair game, the expected value at the end of game $i+1$ is the same as the original winnings.

If A and $\{Z_i\}_{i \geq 0}$ are random variables on the same probability space with $\mathbb{E}[|A|] < \infty$, then $X_i = \mathbb{E}[A \mid Z_i, \ldots, Z_0]$ is called a **Doob martingale**. Let us show that it is indeed a martingale. First, we show that $\mathbb{E}[|X_i|] < \infty$:

$$\begin{aligned} \mathbb{E}[|X_i|] &= \mathbb{E}\left[|\mathbb{E}[A \mid Z_i, \ldots, Z_0]|\right] \\ &\leq \mathbb{E}\left[\mathbb{E}[|A| \mid Z_i, \ldots, Z_0]\right] = \mathbb{E}[|A|] \end{aligned}$$

Second, we show that $\mathbb{E}[X_{i+1} \mid X_i, \ldots, X_0] = X_i$. Note that X_i is a function of $Z_i, \ldots, Z_0$.

$$\begin{aligned} \mathbb{E}[X_{i+1} \mid X_i, \ldots, X_0] &= \mathbb{E}\left[\mathbb{E}[A \mid Z_{i+1}, Z_i, \ldots, Z_0] \mid X_i, \ldots, X_0\right] \\ &= \mathbb{E}\left[\mathbb{E}[A \mid Z_{i+1}] \mid X_i, \ldots, X_0, Z_i, \ldots, Z_0\right] \\ &= \mathbb{E}\left[\mathbb{E}[A \mid Z_{i+1}] \mid Z_i, \ldots, Z_0\right] \\ &= \mathbb{E}\left[A \mid Z_i, \ldots, Z_0\right] = X_i. \end{aligned}$$

There are two important martingales used in the theory of random graphs. The pairs of vertices (or "edges" here) are given some arbitrary order in order to define the edge-exposure martingale.

Definition 5.10.3 *Let* f *be a real-valued function of graphs. For a graph* H, *the* ***edge-exposure martingale*** *is defined as follows: Let* $Z_i(H)$ *be the indicator variable of the event that the graph has the same outcome as* H *for the* i^{th} *pair. Let* $G \sim G(n,p)$, *let* $A = X_0(H) = \mathbb{E}[f(G)]$ *and for* $i = 1, \ldots, m$,

$$\begin{aligned} X_i &= \mathbb{E}[f(G) \mid Z_i, \ldots, Z_1] \\ &= \mathbb{E}[f(G) \mid e_j \in G \textit{ iff } e_j \in H, 1 \leq j \leq i]. \end{aligned}$$

The vertices are given some arbitrary order in order to define the vertex-exposure martingale.

---

[3] No relation.

Definition 5.10.4 *Let f be a real-valued function of graphs. For a graph H, the* ***vertex-exposure martingale*** *is defined as follows: Let $Z_i(H)$ be the indicator variable where the graph has the same outcome as H for the set of pairs $\{\{i,j\} : j = 1, \ldots, i-1\}$. Let $G \sim G(n,p)$, let $A = X_1(H) = \mathbb{E}[f(G)]$ and for $i = 2, \ldots, n$,*

$$\begin{aligned} X_i &= \mathbb{E}[f(G) \mid Z_i, \ldots, Z_1] \\ &= \mathbb{E}[f(G) \mid e_j \in G \textit{ iff } e_j \in H, 1 \le j \le i]. \end{aligned}$$

Clearly, the above are Doob martingales and so it is correct to use the appellation "martingale" with them. An appropriate ordering of the edges shows that the vertex-exposure martingale can be thought of as a subsequence of the edge-exposure martingale.

5.10.5 Azuma's inequality

Azuma's inequality says, among other things, that the coin-flip gambler cannot lose or win too much money. It bears some resemblance to Chernoff bounds as well and this is because the methods are similar.

Theorem 5.10.6 (Azuma's inequality) *Let $0 = X_0, \ldots, X_n$ be a martingale such that*

$$|X_{i+1} - X_i| \le 1 \qquad \forall i, 0 \le i \le n-1.$$

For $\lambda > 0$,

$$\Pr[X_n > \lambda\sqrt{n}] < e^{-\lambda^2/2}.$$

Proof. Let $Y_i = X_i - X_{i-1}$ for $i = 1, \ldots, n$. Therefore $|Y_i| \le 1$ and

$$\begin{aligned} \mathbb{E}[Y_i \mid X_{i-1}, \ldots, X_0] &= \mathbb{E}[X_i \mid X_{i-1}, \ldots, X_0] - \mathbb{E}[X_{i-1} \mid X_{i-1}, \ldots, X_0] \\ &= X_{i-1} - X_{i-1} = 0. \end{aligned}$$

Set $\alpha = \lambda/\sqrt{n}$. Using the exercises, we see that

$$\mathbb{E}\left[e^{\alpha Y_i} \mid X_{i-1}, \ldots, X_0\right] \le \cosh(\alpha) \le e^{\alpha^2/2}.$$

So, now we have a bound on the moment generating function of X_n:

$$\begin{aligned} \mathbb{E}\left[e^{\alpha X_n}\right] &= \mathbb{E}\left[\prod_{i=1}^{n} e^{\alpha Y_i}\right] = \mathbb{E}\left[\mathbb{E}\left[\prod_{i=1}^{n} e^{\alpha Y_i} \,\middle|\, X_{n-1}, \ldots, X_0\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[e^{\alpha Y_n} \mid X_{n-1}, \ldots, X_0\right] \left(\prod_{i=1}^{n-1} e^{\alpha Y_i}\right)\right] \\ &\le \mathbb{E}\left[e^{\alpha^2/2} \left(\prod_{i=1}^{n-1} e^{\alpha Y_i}\right)\right] \\ &\le e^{\alpha^2 n/2}, \end{aligned}$$

where we arrive at this last one by repeatedly applying the conditional expectation technique in the first step.

Computing the probability using Markov's inequality:

$$\begin{aligned}\Pr(X_n \geq \alpha n) &= \Pr\left(e^{\alpha X_n} \geq e^{\alpha^2 n}\right) \leq \frac{\mathbb{E}\left[e^{\alpha X_n}\right]}{e^{\alpha^2 n}} \\ &\leq e^{\alpha^2 n/2} e^{-\alpha^2 n} = e^{-\alpha^2 n/2} = e^{-\lambda^2/2}.\end{aligned}$$

□

5.10.7 Martingales and concentration

The power of this method is that it is remarkably easy to show that certain graph properties are very tightly concentrated around their mean, without ever knowing what the mean might be.

Theorem 5.10.8 *Let $G \sim G(n,p)$ and let f and g be graph invariants with $\mu = \mathbb{E}[f(G)]$ and $\nu = \mathbb{E}[f(G)]$.*

If f has the property that for any graphs G_1 and G_2 that differ at only one vertex (that is, there is a $v_1 \in V(G_1)$ and $v_2 \in V(G_2)$ such that $G_1 \setminus \{v_1\} \approx G_2 \setminus \{v_2\}$), $|f(G_1) - f(G_2)| \leq 1$, then

$$\Pr\left(|f(G) - \mu| > \lambda\sqrt{n-1}\right) < 2e^{-\lambda^2/2}.$$

If g has the property that for any graphs G_1 and G_2 that differ at only one edge, $|g(G_1) - g(G_2)| \leq 1$, then

$$\Pr\left(|g(G) - \nu| > \lambda\sqrt{\binom{n}{2}}\right) < 2e^{-\lambda^2/2}.$$

This was used to show tight concentration of the chromatic number of the random graph in 1987. At the time, the value of the mean was not known.

Theorem 5.10.9 (Shamir and Spencer [SS87]) *Let n and p be arbitrary and set $\mu = \mathbb{E}[\chi(G)]$ where $G \sim G(n,p)$. Then*

$$\Pr\left[|\chi(G) - \mu| > \lambda\sqrt{n-1}\right] < 2e^{-\lambda^2/2}.$$

In 1988, Bollobás computed the expected value of the chromatic number of the random graph. Alon and Spencer's book [AS00] and Bollobás' *Random Graphs* [Bol01] produce more-or-less the original proof. The book *Random Graphs* by Janson, et al. [JŁR00] uses some different inequalities to obtain essentially the same result.

Theorem 5.10.10 (Bollobás [Bol88]) *Let $0 < p < 1$ be fixed, and set $q = 1 - p$ and $d = 1/q$. Then,* **whp** *, $G(n,p)$ is such that*

$$\frac{n}{2\log_d n}\left(1+\frac{\ln\ln n}{\ln n}\right) \leq \chi\left(G(n,p)\right) \leq \frac{n}{2\log_d n}\left(1+\frac{3\ln\ln n}{\ln n}\right).$$

In particular, $\chi(G(n,p)) \sim n/(2\log_d n)$, **whp**.

McDiarmid proved a more strict concentration. The notation $X_n = O_C(a_n)$ means that there exists a constant C such that $|X_n| \leq Ca_n$, **whp**.

Theorem 5.10.11 (McDiarmid [McD89]) *With the setup as in Theorem 5.10.10,*

$$\chi\left(G(n,p)\right) = \frac{n}{2\log_d n - 2\log_d \log_d n + O_C(1)}$$

and the term $O_C(1)$ can be replaced by $O_C(1/p)$ as long as $p > n^{-\delta}$ for each $\delta > 0$.

For sparse random graphs, the concentration is even more striking.

Theorem 5.10.12 (Alon-Krivelevich [AK97]) *For every positive constants ϵ, δ, there exists $n_0 = n_0(\epsilon, \delta)$ such that for every $n > n_0$ and probability $p = n^{-1/2-\delta}$ there is an integer $t = t(n,p,\epsilon)$ such that*

$$\Pr\left(t \leq \chi(G(n,p)) \leq t+1\right) \geq 1-\epsilon.$$

In other words, for every $\alpha > 1/2$, the chromatic number of $G(n,p)$ with $p = n^{-\alpha}$ takes on one of two values, **whp**.

This is an improvement of the result of Łuczak [Łuc91], which proved the same result for $p = n^{-\alpha}$ for all $\alpha > 5/6$.

5.10.13 Another Chernoff-type bound for binomial random variables

One of the problems with Theorem 5.4.4 is that there is a bound on the λ involved. Azuma's inequality frees us of this restriction. Note that if Y_n is a $\mathrm{bin}(n,p)$ random variable, then with $X_n = Y_n - np$, $0 = X_0, X_1, X_2, \ldots, X_n$ can be thought of as a martingale. That is, for $n \geq 1$, $X_n = X_{n-1} + (Z_{n-1} - p)$ where Z_{n-1} is a $\mathrm{Ber}(p)$ random variable.

We can apply Theorem 5.10.6 to this martingale and Corollary 5.10.14

Corollary 5.10.14 *Let n be an integer, $p \in [0,1]$ and $Y_n \sim \mathrm{bin}(n,p)$. Then*

$$\Pr\left(|Y_n - np| > \lambda\sqrt{n}\right) < 2\exp\{-\lambda^2/2\}.$$

Exercises.

(1) For $i = 1, \ldots, m$, let Y_i be a random variable so that $(Y_i+1)/2 \sim \mathrm{Ber}(1/2)$. That is, $\Pr(Y_i = -1) = \Pr(Y_i = 1) = 1/2$. Prove that if $X_0 = 0$ and $X_i = Y_i + \cdots + Y_1$, then $X_0, X_1, \ldots, X_m$ is a martingale.

(2) An amusing philosophical diversion is the **St. Petersburg paradox**. The game is simple. A coin is flipped until heads occurs. If heads occurs on the first flip, the house pays \$2. If the first flip is tails and the second heads, the house pays $\$2^2$=\$4. In general, if heads occurs first on the i^{th} flip, then the house pays $\$2^i$. Would you pay \$1000 to play this game? Why or why not? Robert Martin[4] says that most people would not even pay \$25 to play this game!

(3) Prove, for a positive real number α and a real number y such that $|y| \leq 1$, that

$$e^{\alpha y} \leq \cosh(\alpha) + \sinh(\alpha) y. \tag{5.5}$$

Hence, if Y is a random variable with $|Y| \leq 1$, then $\mathbb{E}[e^{\alpha Y}] \leq \mathbb{E}[\cosh(\alpha) + \sinh(\alpha) Y] = \cosh(\alpha) + \sinh(\alpha)\mathbb{E}[Y]$.

(4) Prove that for any real number x, $\cosh(x) \leq e^{x^2/2}$.

(5) The minimum rank of an n-vertex graph G, denoted $\mathrm{mr}(G)$, is the minimum rank over all $n \times n$ matrices A such that, for distinct i, j, if $A_{ij} = 0$ then vertex i is nonadjacent to vertex j in G. Note that the diagonal entries in A can be arbitrary. A **principle minor** (of size $n-1$) of A is obtained by deleting the i^{th} row and i^{th} column of A. Prove that, for any principle minor, A' of A that

$$\mathrm{rank}(A) - 2 \leq \mathrm{rank}(A') \leq \mathrm{rank}(A).$$

Use this fact to show that, if $G \sim G(n,p)$ then $\mathrm{mr}(G)$ is tightly concentrated around its mean. It has been recently shown that $0.1469n - o(n) \leq \mathbb{E}\left[G(n, 1/2)\right] \leq n/2 + o(n)$. [HHMS10]

5.11 Gems: Entropy Method

One of the more interesting uses of the probabilistic method makes use of the entropy function. Entropy is said to measure the amount of "information" in a system. We will discuss this paradigm a little later. We direct the reader to a manuscript of Radhakrishnan [Rad01] which surveys the main results in an efficient way.

Definition 5.11.1 *Let X be a random variable whose range is $\{x_1, \ldots, x_n\}$. The* ***binary entropy function*** *of X is defined to be*

$$H[X] = H_2[X] \overset{\text{def}}{=} \sum_{i=1}^{n} \Pr(X = x_i) \log_2 \left(\frac{1}{\Pr(X = x_i)} \right).$$

---

[4] I am quite serious, no relation!

This definition can be extended to define $H_b[X]$ by changing the base of the logarithm from 2 to b. It can also be generalized to continuous probability distributions, but such a generalization is

Take note that the set of values that the random variable takes on doesn't matter. For example, for a Bernoulli random variable, $\mathrm{Ber}(p)$ and $\mathrm{Ber}(1-p)$ have the same binary entropy function, $-\log_2\left(p^p(1-p)^{(1-p)}\right)$.

Proposition 5.11.2 *Let X be a random variable whose range is $x_1,\ldots,x_n$. Then,*

$$H[X] \leq \log_2 n$$

with equality if and only if X is the uniform distribution.

Proof. The result is trivial for $n=1$. For $n \geq 2$, we just observe that the function $\varphi(x) = x\log_2(1/x)$ is concave down for $x \in (0,1)$. So we can apply Jensen's inequality which says that if φ is a function that is concave down and X is a random variable, then $\mathbb{E}\left[\varphi(X)\right] \leq \varphi\left(\mathbb{E}[X]\right)$, with equality if and only if X is constant. Consequently,

$$\frac{1}{n}\sum_{i=1}^{n}\varphi(p_i) \leq \varphi\left(\frac{1}{n}\sum_{i=1}^{n}p_i\right) = \varphi\left(\frac{1}{n}\right) = \frac{1}{n}\log_2 n.$$

Therefore, $\sum_{i=1}^{n} p_i\log_2(1/p_i) \leq \log_2 n$, with equality if and only if $p_i = 1/n$ for $i = 1,\ldots,n$. ☐

The reason entropy is useful is found via conditioning. In order to define this, we define a new random variable.

Definition 5.11.3 *Given discrete random variables X and Y, there is a* ***conditional random variable****, X_Y, which is a function of Y. For each value y in the range of Y, the probability mass function of X_y is defined as follows:*

$$\Pr(X_y = x) = \Pr(X = x \mid Y = y).$$

The ***conditional entropy*** *of X given Y is*

$$H[X \mid Y] \overset{\text{def}}{=} \mathbb{E}_Y[H[X_Y]].$$

Proposition 5.11.4 lists some important facts about conditional entropy. We leave the proofs of (2) and (4) to the exercises. In interpreting Proposition 5.11.4, it is important to keep in mind the paradigm that entropy measures uncertainty.

Proposition 5.11.4 *Let X, Y, Z and $X_1,\ldots,X_n$ be discrete random variables.*

(1) $H[XY] \leq H[(X,Y)] = H[X] + H[Y \mid X]$.

(2) $H[(X,Y) \mid Z] = H[X \mid Z] + H[Y \mid (X,Z)]$.

(3) $H[X] \geq H[X \mid f(Y)] \geq H[X \mid Y] \geq 0$ *for any function* f.

(4) $H\left[(X_1, \ldots, X_n) \mid Y\right] \leq \sum_{i=1}^{n} H[X_i \mid Y]$

Proof of Proposition 5.11.4. Let the range of X be $x_1, x_2, \ldots$, the range of Y be $y_1, y_2, \ldots$.

(1):
Let the range of XY be $w_1, w_2, \ldots$ and $q_k = \Pr(XY = w_k)$. Let $p_{ij} = \Pr\left(\{X = x_i\} \wedge \{Y = y_j\}\right)$ and, for all such k,

$$0 \leq \sum_{i,j:x_i y_j = w_k} p_{ij} \log_2\left(\frac{q_k}{p_{ij}}\right) = -\sum_{i,j:x_i y_j = w_k} p_{ij} \log_2 p_{ij} + q_k \log_2 q_k.$$

Summing over all values of k,

$$0 \leq -\sum_{ij} p_{ij} \log_2 p_{ij} + \sum_{k} q_k \log_2 q_k = H\left[(X, Y)\right] - H[XY].$$

Now we turn to the equality,

$$\begin{aligned}
H\left[(X,Y)\right] &= -\sum_{ij} p_{ij} \log_2 p_{ij} \\
&= -\sum_{ij} p_{ij} \log_2\left(\Pr\left(Y = y_j \mid X = x_i\right) \Pr\left(X = x_i\right)\right) \\
&= -\sum_{ij} p_{ij} \log_2 \Pr\left(Y = y_j | X = x_i\right) - \sum_{ij} p_{ij} \log_2 \Pr\left(X = x_i\right) \\
&= \sum_{i} \Pr\left(X = X_i\right) \left\{ -\sum_{j} \Pr\left(Y_{x_i} = y_j\right) \log_2 \Pr\left(Y_{x_i} = y_j\right) \right\} \\
&\quad -\sum_{i} \left\{ \Pr\left(X = x_i\right) \log_2 \Pr\left(X = x_i\right) \sum_{j} \Pr\left(Y = y_j \mid X = x_i\right) \right\} \\
&= \sum_{i} \Pr\left(X = x_i\right) H[Y_{x_i}] - \sum_{i} \Pr\left(X = x_i\right) \log_2 \Pr\left(X = x_i\right) \\
&= \mathbb{E}_X[Y_X] + H[X] \\
&= H[Y \mid X] + H[X].
\end{aligned}$$

(3):
Let the range of X be $x_1, x_2, \ldots$ and let $W = f(Y)$ have the range $w_1, w_2, \ldots$. Let $p_{ik} = \Pr\left(\{X = x_i\} \wedge \{W = w_k\}\right)$. We use the fact that $\ln x \leq x - 1$ for all

positive values of x.

$$\begin{aligned}
&H[X \mid W] - H[X] \\
&\quad = \quad -\sum_k \Pr(W = w_k) \sum_i \Pr(X = x_i \mid W = w_k) \log_2 \Pr(X = x_i \mid W = w_k) \\
&\qquad\qquad + \sum_i \Pr(X = x_i) \log_2 \Pr(X = x_i) \\
&\quad = \quad -\sum_i \sum_k p_{ik} \log_2 \Pr(X = x_i \mid W = w_k) + \sum_i \sum_k p_{ik} \log_2 \Pr(X = x_i) \\
&\quad = \quad \sum_i \sum_k p_{ik} \log_2 \left[\frac{\Pr(X = x_i)}{\Pr(X = x_i \mid W = w_k)} \right] \\
&\quad \leq \quad \frac{1}{\ln 2} \sum_i \sum_k p_{ik} \left[\frac{\Pr(X = x_i) \Pr(W = w_k)}{p_{ik}} - 1 \right] \\
&\quad = \quad \frac{1}{\ln 2} \sum_{i=1}^{n} \sum_{k=1}^{p} \left[\Pr(X = x_i) \Pr(W = w_k) - \Pr(X = x_i \mid W = w_k) \right] = 0
\end{aligned}$$

Similar calculations can show that $\mathbb{E}[X \mid Y] \leq \mathbb{E}[X \mid f(Y)]$. □

The other key tool we need is a lemma attributed to Shearer (see [CGFS86]):

Lemma 5.11.5 (Shearer's Lemma) *Let* $\mathbf{X} = (X_1, \ldots, X_n)$ *be a random variable and let* $\mathcal{A} = \{A_j\}_{j \in J}$ *be a collection of subsets of* $[n]$*, such that each element of* $[n]$ *appears in at least* k *members of* $\mathcal{A}$*. For any subset* $S \subseteq [n]$*,* $\mathbf{X}_S$ *denotes the vector* $(X_i : i \in S)$*. Then,*

$$\sum_{j \in J} H[\mathbf{X}_{A_j}] \geq kH[\mathbf{X}].$$

Proof. Using Proposition 5.11.4

$$\begin{aligned}
H\left[\mathbf{X}_{A_j}\right] &= \sum_{i \in A_j} H\left[X_i \mid (X_\ell : \ell \in A_j, \ell < i)\right] \\
\sum_{j \in J} H\left[\mathbf{X}_{A_j}\right] &= \sum_{j \in J} \sum_{i \in A_j} H\left[X_i \mid (X_\ell : \ell \in A_j, \ell < i)\right] \\
&= \sum_{i=1}^{n} \sum_{A_j \ni i} H\left[X_i \mid (X_\ell : \ell \in A_j, \ell < i)\right] \\
&\geq \sum_{i=1}^{n} \sum_{A_j \ni i} H\left[X_i \mid (X_\ell : \ell < i)\right] \qquad (5.6)
\end{aligned}$$

Since each $i \in [n]$ appears in at least k of the A_j's, the right hand side of (5.6) is at least $k \sum_{i=1}^{n} H\left[X_i \mid (X_\ell : \ell < i)\right]$. By repeated application of Proposition 5.11.4 (1), we may conclude that

$$\sum_{i=1}^{n} H\left[X_i \mid (X_\ell : \ell < i)\right] = H[\mathbf{X}].$$

Putting this value into (5.6) finishes the proof. □

Theorem 5.11.6 by Kahn [Kah01] is a classical example of how entropy can be used to solve graph theory problems. We want to also note a generalization to graph homomorphisms by Galvin and Tetali [GT04].

Theorem 5.11.6 (Kahn [Kah01]) *Let $G = (A, B; E)$ be a d-regular bipartite graph such that $|A| = |B| = n/2$ and d divides $n/2$. If $\mathcal{I}(G)$ denotes the number of independent sets of G, then*

$$|\mathcal{I}(G)| \leq \left(2^{d+1} - 1\right)^{n/(2d)}, \tag{5.7}$$

with equality if and only if G consists of $n/(2d)$ copies of $K_{d,d}$.

Proof. Let I be a random variable representing an independent set of G chosen uniformly at random. For $v \in V(G)$, let X_v be an indicator random variable indicating whether I is an independent set containing v. That is, $X_v = 1$ if $I \ni v$ and is 0 otherwise. For any subset S of the vertices, we define a vector $X_S = (X_v : v \in S)$ and can associate I with its characteristic vector.

By first applying Proposition 5.11.4 (1), and then applying Shearer's Lemma, Lemma 5.11.5, we have

$$\begin{aligned} \log_2 |\mathcal{I}(G)| = H[I] &= H[X_A \mid X_B] + H[X_B] \\ &\leq H[X_A \mid X_B] + \frac{1}{d}\sum_{v \in A} H[X_{N(v)}]. \end{aligned}$$

By applying Proposition 5.11.4 (4), we see that

$$\begin{aligned} \log_2 |\mathcal{I}(G)| &\leq \sum_{v \in A} H[X_v \mid X_B] + \frac{1}{d}\sum_{v \in A} H[X_{N(v)}] \\ &= \sum_{v \in A}\left(H[X_v \mid X_B] + \frac{1}{d}H[X_{N(v)}]\right) \end{aligned} \tag{5.8}$$

For each $v \in A$, define the random variable ν_v as follows:

$$\nu_v = \begin{cases} 1, & \text{if } X_{N(v)} \neq \mathbf{0}\text{; and} \\ 0, & \text{if } X_{N(v)} = \mathbf{0}. \end{cases}$$

That is, ν_v is 1 if the independent set, I, contains at least one member of $N(v)$ and ν_v is 0 if the independent set contains no member of $N(v)$.

For each $v \in A$, let $q = \Pr(\nu_v = 0)$. By Proposition 5.11.4(3),

$$H[X_v \mid X_{N(v)}] \leq H[X_v \mid \nu_v].$$

By definition, $H[X_v \mid \nu_v] = \sum_{i=0}^{1} \Pr(\nu_v = i) H[X_v \mid \nu_v = i]$. If $\nu_v = 1$, then X_v is identically zero and its entropy function is 0. So, by Proposition 5.11.2,

$$H[X_v \mid X_{N(v)}] \leq H[X_v \mid \nu_v] = qH[X_v \mid \nu_v = 0] \leq q. \tag{5.9}$$

Furthermore, by Proposition 5.11.4 (1),

$$\begin{aligned} H[X_{N(v)}] = H[X_{N(v)}\nu_v] &\leq H[\nu_v] + H[X_{N(v)} \mid \nu_v] \\ &\leq H(q) + (1-q)\log_2(2^d - 1). \end{aligned} \tag{5.10}$$

Plugging (5.9) and (5.11) into (5.8), we obtain

$$\begin{aligned} \log_2 |\mathcal{I}(G)| &\leq \sum_{v \in A} \left(q + \frac{1}{d} \left[H(q) + (1-q)\log_2(2^d - 1) \right] \right) \\ &= \frac{n}{2} \left(q + \frac{1}{d} \left[H(q) + (1-q)\log_2(2^d - 1) \right] \right) \end{aligned} \tag{5.11}$$

Because the maximum of (5.11) occurs for $q = 2^d/(2^{d+1} - 1)$, we obtain

$$\log_2 |\mathcal{I}(G)| \leq \frac{n}{2} \left(\frac{1}{d} \log_2 \left(2^{d+1} - 1 \right) \right).$$

Exponentiate both sides and the theorem is proven. $\square$

After almost two decades of being open, a conjecture of Alon (made in 1991) and Kahn (restated in 2001) was that inequality (5.7) holds for all d-regular n-vertex graphs for which n is even and d divides $n/2$. The proof by Yufei Zhao uses elementary methods.

Theorem 5.11.7 (Yufei Zhao [Zha]) *Let G be a d-regular n-vertex graph such that n is even and d divides $n/2$. If $\mathcal{I}(G)$ denotes the number of independent sets of G, then*

$$|\mathcal{I}(G)| \leq \left(2^{d+1} - 1 \right)^{n/(2d)},$$

with equality if and only if G consists of $n/(2d)$ copies of $K_{d,d}$.

Proof. Given G on vertex set $V = \{v_1, \ldots, v_n\}$, first look at the graph $G+G$ – the disjoint union of two copies of G – and it is clear that $|\mathcal{I}(G)|^2 = |\mathcal{I}(G+G)|$. Next consider the graph $H = G \times K_2$. This is a bipartite graph $H = (X, Y; E)$ where X and Y are copies of V and $x_i \in X$ is adjacent to $y_j \in Y$ if and only if $v_i \sim v_j$.

We want to create an injection ϕ from $\mathcal{I}(G+G)$ to $\mathcal{I}(G \times K_2)$. The way this is done is as follows. Let the independent set be $A + B$ with $A \subseteq X$ and $B \subseteq Y$. When A and B are viewed in a natural way as subsets of $V(G)$ – call the corresponding sets in $V(G)$ A' and B', respectively – the set $A' \cup B' \subseteq V(G)$ is bipartite in G. In particular, any edge must have one endpoint in $A' - B'$ and the other in $B' - A'$.

For any independent sets A', B' in $V(G)$, let S be the lexicographically first set in $A' \triangle B'$ such that every edge in $A' \cup B'$ has exactly one endpoint in S and one endpoint in $(A' \triangle B') - S$. As we have seen, such a set much exist.

For independent set $A + B$ in $G + G$ such that $A \subseteq X$ and $B \subseteq Y$, construct $\phi(A + B)$ as follows:

> Consider the corresponding sets A' and B' as subsets of $V(G)$. If $v_i \in A' \cap B'$, then the corresponding vertices $x_i \in X$ and $y_i \in Y$ are in $\phi(A + B)$. If $v_i \in A' \triangle B'$, then we include x_i if $v_i \in A' - S$ or $v_i \in B' \cap S$. On the other hand, we include y_i if $v_i \in B' - S$ or $v_i \in A' \cap S$. (To see it another way, we want to map A to its image in X and B to its image in Y, but we switch the vertex from X to Y and vice versa if it's in S.)

Since S is uniquely defined by $A' \triangle B'$, we can find the inverse image of every independent set in $G \times K_2$ because they are all of the form $A' + B'$. We do this by finding the unique S in $A' \triangle B'$. Given $A' \triangle B'$ and S, we are able to find the inverse image to find $A + B$ in the graph $G + G$.

By applying Kahn's theorem, Theorem 5.11.6, to $\mathcal{I}(G \times K_2)$ we have that

$$|\mathcal{I}(G)|^2 = |\mathcal{I}(G + G)| = |\mathcal{I}(G \times K_2)| \leq \left(2^{d+1} - 1\right)^{(2n)/(2d)}.$$

$\square$

Exercises.

(1) Prove Proposition 5.11.4 (2) and (4).

Chapter 6

Szemerédi's Regularity Lemma

6.1 Origins

This chapter begins our introduction to Szemerédi's regularity lemma [Sze78] originally used to prove the following:

Theorem 6.1.1 (Szemerédi [Sze75]) *For all $\epsilon > 0$ and positive integers $k \geq 1$, there exists $N(k, \epsilon)$ such that if $n > N(k, \epsilon)$ and R is a set of non-negative integers not exceeding n with $|R| > \epsilon n$, then R contains an arithmetic progression with k terms.*

This generalizes an old theorem by van der Waerden in 1927. The following is a basic consequence of the more general van der Waerden theorem.

Theorem 6.1.2 (van der Waerden [vdW27]) *If $\mathbb{Z}$ is colored by $k \geq 1$ colors, then there exists a monochromatic arithmetic progression of arbitrary length.*

So, Szemerédi shows that the monochromatic arithmetic progression occurs in each color class that is not trivially sparse (we will not bother to define positive upper density). Out of Theorem 6.1.1 came the celebrated Regularity Lemma.

There are two excellent surveys which contain much of the information we present here. The first is by Komlós and Simonovits [KS96] and the second is by Komlós, Shokoufandeh, Simonovits and Szemerédi [KSSS02]. Although the second contains updates on results covered in [KS96], the first has some basic results that the beginning student may find even more valuable than subsequent developments.

6.2 Epsilon-regular pairs

6.2.1 Random pairs

Let us construct a bipartite graph G on (A, B), $|A| = |B| = n$ with the property that for any $a \in A$ and $b \in B$, the indicator variable for the events $\{ab \text{ is an edge}\}$ are independent $\mathrm{Ber}(p)$ random variables. We will compute two probabilities. Let us recall the Chernoff bound, Corollary 5.4.7, in particular if $X \sim \mathrm{bin}(n, p)$, then for all λ, $0 \le \lambda \le 2\sqrt{np(1-p)}$,

$$\Pr\left(|X - np| \ge \lambda\sqrt{np(1-p)}\right) \le 2\exp\{\lambda^2/4\}.$$

We leave the following proposition as an exercise.

Proposition 6.2.2 *For any $p \in (0,1)$, if $G \sim G(n,p)$, then,* **whp***,*

$$np - 3\sqrt{p(1-p)n\ln n} \le \delta(G) \le \Delta(G) \le np + 3\sqrt{p(1-p)n\ln n}.$$

Next, we have a little more complicated condition. For any pair (S, T), $S \cap T = \emptyset$, we say that the **density of** (S,T) is $d(S,T) := \frac{e(S,T)}{|S||T|}$, where $e(S,T)$ is the number of edges between S and T.

We want to compute the probability that for any $X \subset A$ and $Y \subset B$ with $|X| \ge \epsilon n$ and $|Y| \ge \epsilon n$ that

$$|d(X,Y) - d(A,B)| > \epsilon.$$

Note that $e(X,Y)$ is a $\mathrm{bin}(|X||Y|, p)$ random variable. For simplicity, assume that $|A| = |B| = n$.

Therefore, we can use Boole's inequality.

$$\begin{aligned} &\Pr\left(\exists (X,Y) : |d(X,Y) - d(A,B)| > \epsilon\right) \\ &\le \binom{|A|}{\ge \epsilon|A|}\binom{|B|}{\ge \epsilon|B|} \Pr\left(|d(X,Y) - d(A,B)| > \epsilon\right) \\ &\le 2^n \cdot 2^n \left[\Pr\left(|d(X,Y) - p| > \epsilon/2\right) + \Pr\left(|d(A,B) - p| > \epsilon/2\right)\right]. \end{aligned}$$

In order to bound $\Pr\left(|d(X,Y) - p| > \epsilon/2\right)$, we use the Chernoff bound,

$$\begin{aligned} \Pr\left(|d(A,B) - p| > \epsilon/2\right) &\le 4^n \Pr\left(|e(X,Y) - p|X||Y|| > \frac{\epsilon}{2}|X||Y|\right) \\ &< 4^n \cdot 2e^{-\epsilon^2|X||Y|/8} \\ &\le 4^n \cdot 2e^{-\epsilon^4 n^2/8}. \end{aligned}$$

Similarly, $\Pr\left(|d(A,B) - p| > \epsilon/2\right) \le 4^n \cdot 2e^{-\epsilon^2|A||B|/8}$. Thus, for fixed ϵ, the probability that there exists any pair (X,Y) with density outside of $[p-\epsilon, p+\epsilon]$ goes to 0 as $n \to \infty$.

6.2.3 Regular pairs

Definition 6.2.4 *Let A and B be sets of vertices with $A \cap B = \emptyset$. The number of edges in the pair (A, B) is denoted $e(A, B)$. Recall that the* ***density of*** *(A, B), denoted $d(A, B)$ is $\frac{e(A,B)}{|A||B|}$. The pair is an* ***ϵ-regular pair of density*** *d if $d = d(A, B)$ and, for any $X \subseteq A$ and $Y \subseteq B$ with $|X| \geq \epsilon|A|$ and $|Y| \geq \epsilon|B|$,*

$$|d(X, Y) - d| \leq \epsilon.$$

We will sometimes relax the definition of a graph here. We will often refer to the pair (A, B) when we formally mean the bipartite graph G induced by edges that have one endpoint in A and the other in B.

Exercises.

(1) Use a Chernoff bound to prove Proposition 6.2.2.

6.3 The regularity lemma

Szemerédi's groundbreaking theorem (Theorem 6.1.1) on arithmetic progressions in integer sequences of positive upper density came the following lemma, written separately in 1978.

Theorem 6.3.1 (Szemerédi's regularity lemma [Sze78]) *For every $\epsilon > 0$ and positive integer m, there exist two integers $M = M(\epsilon, m)$ and $N = N(\epsilon, m)$ with the property that, for every graph G with $n \geq N$ vertices, there exists a partition of the vertex set into $\ell + 1$ classes $V = V_0 + V_1 + \cdots + V_\ell$ such that*

- $m \leq \ell \leq M$
- $|V_0| < \epsilon n$
- $|V_1| = |V_2| = \cdots = |V_\ell|$
- *For distinct $i, j \neq 0$, all but at most $\epsilon\ell^2$ of the pairs (V_i, V_j) are ϵ-regular of some density.*

It almost seems as if this is too general to be helpful. In fact, it is arguably the most powerful tool in all of graph theory. Note that the numbers M and N only depend on ϵ and m. In many applications, $m = 1$ or $m = \lceil \epsilon^{-1} \rceil$ and we see that the number of vertices is just larger than some constant, depending only on ϵ.

6.4 Proving SzemRegLem

Before we prove SzemRegLem itself, let us use some definitions:

Definition 6.4.1

- *The sets* $V_1, \ldots, V_\ell$ *are called* ***clusters***
- *The set* V_0 *is called the* ***exceptional set*** *or* ***leftover set****.*
- *If a pair is not* ϵ*-regular, we often refer to it as* ϵ***-irregular****.*
- *We call a partition* $V = V_0 + V_1 + \cdots + V_\ell$ *an* ***equitable partition*** *if* $|V_1| = \cdots = |V_\ell|$*.*
- *Let* V *have an equitable partition,* P*, labelled* $V = V_0 + V_1 + \cdots + V_\ell$*. The* ***index of*** P *is defined by*

  $$\operatorname{index}(P) = \frac{1}{\ell^2} \sum_{i=1}^{\ell} \sum_{j=i+1}^{\ell} d^2(V_i, V_j)$$

  Observe that $0 \le \operatorname{index}(P) < 1/2$*.*

The main lemma that proves the regularity lemma simply says that an ϵ-irregular equitable partition can be refined so that the index increases by the fixed constant $\epsilon^5/20$. This can be done at most $10\epsilon^{-5}$ times and so

Lemma 6.4.2 (Main Lemma) *Let* $G = (V, E)$ *be a graph on* n *vertices. Let* P *be an equitable partition of* $V = V_0 + V_1 + \cdots + V_\ell$*, with exceptional class* V_0*. Let* ℓ *be a positive integer with* $4^\ell > 200\epsilon^{-5}$*.*

If more than $\epsilon\ell^2$ *pairs* (V_i, V_j) *with* $1 \le i < j \le \ell$ *are* ϵ*-irregular, then there exists an equitable partition* Q *of* V *into* $1 + \ell 4^\ell$ *classes with* $|Q_0| \le |V_0| + n/4^\ell$ *such that* $\operatorname{index}(Q) \ge \operatorname{index}(P) + \epsilon^5/20$*.*

Furthermore, Q *is a refinement of* P*, which is to say each nonexceptional cluster of* Q *lies within a cluster of* P*.*

Using the Main Lemma, let us prove SzemRegLem:

Proof of the Regularity Lemma. Let s be the smallest positive integer such that $4^s > 200\epsilon^{-5}$, $s \ge m$ and $s \ge 2/\epsilon$.

Define $f(t) \overset{\text{def}}{=} \begin{cases} s, & \text{if } t = 0; \\ f(t-1)4^{f(t-1)}, & \text{otherwise.} \end{cases}$

Let t be the largest integer such that there exists an equitable partition P of V into $1 + f(t)$ clusters such that $\operatorname{index}(P) \ge t\epsilon^5/20$ and the size of the exceptional class is at most $\epsilon n \left(1 - 2^{-(t+1)}\right)$.

We now claim that P is ϵ-regular (i.e., all but $\epsilon\ell^2$ pairs are ϵ-regular). Otherwise, the Main Lemma (Lemma 6.4.2) implies there exists a Q with index at least $(t+1)\epsilon^5/20$ with at least $1 + f(t)4^{f(t)} = 1 + f(t+1)$ clusters (including the exceptional set), and with exceptional set size at most $\epsilon n \left(1 - 2^{-(t+1)}\right) + \frac{n}{4^{f(t)}}$. We claim that this is at most $\epsilon n \left(1 - 2^{-(t+2)}\right)$, which would contradict the maximality of t.

To verify this, we need $4^{f(t)} \ge 2^{t+2}/\epsilon$. We leave it to the reader to verify, by induction on t, that, with $s \ge 2/\epsilon \ge 1$ and $f(t)$ defined as above, that

$$4^{f(t)} \ge s2^{t+1} \ge 2^{t+2}\frac{1}{\epsilon}.$$

□

6.4.3 Proof of the Main Lemma

In order to prove Lemma 6.4.2, we need two general tools. The first is the defect form of Cauchy-Schwarz (Lemma 6.4.4) and the second is called Continuity of Density (Lemma 6.4.5).

Cauchy-Schwarz, defect form

There are two preliminary lemmas which we will use. The first is a generalization of the summation form of Cauchy-Schwarz:

Lemma 6.4.4 (Cauchy-Schwarz inequality, defect form) *If $X_1, \ldots, X_m$ are real numbers such that, for some $m < n$,*

$$\sum_{k=1}^{m} X_k = \frac{m}{n} \sum_{k=1}^{n} X_k + \delta$$

then

$$\sum_{k=1}^{n} X_k^2 \geq \frac{1}{n} \left(\sum_{k=1}^{n} X_k \right)^2 + \frac{\delta^2 n}{m(n-m)}.$$

The original non-defect form of the Cauchy-Schwarz inequality was proven for sums by Cauchy [Cau21] and the inequality for integrals was observed by his student Bouniakowsky [Bou59]. Schwarz [Sch88] extended it to a general inner product space.

Proof. Let $\overline{X} = \frac{1}{n} \sum_{k=1}^{n} X_k$ and let $Y_k = X_k - \overline{X}$.

Observe that

$$\sum_{k=1}^{n} Y_k = \sum_{k=1}^{n} \left(X_k - \overline{X} \right) = n\overline{X} - n\overline{X} = 0,$$

but if we just count the first m terms,

$$\sum_{k=1}^{m} Y_k = \sum_{k=1}^{m} \left(X_k - m\overline{X} \right) = m\overline{X} + \delta - m\overline{X} = \delta.$$

Summarizing,

$$\sum_{k=1}^{m} Y_k = \delta \qquad \text{and} \qquad \sum_{k=m+1}^{n} Y_k = -\delta \tag{6.1}$$

In addition,

$$\sum_{k=1}^{n} Y_k^2 = \sum_{k=1}^{n} \left(X_k^2 - 2X_k\overline{X} + (\overline{X})^2 \right) = \sum_{k=1}^{n} X_k^2 - n\overline{X}^2 \tag{6.2}$$

Let $S \subseteq \{1, \dots, n\}$, then the non-defect form of Cauchy-Schwarz (or Jensen's inequality for convex functions) gives that, if $\vec{y}$ is a vector with indices in S where the k^{th} entry is Y_k and $\mathbf{1}$ is the all-ones vector with $|S|$ indices, then

$$\left|\sum_S 1 \cdot Y_k\right|^2 = |\langle \mathbf{1}, \vec{y}\rangle|^2 \le \|\mathbf{1}\|^2 \, \|\vec{y}\|^2 = \left(\sum_S 1^2\right)\left(\sum_S Y_k^2\right) = |S| \sum_S Y_k^2.$$

So, using (6.1),

$$\sum_{k=1}^{m} Y_k^2 \ge \frac{1}{m}\left(\sum_{k=1}^{m} Y_k\right)^2 = \frac{\delta^2}{m} \tag{6.3}$$

$$\sum_{k=m+1}^{n} Y_k^2 \ge \frac{1}{n-m}\left(\sum_{k=m+1}^{n} Y_k\right)^2 = \frac{\delta^2}{n-m} \tag{6.4}$$

Using (6.2) as well as (6.3) and (6.4), we have

$$\sum_{k=1}^{n} X_k^2 - n\overline{X}^2 = \sum_{k=1}^{n} Y_k^2 \ge \frac{\delta^2}{m} + \frac{\delta^2}{n-m}.$$

Therefore,

$$\sum_{k=1}^{n} X_k^2 \ge n\overline{X}^2 + \frac{\delta^2 n}{m(n-m)}.$$

□

Continuity of density

The second preliminary lemma says that if a subpair is achieved by deleting a few vertices, the density does not differ by much from that of the original pair.

Lemma 6.4.5 (Continuity of density) *For $0 \le \delta \le 1/2$, if $X' \subseteq X$ and $Y' \subseteq Y$ with $|X'| \ge (1-\delta)|X|$ and $|Y'| \ge (1-\delta)|Y|$, then*

(1) $|d(X,Y) - d(X',Y')| < 2\delta$

(2) $\left|d^2(X,Y) - d(X',Y')\right| < 4\delta$

Proof. Note that it will be sufficient to prove that

$$|d(X,Y) - d(X',Y')| \le \left(1 - \frac{|X'|}{|X|}\right) + \left(1 - \frac{|Y'|}{|Y|}\right) \tag{6.5}$$

because substituting $(1-\delta)|X|$ and $(1-\delta)|Y|$ for $|X'|$ and $|Y'|$ will give the upper bound of 2δ immediately for (1). Moreover, (2) follows directly from (1) because

$$\begin{aligned} |d^2(X,Y) - d^2(X',Y')| &= |d(X,Y) + d(X',Y')||d(X,Y) - d(X',Y')| \\ &\le 2|d(X,Y) - d(X',Y')|. \end{aligned}$$

As to proving (6.5), we observe that, by the triangle inequality,

$$\begin{aligned}
|d(X,Y) - d(X',Y')| &= \frac{1}{2}\Big| d(X,Y) - d(X,Y') + d(X,Y) - d(X',Y) \\
&\qquad + d(X,Y') - d(X',Y') + d(X',Y) - d(X,Y) \Big| \\
&\leq \frac{1}{2}\Big| d(X,Y) - d(X,Y') \Big| + \frac{1}{2}\Big| d(X',Y) - d(X',Y') \Big| \\
&\quad + \frac{1}{2}\Big| d(X,Y) - d(X',Y) \Big| + \frac{1}{2}\Big| d(X,Y') - d(X',Y') \Big|
\end{aligned}$$

Let us bound one of the four terms, the rest follow similarly. The upper bound is:

$$\begin{aligned}
d(X,Y) - d(X,Y') &= \frac{e(X,Y)}{|X||Y|} - \frac{e(X,Y')}{|X||Y'|} \\
&\leq \frac{e(X,Y') + |X||Y \setminus Y|}{|X||Y|} - \frac{e(X,Y')}{|X||Y'|} \\
&= (1 - d(X,Y'))\left(1 - \frac{|Y'|}{|Y|}\right)
\end{aligned}$$

The lower bound is:

$$\begin{aligned}
d(X,Y) - d(X,Y') &\geq \frac{e(X,Y')}{|X||Y|} - \frac{e(X,Y')}{|X||Y'|} \\
&= -d(X,Y')\left(1 - \frac{|Y'|}{|Y|}\right)
\end{aligned}$$

So, $|d(X,Y) - d(X,Y')| \leq 1 - |Y'|/|Y|$ and the other densities are bounded similarly. Hence,

$$|d(X,Y) - d(X',Y')| \leq \left(1 - \frac{|Y'|}{|Y|}\right) + \left(1 - \frac{|X'|}{|X|}\right)$$

which was our goal in (6.5). □

Proving the Main Lemma

Proof of Lemma 6.4.2, the Main Lemma.

Partitioning V_i into atoms

Consider any ϵ-irregular pair (V_i, V_j), $1 \leq i < j \leq \ell$. Then we can choose $X = X(i,j)$ and $Y = Y(i,j)$ such that $X \subseteq V_i$, $Y \subseteq V_j$ $|X| = |Y| = \lceil \epsilon |V_i| \rceil = \lceil \epsilon |V_j| \rceil$ and

$$|d(X,Y) - d(V_i, V_j)| > \epsilon.$$

For every i, $1 \le i \le \ell$, we define an equivalence relation over V_i as follows:

$$\forall x, z \in V_i, \qquad x \equiv z \Longleftrightarrow x \in X(i,j) \text{ whenever } z \in X(i,j), \forall i \neq j$$

and $X(i,j) = Y(j,i)$ for $j < i$). The equivalence classes are called *atoms*. Each cluster V_i has at most $2^{\ell-1}$ atoms. Let

$$L = \left\lfloor \frac{|V_i|}{4^\ell} \right\rfloor, \qquad 1 \le i \le \ell.$$

Create the partition Q so that

(1) each member of Q has cardinality L.

(2) each atom A has exactly $\lfloor |A|/L \rfloor$ members of Q,

(3) each cluster V_i, $1 \le i \le \ell$ has exactly $\lfloor |V_i|/L \rfloor$ members of Q,

This is easy to do by arbitrarily partitioning each A into pieces of size L. Combine the remaining (at most $L-1$) vertices from each atom A and collect them together. Partition this set arbitrarily into sets of size L, which leaves at most $L-1$ leftover vertices. Finally, add these to the exceptional set.

Note that, as long as $L \ge 4^\ell$, then $\lfloor |V_i|/L \rfloor = 4^\ell$, $1 \le i \le \ell$, so each V_i contains exactly 4^ℓ elements of Q and Q contains exactly $\ell 4^\ell$ members.

Furthermore, there are ℓ clusters and at most $L-1$ are placed from each into the exceptional set, the new exceptional set is now of size at most

$$|V_0| + \ell L \le |V_0| + \ell \left\lfloor \frac{n/\ell}{4^\ell} \right\rfloor \le |V_0| + \frac{n}{4^\ell}.$$

Let $q = 4^\ell$ and each member of Q in V_i $(1 \le i \le \ell)$ be

$$V_i(s), \qquad s = 1, \ldots, q = 4^\ell$$

and

$$V_i^* = \bigcup_{s=1}^{q} V_i(s).$$

Then,

$$\begin{aligned} |V_i^*| \ge |V_i| - (L-1) &= |V_i| - \left\lfloor \frac{|V_i|}{4^\ell} \right\rfloor + 1 \\ &\ge |V_i| \left(1 - \frac{1}{4^\ell}\right) + 1 \\ &\ge |V_i| \left(1 - \frac{\epsilon^5}{200}\right) \end{aligned}$$

Index change for all pairs

By continuity of density, for $1 \le i < j \le \ell$,

$$\left| d^2(V_i^*, V_j^*) - d^2(V_i, V_j) \right| < 4\frac{\epsilon^5}{200} = \frac{\epsilon^5}{50}. \tag{6.6}$$

So, given $1 \le i < j \le \ell$,

$$\begin{aligned}
\frac{1}{q^2}\sum_{s=1}^{q}\sum_{t=1}^{q} d^2(V_i(s), V_j(t)) &\ge \left[\frac{1}{q^2}\sum_{s=1}^{q}\sum_{t=1}^{q} d(V_i(s), V_j(t))\right]^2 && (6.7)\\
&= \left[\frac{1}{q^2 L^2}\sum_{s=1}^{q}\sum_{t=1}^{q} e(V_i(s), V_j(t))\right]^2\\
&= \left[\frac{1}{q^2 L^2} e(V_i^*, V_j^*)\right]^2\\
&= \left[d(V_i^*, V_j^*)\right]^2\\
&> d^2(V_i, V_j) - \frac{\epsilon^5}{50} && (6.8)
\end{aligned}$$

Inequality (6.7) comes from the (non-defect version of) Cauchy-Schwarz. Inequality (6.8) comes from inequality (6.6).

Index change for ϵ-irregular pairs

But now, consider an ϵ-irregular pair (V_i, V_j) with $1 \le i < j \le \ell$. Let $X = X(i,j)$ and $Y = Y(i,j)$. Recall $V_i^* = \bigcup_{s=1}^q V_i(s)$, $V_j^* = \bigcup_{t=1}^q V_j(t)$. Without loss of generality, we may assume that the members of Q in X and Y are, respectively,

$$\begin{aligned}
&V_i(s) \subseteq X, && s = 1, \dots, r_X\\
&V_j(t) \subseteq Y, && t = 1, \dots, r_Y
\end{aligned}$$

As a result,

$$\begin{aligned}
\left|\bigcup_{s=1}^{r_X} V_i(s)\right| &> |X| - L2^{\ell-1} && (6.9)\\
&> |X| - 2^\ell L\\
&\ge |X| - \frac{|V_i|}{2^\ell} && (6.10)\\
&\ge |X| - \frac{|X|}{\epsilon 2^\ell}\\
&\ge |X| - \frac{|X|}{\epsilon\sqrt{200\epsilon^{-5}}} \ge |X|\left(1 - \frac{\epsilon^{3/2}}{10\sqrt{2}}\right)\\
&\ge |X|\left(1 - \frac{\epsilon}{100}\right),
\end{aligned}$$

as long as $\epsilon \le 0.02$.

Inequality (6.9) comes from the fact that X splits into at most $2^{\ell-1}$ atoms and for each atom, there are less than L vertices not in one such atom.

Inequality (6.10) comes from $L = \lfloor |V_i|/4^\ell \rfloor$.

Therefore,

$$r_X = \frac{|\bigcup_{i=1}^{r_X} V_i(s)|}{L} > \frac{|X|(1-\epsilon/100)}{\lfloor |V_i|/q \rfloor} \geq \frac{|X|}{|V_i|} q \left(1 - \frac{\epsilon}{100}\right) \geq \epsilon q \left(1 - \frac{\epsilon}{100}\right).$$

Analogously, $\left|\bigcup_{j=1}^{r_Y} V_j(t)\right| > |Y|(1-\epsilon/100)$ and $r_Y > \epsilon q(1-\epsilon/100)$. Let $r = \min\{r_X, r_Y\}$, then

$$r > \epsilon q \left(1 - \frac{\epsilon}{100}\right). \tag{6.11}$$

If you define

$$X^* = \bigcup_{i=1}^{r} V_s(i) \subseteq X \qquad \text{and} \qquad Y^* = \bigcup_{j=1}^{r} V_t(j) \subseteq Y,$$

then, if $r_X \geq r_Y = r$, $|Y^*| = r\frac{|Y^*|}{r} = r\frac{|X^*|}{r_X} \leq |X^*|$. Hence, regardless of whether r_X or r_Y is bigger,

$$|X^*| > |X| \left(1 - \frac{\epsilon}{100}\right), \qquad \text{and} \qquad |Y^*| > |Y| \left(1 - \frac{\epsilon}{100}\right).$$

By continuity of density,

$$|d(X^*, Y^*) - d(X, Y)| < 2\frac{\epsilon}{100} < \frac{\epsilon}{4} \tag{6.12}$$

and so the triangle inequality, (6.12) and (6.6),

$$\left|d(X^*, Y^*) - d(V_i^*, V_j^*)\right| > \epsilon - \frac{\epsilon}{4} - \frac{\epsilon^5}{50} > \frac{\epsilon}{2} \tag{6.13}$$

Now we have the following identity:

$$\underbrace{\sum_{s=1}^{r}\sum_{t=1}^{r} d\left(V_i(s), V_j(t)\right)}_{r^2 d(X^*, Y^*)} = \underbrace{\frac{r^2}{q^2}\sum_{s=1}^{q}\sum_{t=1}^{q} d\left(V_i(s), V_j(t)\right)}_{r^2 d(V_i^*, V_j^*)} + r^2 d(X^*, Y^*) - r^2 d(V_i^*, V_j^*) \tag{6.14}$$

So, applying Lemma 6.4.4, the defect form of Cauchy-Schwarz to (6.14) with $n = q^2$, $m = r^2$ and $\delta = r^2 d(X^*, Y^*) - r^2 d(V_i^*, V_j^2)$, we get

$$\begin{aligned}
\sum_{s=1}^{q}\sum_{t=1}^{q} d^2\left(V_i(s), V_j(t)\right) \quad &\geq \quad \frac{1}{q^2}\left[\sum_{s=1}^{q}\sum_{t=1}^{q} d(V_i(s), V_j(t))\right]^2 \\
&\quad + \frac{\left(r^2 d(X^*, Y^*) - r^2 d(V_i^*, V_j^*)\right)^2 q^2}{r^2(q^2 - r^2)} \\
&= \quad \frac{1}{q^2}\left(q^2 d(V_i^*, V_j^*)\right)^2 \\
&\quad + \frac{r^2 q^2}{q^2 - r^2}\left(d(X^*, Y^*) - d(V_i^*, V_j^*)\right)^2.
\end{aligned}$$

So,

$$
\begin{aligned}
\frac{1}{q^2}\sum_{s=1}^{q}\sum_{t=1}^{q} d^2\left(V_i(s), V_j(t)\right) &\geq d^2(V_i^*, V_j^*) + \frac{r^2}{q^2 - r^2}\left(\frac{\epsilon}{2}\right)^2 && (6.15)\\
&\geq \left(d^2(V_i, V_j) - \frac{\epsilon^5}{50}\right) \\
&\quad + \frac{(\epsilon q(1-\epsilon/100))^2}{q^2 - (\epsilon q(1-\epsilon/100))^2}\frac{\epsilon^2}{4} && (6.16)\\
&\geq d^2(V_i, V_j) - \frac{\epsilon^5}{50} + \frac{\epsilon^4}{16} && (6.17)
\end{aligned}
$$

Inequality (6.15) comes from (6.13), and inequality (6.16) comes from (6.6) and (6.11). Inequality (6.17) requires $\frac{\epsilon^2(1-\epsilon/100)^2}{1-\epsilon^2(1-\epsilon/100)^2} > \frac{1}{4}$, which holds if $\epsilon < 0.4492$.

Combining the results

Now we use the fact that there are at least $\epsilon\ell^2$ pairs for which (6.17) holds. And we also use (6.8) for all pairs.

$$
\begin{aligned}
\text{index}(Q) &\geq \frac{1}{\ell^2}\sum_{i=1}^{\ell}\sum_{j=i+1}^{\ell}\left[\frac{1}{q^2}\sum_{s=1}^{q}\sum_{t=1}^{q} d^2(V_i(s), V_j(t))\right]\\
&\geq \frac{1}{\ell^2}\sum_{i=1}^{\ell}\sum_{j=i+1}^{\ell}\left[d^2(V_i, V_j) - \frac{\epsilon^5}{50}\right] + \frac{1}{\ell^2}\left(\epsilon\ell^2\frac{\epsilon^4}{16}\right)\\
&= \text{index}(P) - \frac{\binom{\ell}{2}}{\ell^2}\frac{\epsilon^5}{50} + \frac{\epsilon^5}{16}\\
&\geq \text{index}(P) - \frac{\epsilon^5}{100} + \frac{\epsilon^5}{16}\\
&\geq \text{index}(P) + \frac{\epsilon^5}{20}.
\end{aligned}
$$

□

Exercises.

(1) Prove, by induction on t, that, with $s \geq 2/\epsilon \geq 1$, $f(0) = s$ and $f(t) = f(t-1)4^{f(t-1)}$ if $t \geq 1$, that $4^{f(t)} \geq s2^{t+1} \geq 2^{t+2}\frac{1}{\epsilon}$.

6.5 Gems: Smoothed analysis of graphs

Chapter 7

Properties of Epsilon-Regular Pairs

There are some interesting properties of regular pairs emerge directly from the definition and we leave most of them as an exercise.

Proposition 7.0.1 *Let* $\epsilon' > \epsilon$. *If* (A, B) *is an* ϵ*-regular pair, then* (A, B) *is an* ϵ'*-regular pair.*

Proposition 7.0.2 *Let* G *be graph defined by the pair* (A, B) *and let* $\overline{G}$ *denote the* ***bipartite complement*** *of* G. *That is,* $\overline{G}$ *is a bipartite graph defined on the pair* (A, B) *with the property that for any* $a \in A$ *and* $b \in B$, *then* $a \sim_{\overline{G}} b$ *iff* $a \not\sim_G b$.

If G *is* ϵ*-regular with density* d, *then* $\overline{G}$ *is* ϵ*-regular with density* $1 - d$.

Proposition 7.0.3 *Let* (A, B) *be an* ϵ*-regular pair with density* d. *Let* A' *be the set of vertices with* $\deg(a) \in [(d - \epsilon)|B|, (d + \epsilon)|B|]$. *Then,* $|A'| \geq (1 - 2\epsilon)|A|$.

Exercises.

(1) Prove Proposition 7.0.1.

(2) Prove Proposition 7.0.2.

(3) Prove Proposition 7.0.3.

7.1 The Intersection Property

Proposition 7.1.1 (Intersection Property) *Let* (A, B) *be* ϵ*-regular with density* d, $0 < \epsilon < d < 1$ *and* $t \geq 1$ *be an integer. If* $Y \subseteq B$ *and* $(d-\epsilon)^{t-1}|Y| \geq \epsilon|B|$, *then*

$$\#\left\{(a_1, a_2, \ldots, a_t) \in A^t : \left|Y \cap \left(\bigcap_{i=1}^{t} N(a_i)\right)\right| < (d - \epsilon)^t|Y|\right\} \leq t\epsilon|A|^t.$$

Proof. This is proven by induction on t. Let $t = 1$ and $X = \{a \in A : |Y \cap N(x)| < (d-\epsilon)|Y|\}$. The fact that (A, B) is ϵ-regular with density d gives that $|X| < \epsilon|A|$, otherwise, since $|Y| \geq \epsilon|B|$, $d(X, Y) < (d - \epsilon)$, a contradiction to the definition of X.

Suppose the statement is true for $t - 1$.

Let $\mathcal{X} = \left\{(a_1, \ldots, a_t) \in A^t : \left|Y \cap \left(\bigcap_{i=1}^{t} N(a_i)\right)\right| < (d-\epsilon)^t|Y|\right\}$, let $\mathcal{X}_1 = \left\{(a_1, \ldots, a_{t-1}) \in A^t : \left|Y \cap \left(\bigcap_{i=1}^{t-1} N(a_i)\right)\right| < (d-\epsilon)^{t-1}|Y|\right\}$.

Let $\vec{x} = (a_1, \ldots, a_{t-1}) \notin \mathcal{X}_1$ and $Y' = Y \cap \bigcap_{i=1}^{t-1} N(x_i)$. Let $X' = X'(\vec{x}) = \{a \in A : |Y' \cap N(x)| < (d-\epsilon)|Y'|\}$. The fact that (A, B) is ϵ-regular with density d gives that $|X'| < \epsilon|A|$, otherwise, since $|Y'| \geq \epsilon|B|$, $d(X', Y') < (d-\epsilon)$, a contradiction to the definition of X.

By the inductive hypothesis,

$$|\mathcal{X}| \leq |\mathcal{X}_1||A| + \sum_{\vec{x} \in \mathcal{X} \setminus \mathcal{X}_1} |X'(\vec{x})| < (t-1)\epsilon|A|^{t-1} \cdot |A| + |A|^{t-1} \cdot (\epsilon|A|) = t\epsilon|A|^t.$$

□

A similar proof to that of the Intersection Property gives that there is only a small number of t-tuples whose common intersection is too large.

Proposition 7.1.2 *Let (A, B) be ϵ-regular with density d, $0 < \epsilon < d < 1$ and $k \geq 1$ be an integer. If $Y \subseteq B$ and $(d+\epsilon)^{t-1}|Y| \geq \epsilon|B|$, then*

$$\#\left\{\vec{a} = (a_1, a_2, \ldots, a_t) \in A^t : \left|Y \cap \left(\bigcap_{i=1}^{t} N(x_i)\right)\right| > (d+\epsilon)^t|Y|\right\} \leq t\epsilon|A|^t.$$

One application of the intersection property is that complete bipartite graphs are present in regular pairs.

Corollary 7.1.3 *Fix positive integers k, k'. If (A, B) is an ϵ-regular pair with density d, $\epsilon \ll d$ and $n = |A| = |B|$, then if n is large enough, (A, B) has a copy of $K_{k,k'}$*

The intersection property can be generalized as well. The condition on the size of Y could be tightened, but it is unnecessary for applications.

Proposition 7.1.4 (Generalized Intersection Property) *Let (A, B) be ϵ-regular with density d, $0 < \epsilon < d < 1$ and $k \geq 1$ and $\ell \in \{0, 1, \ldots, k\}$ be integers. If $Y \subseteq B$ and $(\min\{d, 1-d\} - \epsilon)^{k-1}|Y| \geq \epsilon|B|$, then*

$$\#\left\{\vec{x} \in A^k : \left|Y \cap \left(\bigcap_{i=1}^{\ell} N(x_i)\right) \cap \left(\bigcap_{i=\ell+1}^{k} \overline{N(x_i)}\right)\right| \right.$$
$$\left. \notin \left[(d-\epsilon)^\ell(1-d-\epsilon)^{k-\ell}|Y|, (d+\epsilon)^\ell(1-d+\epsilon)^{k-\ell}|Y|\right]\right\} \leq 2k\epsilon|A|^k.$$

Note: Similar to Corollary 7.1.3, the Generalized Intersection Property implies that, in an ϵ-regular pair (A, B) for which $\epsilon \ll \min\{d, 1-d\}$ and $|A|$ and $|B|$ are sufficiently large, then any fixed sized bipartite subgraph H is in (A, B) as an **induced subgraph**.

Finally, we take special notice of Proposition 7.1.1 for the special case where $k = 2$ and $Y = B$:

Corollary 7.1.5 *Let (A, B) be ϵ-regular with density d, $0 < \epsilon < d < 1$. If $(d-\epsilon) \geq \epsilon$, then*

$$\#\left\{(x_1, x_2) \in A^2 : |N(x_1) \cap N(x_2)| < (d-\epsilon)^2 |B|\right\} \leq 2\epsilon |A|^2.$$

Exercises.

1. Prove Proposition 7.1.2.
2. Prove the following, which leads directly to Corollary 7.1.3: Let k, k' be positive integers. If (A, B) is ϵ-regular with density d such that
   - $(d-\epsilon)^{k-1} \geq \epsilon$.
   - $k\epsilon |A|^k < (|A|)_k$,
   - $(d-\epsilon)^k |B| \geq k'$, and

   Then (A, B) contains a copy of $K_{k,k'}$ with k vertices in A and k' vertices in B. In particular, if $\epsilon < 1/k$ and $d \geq \epsilon^{1/(k-1)} + \epsilon$ and $|A|, |B|$ are large enough, then (A, B) has a $K_{k,k'}$.
3. Prove Proposition 7.1.4.

7.2 Subsets in regular pairs

Recall that a pair (A, B) is ϵ-irregular of density $d = d(A, B)$ if there exist $X \subseteq A$, $Y \subseteq B$ with $|X| \geq \epsilon |A|$ and $|Y| \geq \epsilon |B|$ and $|d(X, Y) - d| > \epsilon$.

Proposition 7.2.1 *Let (A, B) be an ϵ-irregular pair of density d. Then, there exist $X' \subseteq A$, $Y' \subseteq B$ with $|X'| = \lceil \epsilon |A| \rceil$ and $|Y'| = \lceil \epsilon |B| \rceil$ and $|d(X', Y') - d| > \epsilon$.*

The proof uses (7.1) and averaging.

$$\sum_{\substack{X' \subseteq A \\ |X'| = \lceil \epsilon |A| \rceil}} \sum_{\substack{Y' \subseteq A \\ |Y'| = \lceil \epsilon |B| \rceil}} e(X', Y') = \binom{|X| - 1}{\lceil \epsilon |A| \rceil - 1} \binom{|Y| - 1}{\lceil \epsilon |B| \rceil - 1} e(X, Y) \tag{7.1}$$

The so-called Slicing Lemma is a straightforward way to show that arbitrary subpairs are regular if they are large enough. The Slicing Lemma has a number of applications, which we will see in the next chapter.

Lemma 7.2.2 (Slicing lemma) *Given* ϵ, α, d *such that* $0 < \epsilon < \alpha < 1$ *and* $d, 1-d \geq \max\{2\epsilon, \epsilon/\alpha\}$*. Let* (A, B) *be an* ϵ*-regular pair with density* d*,* $A' \subseteq A$ *with* $|A'| \geq \alpha|A|$ *and* $B' \subseteq B$ *with* $|B'| \geq \alpha|B|$*. Then* (A', B') *is* ϵ'*-regular with* $\epsilon' = \max\{2\epsilon, \epsilon/\alpha\}$ *and density in* $[d-\epsilon, d+\epsilon]$*.*

Exercises.

(1) Use (7.1) to prove Proposition 7.2.1.

(2) Prove the Slicing Lemma (Lemma 7.2.2).

7.3 Mean and variance implies regularity

7.4 Gems: Random slicing and fractional packing

Chapter 8

Subgraph Applications of the Regularity Lemma

For many of these applications, we will be investigating a number of properties of large graphs. As such, the notation G_n denotes a graph G on n vertices.

There is some important notation that comes into play when the regularity lemma is used:

Definition 8.0.1 *We say that ϵ is small enough relative to δ and write $\epsilon \ll \delta$ if there is a function f of δ such that $\epsilon < f(\delta)$.*

The "$\ll$" notation is used for parameters in many theorems that use the regularity lemma.

8.1 Erdős-Stone-Simonovits

Theorem 8.1.1 is due to Erdős and Stone in 1946. It was generalized by Erdős and Simonovits (Theorem 8.1.2) in 1966 and is, hence, usually called Erdős-Stone-Simonovits. The theorem predates Szemerédi's regularity lemma and did not use it. A proof of Erdős-Stone that does not use the regularity lemma can be found in *Combinatorial Geometry* by Pach and Agarwal [PA95].[1]

For integers $p \geq 2$ and $t \geq 1$, we use $K_p(t)$ to denote the complete p-partite graph with t vertices in each partite set.

Theorem 8.1.1 (Erdős-Stone, [ES46]) *Let $p \geq 2$ and $t \geq 1$ be integers and let $\epsilon > 0$. There exists an $n_0 = n_0(p, t, \epsilon)$ such that if a graph G has at least $n \geq n_0$ vertices and $e(G) \geq \left(1 - \frac{1}{p-1} + \epsilon\right) \frac{n_0^2}{2}$, then there exists a copy of $K_p(t)$ in G.*

---

[1] Is it true?

In the case of $t = 1$, this is nothing more than an immediate consequence of Turán's theorem (Theorem 3.3.1):

$$e(G_n) \geq \left(1 - \frac{1}{p-1}\right)\frac{n^2}{2} \qquad \Rightarrow \qquad G_n \supseteq K_p.$$

For a family of graphs $\mathcal{F}$, we use the extremal notation $\mathrm{ex}(n, \mathcal{F})$ to denote the maximum number of edges in a graph G_n that has no $F \in \mathcal{F}$ as a subgraph.

Theorem 8.1.2 (Erdős-Simonovits, [ES66]) *Let $\mathcal{F}$ be a family of graphs such that* $\min\{\chi(F) : F \in \mathcal{F}\} = p > 2$

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

For our proof of Erdős-Stone, we will use Lemma 8.1.3, a statement which will be following will be generalized later by the Key Lemma, but it is worth investigating it in detail so as to see how Szemerédi's regularity lemma is applied in these contexts.

Lemma 8.1.3 *Let $p \geq 2$ and $t \geq 1$ be integers. Then, for all $d \in (0,1)$, there exist $\epsilon_0 = \epsilon_0(d,t,p)$, and $L_0 = L_0(d,t,p)$ such that whenever $\epsilon \leq \epsilon_0$ and $(V_1, \ldots, V_p)$ is a p-tuple that is pairwise ϵ-regular with density at least d and $|V_1| = \ldots = |V_p| = L \geq L_0$, it contains a copy of $K_p(t)$.*

I.e., if $\epsilon \ll d$ and L is large enough, then a copy of $K_p(t)$ is present in a balanced p-tuple with enough vertices in each part that pairwise is ϵ-regular with density at least d contains a copy

Proof. We will proceed by induction on p. For the base case of $p = 2$, we have a pair which is ϵ-regular with density $d \ll \epsilon$, so we just use the intersection property (Proposition 7.1.1). Here we need $(d-\epsilon)^{t-1}L \geq \epsilon L$, $(d-\epsilon)^t L \geq t$ and $t\epsilon L^t < L^t$ in order to force a copy of $K_{t,t} = K_2(t)$. Choosing $\epsilon_0(d,t,2) = \min\{(d/2)^t, 1/t\}$ and $L_0(d,t,2) = t/(d/2)^t$ suffices

Now, suppose the statement is true for $p \geq 2$, we have $\epsilon_0 = \epsilon_0(d,t,p)$ and $L_0 = L_0(d,t,p)$, so we need to prove it for $p+1$, and obtain $\epsilon_0(d,t,p+1)$ and $L_0(d,t,p+1)$.

Again, we use the intersection property. If $(d-\epsilon)^{t-1}L \geq \varepsilon L$ then we can apply the intersection property to every pair (V_i, V_{p+1}) for $i = 1, \ldots, p$. Consequently,

$$\#\left\{\vec{a} \in V_{p+1}^t : \left|\bigcap_{i=1}^{t} N(a_i)\right| \leq (d-\epsilon)^t L \text{ for any } V_i \text{ with } 1 \leq i \leq p\right\} \leq tp\epsilon L^t.$$

If $tp\epsilon < 1$, $\exists$ t vertices with neighborhoods of size at least $(d-\epsilon)^t L$ in each of $V_1, \ldots, V_p$.

Let $V_i' \subseteq V_i$, with $|V_i'| \geq \lceil (d-\epsilon)^t L \rceil$. As long as $\lceil (d-\epsilon)^t L \rceil \geq \epsilon L$, for $i \neq j$, the pair (V_i', V_j') is $\max\left\{2\epsilon, \frac{\epsilon}{(d-\epsilon)^t}\right\}$-regular of density greater than $(d-\epsilon)$.

We need $(d-\epsilon)^t \geq \epsilon$. If $2\epsilon < \epsilon_0(d-\epsilon, t, p)$ and $\dfrac{\epsilon}{(d-\epsilon)^t} < \epsilon_0(d-\epsilon, t, p)$ and $(d-\epsilon)^t < L_0(d-\epsilon, t, p)$, then we can apply the inductive hypothesis, and see that $(V_1', \dots, V_p')$ contains $K_p(t)$, giving a copy of $K_{p+1}(t)$ in the original graph.

Summarizing the requirements for ϵ_0 and L_0, ensure that

$$\epsilon_0(d, t, p+1) \leq \min\left\{1/2, (d-\epsilon)^t\right\} \cdot \epsilon_0(d, t, p)$$

and

$$L_0(d, t, p+1) \leq \frac{1}{(d-\epsilon)^t} L_0(d-\epsilon, t, p).$$

Setting $\epsilon_0(d, t, p) = \min\{(d/p)^t, 1/((p-1)t)\}$ and $L_0(d, t, p) = t/(d/p)^t$ suffices. □

It takes a good deal of work to determine values of ϵ_0 and L_0 to make sure that the theorem works, but if we simply say that $\epsilon \ll d$ and L is large enough, then the existence of $K_p(t)$ is guaranteed.

Now we return to Erdős-Stone:

Proof of Theorem 8.1.1. We construct a graph on the clusters $V_1, \dots, V_\ell$ of the Szemerédi partition of G_n. This is a common technique and this graph is called variously the **cluster graph**, **reduced graph** or **Szemerédi graph**. For this proof, we will denote it as G^{Sz}. The literature sometimes refers to this graph as G_r.

Apply Szemerédi's regularity lemma to G with ε' (which we will choose later) and $m = \lceil (\varepsilon')^{-1} \rceil$. For some $d \in (0,1)$ (again, chosen later) and Szemerédi partition $V_0 + V_1 + \cdots + V_\ell$, define G^{Sz} on the vertex set $v_1, \dots, v_\ell$ such that $v_i \sim v_j$ if and only if the pair (V_i, V_j) is ϵ-regular with density at least d.

First, we will create G_n' by deleting edges from G_n that are (1) incident to V_0, (2) in an irregular pair or (3) inside a cluster. The number of such edges is at most

(1) $\epsilon' n \cdot n \leq \epsilon' n^2$

(2) $\epsilon' \ell^2 \cdot L^2 \leq \epsilon' n^2$

(3) $\ell \binom{L}{2} \leq \frac{(\ell L)^2}{2\ell} \leq \frac{n^2}{2m} \leq \epsilon' n^2/2$.

Consequently the total amount of such edges is less than $3\epsilon' n^2$.

Now, we count the number of edges in G^{Sz}.

$$e(G) - 3\epsilon' n^2 \leq e(G_n') \leq \binom{\ell}{2} dL^2 + e(G^{Sz})(L^2 - dL^2) \leq \frac{d}{2} n^2 + e(G^{Sz})(1-d)L^2.$$

Consequently,

$$\begin{aligned} e(G^{Sz}) &\geq \frac{e(G) - (3\epsilon' + d/2)n^2}{(1-d)L^2} \\ &\geq \frac{(1 - 1/(p-1) + \epsilon)(n^2/2) - (3\epsilon' + d/2)}{(1-d)L^2} \\ &\geq \left(1 - \frac{1}{p-1}\right)\frac{n^2}{2} + \left(\frac{\epsilon}{2} - 3\epsilon' - \frac{d}{2}\right)\ell^2. \end{aligned}$$

So, if $3\epsilon' + d/2 < \epsilon$, then we can apply Turán's theorem to G^{Sz}, which implies G^{Sz} has a p-clique. We can apply Lemma 8.1.3 to the clusters that correspond to this clique, which yields the copy of $K_p(t)$. □

8.2 Degree form and number of copies of a graph

Szemerédi's regularity lemma has a number of different forms, the most useful for embedding problems is

Theorem 8.2.1 (Regularity lemma, degree form) *For every $\epsilon > 0$, there is an $M = M(\epsilon)$ such that if $G = (V, E)$ is any graph on n vertices and $d \in [0, 1]$ is any real number, then there is a partition of the vertex set V into $\ell+1$ clusters $V_0 + V_1 + \cdots + V_\ell$ and there is a subgraph $G' \subseteq G$ such that*

- $\ell \leq M$
- $|V_0| < \epsilon n$
- $|V_1| = |V_2| = \cdots = |V_\ell| \leq \epsilon n$
- $\deg_{G'}(v) > \deg_G(v) - (d + \epsilon)n$
- $e(G'[V_i]) = 0$
- ***ALL*** *pairs of clusters are ϵ-regular with density either 0 or at least d.*

We leave the proof of Theorem 8.2.1 to the exercises, but note that, like the original form (Theorem 6.3.1), this theorem will refine an existing equipartition. For example, if G is bipartite, we may assume that each of the clusters $V_1, \ldots, V_\ell$ spans only one partite set. Only the leftover set V_0 will span multiple partite sets.

8.2.2 Number of copies of a graph

The survey KS96 gives a number of immediate and insightful applications of the degree form. We use $\|H \to G_n\|$ to denote the number of embeddings of H in G_n. I.e., the number of copies of H in G_n.

Theorem 8.2.3 (Number of copies of H) *Let H be a graph with h vertices and chromatic number p. Let $\beta < 0$ and $\epsilon = (\beta/6)^h$. If n is large enough and $e(G_n) \geq \left(1 - \frac{1}{p-1} + \beta\right) \frac{n^2}{2}$, then $\|H \to G_n\| \geq \left(\frac{\epsilon n}{M(\epsilon)}\right)^h$, where $M(\epsilon)$ is given by Theorem 8.2.1.*

The proof of Theorem 8.2.3 is left as an exercise, given the following Key Lemma:

Lemma 8.2.4 (Key Lemma) *Let $d > \epsilon \geq 0$. Given a graph R and a positive integer t, the graph $R(t)$ is constructed by replacing each vertex of R by t vertices and each edge by $K_{t,t}$ (nonedges are replaced with empty bipartite graphs). Given a graph R and a positive integer L, construct G by replacing each vertex of R by L vertices and each edge with ϵ-regular pairs of density at least d. Let H be a subgraph of $R(t)$, $h = |V(H)|$ and maximum degree $\Delta = \Delta(H) > 0$. Let $\delta = d - \epsilon$ and $\epsilon_0 = \frac{\delta^\Delta}{\Delta+2}$. If $\epsilon < \epsilon_0$ and $t - 1 \leq \epsilon_0 L$, then*

$$\|H \to G\| > (\epsilon_0 L)^h.$$

Proof. We will prove the following statement and leave it to the reader to conclude that it implies the statement of the Key Lemma.

$$\text{If } t-1 \leq (\delta^\Delta - \Delta\epsilon)L, \text{ then } \|H \to G\| > \left((\delta^\Delta - \Delta\epsilon)L - (t-1)\right)^h.$$

Define $V(H) = \{v_1, \ldots, v_h\}$ and assign each v_i to a cluster according to some embedding of H in $R(t)$. We will choose $v_1, v_2, \ldots, v_h$ in sequence. At stage $i \geq 1$, we denote $C_{i,j}$ to be the subset of the cluster of vertices that are eligible to be chosen for v_j, given that $v_1, \ldots, v_i$ have already been chosen. Hence, $C_{0,j} \supseteq C_{1,j} \supseteq \cdots \supseteq C_{j-1,j}$.

At stage $i \geq 1$, we do two things:

(1) Choose $v_i \in C_{i-1,i}$ such that $\deg_{C_{i-1,j}}(v_i) \geq \delta |C_{i-1,j}|$ for all $j > i$ with $v_j \sim_H v_i$.

(2) Update $C_{i,j} \overset{\text{def}}{=} C_{i-1,j} \cap N(v_i)$ for all $j > i$ with $v_j \sim_H v_i$.

For $j > i$, let $d_{i,j} = |\{k \in \{1, \ldots, i\} : v_k \sim_H v_j\}|$. So, $|C_{i,j}| \geq \delta^{d_{i,j}} L$. Consequently, $|C_{i,j}| \geq \delta^{d_{i,j}} L \geq \delta^\Delta L > \epsilon L$. So, throughout this process, every C set is larger than ϵL. To see that the process completes with an embedding of H, v_i can be chosen because of the ϵ-regularity of the clusters. If $S \subset C_{i-1,i}$ is the set of vertices with fewer than $\delta |C_{i-1,j}|$ neighbors in $C_{i-1,j}$, then the density is less than $\delta = d - \epsilon$, implying $|S| < \epsilon L$. There are at most Δ such pairs for any i.

As a result, there are at most $\Delta\epsilon L + (t-1)$ vertices in $C_{i-1,i}$ that v_i cannot be assigned to. The $t-1$ accounts for the fact that v_i cannot be any of $v_1, \ldots, v_{i-1}$, which might have come from the same cluster. Therefore, we have $\left(\delta^\Delta - \Delta\epsilon\right) L - (t-1)$ choices for v_i. Consequently, there are $\left((\delta^\Delta - \Delta\epsilon)L - (t-1)\right)^h$ embeddings. $\square$

One thing that makes Theorem 8.2.3 so striking is that, trivially, $\|H \to G_n\| \leq n^h$, but there is a G_n with minimum degree $\left(1 - \frac{1}{p-1} - o(1)\right)\frac{n^2}{2}$, namely the Turán graph, that has no copies of H. But if the minimum degree is increased slightly, the number of embeddings is within a constant of the maximum possible.

Another striking observation, the details of which we leave to the reader, is in some ways, the contrapositive of Theorem 8.2.3. It says that we can choose a small set of edges to cover all copies of H, even if the number of such copies is proportional to n^h.

Theorem 8.2.5 (Covering copies of H by edges) *For every $\beta > 0$ and H a graph on h vertices, there exists $\gamma = \gamma(\beta, H) > 0$ such that if $\|H \to G_n\| \leq \gamma n^h$, then by deleting at most βn^2 edges, G_n can be made H-free.*

Exercises.

(1) Prove the degree form of the Regularity Lemma from the original statement. **Hint:** Apply Theorem 6.3.1 with parameters $\epsilon = \epsilon_{\mathrm{DF}}^2$ and $m = \lceil \epsilon_{\mathrm{DF}} \rceil$. Follow the proof of Theorem 8.1.1.

(2) Prove Theorem 8.2.3 using the Key Lemma, Lemma 8.2.4. **Hint:** Follow the proof of Theorem 8.1.1.

8.3 Blow-up lemma

8.3.1 Alon-Yuster

8.3.2 Embedding theorems

8.3.3 Zhao's theorem on bipartite tiling

8.3.4 Tripartite version of Hajnal-Szemerédi

Chapter 9

Induced Subgraph Applications of the Regularity Lemma

9.6 Expander graphs

Bibliography

[AK97] Noga Alon and Michael Krivelevich, *The concentration of the chromatic number of random graphs*, Combinatorica **17** (1997), no. 3, 303–313. MR MR1606020 (98m:05173)

[AKS80] Miklós Ajtai, János Komlós, and Endre Szemerédi, *A note on Ramsey numbers*, J. Combin. Theory Ser. A **29** (1980), no. 3, 354–360. MR MR600598 (82a:05064)

[AKS81] ______, *A dense infinite Sidon sequence*, European J. Combin. **2** (1981), no. 1, 1–11. MR MR611925 (83f:10056)

[AS00] Noga Alon and Joel Spencer, *The probabilistic method*, second ed., Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience [John Wiley & Sons], New York, 2000, With an appendix on the life and work of Paul Erdős. MR MR1885388 (2003f:60003)

[Ber08] Felix Bernstein, *Zur theorie der trigonometrische reihen*, Leipz. Ber **60** (1908), 325–328.

[BF92] László Babai and Péter Frankl, *Linear algebra methods in combinatorics with applications to geometry and computer science*, University of Chicago Department of Computer Science, Chicago, IL, 1992, (Preliminary Version 2).

[Bir46] Garrett Birkhoff, *Three observations on linear algebra*, Univ. Nac. Tucumán. Revista A. **5** (1946), 147–151. MR MR0020547 (8,561a)

[Bol88] Béla Bollobás, *The chromatic number of random graphs*, Combinatorica **8** (1988), no. 1, 49–55. MR MR951992 (89i:05244)

[Bol98] ______, *Modern graph theory*, Graduate Texts in Mathematics, vol. 184, Springer-Verlag, New York, 1998. MR MR1633290 (99h:05001)

[Bol01] ______, *Random graphs*, second ed., Cambridge Studies in Advanced Mathematics, vol. 73, Cambridge University Press, Cambridge, 2001. MR MR1864966 (2002j:05132)

[Bou59] V.Y. Bouniakowsky, *Sur quelques inegalités concernant les intégrales aux différences finies*, Mem. Acad. Sci. St. Petersbourg **I** (1859), no. 9, 1–18.

[CAH63] Keresztély Corradi and A. András Hajnal, *On the maximal number of independent circuits in a graph*, Acta Math. Acad. Sci. Hungar. **14** (1963), 423–439. MR MR0200185 (34 #84)

[Cam99] Peter J. Cameron, *Permutation groups*, London Mathematical Society Student Texts, vol. 45, Cambridge University Press, Cambridge, 1999. MR MR1721031 (2001c:20008)

[Cau21] Augustin-Louis Cauchy, *Oeuvres 2*, vol. III, 1821.

[CGFS86] Fan R.K. Chung, Ronald L. Graham, Péter Frankl, and James B. Shearer, *Some intersection theorems for ordered sets and graphs*, J. Combin. Theory Ser. A **43** (1986), no. 1, 23–37. MR MR859293 (87k:05002)

[Dil50] Robert P. Dilworth, *A decomposition theorem for partially ordered sets*, Ann. of Math. (2) **51** (1950), 161–166. MR MR0032578 (11,309f)

[Dir52] Gabriel A. Dirac, *Some theorems on abstract graphs*, Proc. London Math. Soc. (3) **2** (1952), 69–81. MR MR0047308 (13,856e)

[EFS56] Peter Elias, Amiel Feinstein, and Claude E. Shannon, *A note on the maximum flow through a network*, IRE Transactions on Information Theory **2** (1956), no. 4, 117–119.

[Ege31] Eugene (Jenő) Egerváry, *On combinatorial properties of matrices*, Mat. Lapok **38** (1931), 16–28, (Hungarian with German summary).

[EH89] Pál Erdős and András Hajnal, *Ramsey-type theorems*, Discrete Appl. Math. **25** (1989), no. 1-2, 37–52, Combinatorics and complexity (Chicago, IL, 1987). MR MR1031262 (90m:05091)

[EHR65] Pál Erdős, András Hajnal, and Richard Rado, *Partition relations for cardinal numbers*, Acta Math. Acad. Sci. Hungar. **16** (1965), 93–196. MR MR0202613 (34 #2475)

[EL75] Pál Erdős and László Lovász, *Problems and results on* 3*-chromatic hypergraphs and some related questions*, Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vol. II, North-Holland, Amsterdam, 1975, pp. 609–627. Colloq. Math. Soc. János Bolyai, Vol. 10. MR MR0382050 (52 #2938)

[ER52] Pál Erdös and Richard Rado, *Combinatorial theorems on classifications of subsets of a given set*, Proc. London Math. Soc. (3) **2** (1952), 417–439. MR MR0065615 (16,455d)

[ER60] Pál Erdős and Alfréd Rényi, *On the evolution of random graphs*, Magyar Tud. Akad. Mat. Kutató Int. Közl. **5** (1960), 17–61. MR MR0125031 (23 #A2338)

[Erd47] Pál Erdös, *Some remarks on the theory of graphs*, Bull. Amer. Math. Soc. **53** (1947), 292–294. MR MR0019911 (8,479d)

[Erd59] Pál Erdős, *Graph theory and probability*, Canad. J. Math. **11** (1959), 34–38. MR MR0102081 (21 #876)

[Erd61] ______, *Graph theory and probability. II*, Canad. J. Math. **13** (1961), 346–352. MR MR0120168 (22 #10925)

[Erd70] ______, *On the graph theorem of Turán*, Mat. Lapok **21** (1970), 249–251 (1971). MR MR0307975 (46 #7090)

[ES46] Pál Erdős and Arthur H. Stone, *On the structure of linear graphs*, Bull. Amer. Math. Soc. **52** (1946), 1087–1091. MR MR0018807 (8,333b)

[ES66] Pál Erdős and Miklós Simonovits, *A limit theorem in graph theory*, Studia Sci. Math. Hungar **1** (1966), 51–57. MR MR0205876 (34 #5702)

[ES72] Pál Erdős and Endre Szemerédi, *On a Ramsey type theorem*, Period. Math. Hungar. **2** (1972), 295–299, Collection of articles dedicated to the memory of Alfréd Rényi, I. MR MR0325446 (48 #3793)

[FF56] Lester R. Ford Jr. and Delbert R. Fulkerson, *Maximal flow through a network*, Canad. J. Math. **8** (1956), 399–404. MR MR0079251 (18,56h)

[Fro17] Georg Frobenius, *Über zerlegbare Determinanten*, Sitzungsber. König. Preuss. Adad. Wiss. **XVIII** (1917), 274–277.

[GT04] David Galvin and Prasad Tetali, *On weighted graph homomorphisms*, Graphs, morphisms and statistical physics, DIMACS Ser. Discrete Math. Theoret. Comput. Sci., vol. 63, Amer. Math. Soc., Providence, RI, 2004, pp. 97–104. MR 2056231 (2005c:05015)

[GY68] Jack E. Graver and James Yackel, *Some graph theoretic results associated with Ramsey's theorem*, J. Combinatorial Theory **4** (1968), 125–175. MR MR0225685 (37 #1278)

[Haf03] Paul R. Hafner, *The Hoffman-Singleton graph and its automorphisms*, J. Algebraic Combin. **18** (2003), no. 1, 7–12. MR MR2002216 (2004f:05078)

[Hal35] Philip Hall, *On representation of subsets*, J. London Math. Soc. **10** (1935), 26–30.

[HHMS10] H. Tracy Hall, Leslie Hogben, Ryan Martin, and Bryan Shader, *Expected values of parameters associated with the minimum rank of a graph*, Linear Algebra Appl. (2010), 17pp., to appear.

[HHPa] Hatos-Hall Productions, *The Monty Hall problem*, `http://www.letsmakeadeal.com/problem.htm`, accessed 18 February, 2008.

[HHPb] ______, *The official Let's Make A Deal website*, `http://www.letsmakeadeal.com/`, accessed 18 February, 2008.

[HS60] Alan J. Hoffman and Robert R. Singleton, *On Moore graphs with diameters* 2 *and* 3, IBM J. Res. Develop. **4** (1960), 497–504. MR MR0140437 (25 #3857)

[HS70] András Hajnal and Endre Szemerédi, *Proof of a conjecture of P. Erdős*, Combinatorial theory and its applications, II (Proc. Colloq., Balatonfüred, 1969), North-Holland, Amsterdam, 1970, pp. 601–623. MR MR0297607 (45 #6661)

[HW73] Carl Hierholzer and Christian Wiener, *Ueber die Möglichkeit, einen Linenzug ohne Wiederholung und ohne Unterbrechung zu umfahren*, Math. Ann. **6** (1873), no. 1, 30–32. MR MR1509807

[JŁR00] Svante Janson, Tomasz Łuczak, and Andrej Ruciński, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000. MR MR1782847 (2001k:05180)

[Kah01] Jeff Kahn, *An entropy approach to the hard-core model on bipartite graphs*, Combin. Probab. Comput. **10** (2001), no. 3, 219–237. MR 1841642 (2003a:05111)

[Kim95] Jeong Han Kim, *The Ramsey number* $R(3,t)$ *has order of magnitude* $t^2/\log t$, Random Structures Algorithms **7** (1995), no. 3, 173–207. MR MR1369063 (96m:05140)

[KK08] Hal A. Kierstead and Alexandr V. Kostochka, *A short proof of the Hajnal-Szemerédi theorem on equitable colouring*, Combin. Probab. Comput. **17** (2008), no. 2, 265–270. MR MR2396352 (2009a:05071)

[Kön31] Dénes König, *Graphen und Matrizen*, Mat. Lapok **38** (1931), 116–119.

[KS96] János Komlós and Miklós Simonovits, *Szemerédi's regularity lemma and its applications in graph theory*, Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993), Bolyai Soc. Math. Stud., vol. 2, János Bolyai Math. Soc., Budapest, 1996, pp. 295–352. MR MR1395865 (97d:05172)

[KSSS02] János Komlós, Ali Shokoufandeh, Miklós Simonovits, and Endre Szemerédi, *The regularity lemma and its applications in graph theory*, Theoretical aspects of computer science (Tehran, 2000), Lecture Notes in Comput. Sci., vol. 2292, Springer, Berlin, 2002, pp. 84–112. MR MR1966181 (2004d:05106)

[Lef87] Hanno Lefmann, *A note on Ramsey numbers*, Studia Sci. Math. Hungar. **22** (1987), no. 1-4, 445–446. MR MR932230 (89d:05132)

[Lov75] László Lovász, *Three short proofs in graph theory*, J. Combinatorial Theory Ser. B **19** (1975), no. 3, 269–271. MR MR0396344 (53 #211)

[LR93] Hanno Lefmann and Vojtěch Rödl, *On canonical Ramsey numbers for coloring three-element sets*, Finite and infinite combinatorics in sets and logic (Banff, AB, 1991), NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., vol. 411, Kluwer Acad. Publ., Dordrecht, 1993, pp. 237–247. MR MR1261209 (94k:05211)

[LR95] ______, *On Erdős-Rado numbers*, Combinatorica **15** (1995), no. 1, 85–104. MR MR1325273 (96a:05003)

[Łuc91] Tomasz Łuczak, *A note on the sharp concentration of the chromatic number of random graphs*, Combinatorica **11** (1991), no. 3, 295–297. MR MR1122014 (92k:05118)

[McD89] Colin McDiarmid, *On the method of bounded differences*, Surveys in combinatorics, 1989 (Norwich, 1989), London Math. Soc. Lecture Note Ser., vol. 141, Cambridge Univ. Press, Cambridge, 1989, pp. 148–188. MR MR1036755 (91e:05077)

[Men27] Karl Menger, *Zur allgemeinen kurventheorie*, Fund. Math. **10** (1927), 95–115.

[Mos04] Ivan Moscovich, *The Monty Hall problem & other puzzles*, Sterling, 2004.

[MŠ] Martin Mačaj and Jozef Širáň, *Search for properties of the missing Moore graph*, Linear Algebra and its Applications, To appear.

[Ore60] Oystein Ore, *Note on Hamilton circuits*, Amer. Math. Monthly **67** (1960), 55. MR MR0118683 (22 #9454)

[PA95] János Pach and Pankaj K. Agarwal, *Combinatorial geometry*, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons Inc., New York, 1995, A Wiley-Interscience Publication. MR MR1354145 (96j:52001)

[Rad94] Stanisław P. Radziszowski, *Small Ramsey numbers*, Electron. J. Combin. **1** (1994), Dynamic Survey 1, 30 pp. (electronic), accessed 28 Jan 2010. MR MR1670625 (99k:05117)

[Rad01] Jaikumar Radhakrishnan, *Entropy and counting*, IIT Kharagpur, Golden Jubilee Volume on Computational Mathematics, Modelling and Algorithms (J.C. Mishra, ed.), Narosa Publishers, New Delhi, 2001.

[Ram30] Frank P. Ramsey, *On a problem of formal logic*, Proc. Lond. Math. Soc. **30** (1930), 264–286.

[Sch88] H. A. Schwarz, *Ueber ein Flachen kleinsten Flacheninhalts betreffendes Problem der Variationsrechnung*, Acta Societatis scientarum Fennicae **XV** (1888), 316–362.

[She83] James B. Shearer, *A note on the independence number of triangle-free graphs*, Discrete Math. **46** (1983), no. 1, 83–87. MR MR708165 (85b:05158)

[She91] ______, *A note on the independence number of triangle-free graphs. II*, J. Combin. Theory Ser. B **53** (1991), no. 2, 300–307. MR MR1129557 (92k:05073)

[Sin68] Robert R. Singleton, *There is no irregular Moore graph*, Amer. Math. Monthly **75** (1968), 42–43. MR MR0225679 (37 #1272)

[Spe78] Joel Spencer, *Asymptotic lower bounds for Ramsey functions*, Discrete Math. **20** (1977/78), no. 1, 69–76. MR MR0491337 (58 #10600)

[SS87] Eli Shamir and Joel Spencer, *Sharp concentration of the chromatic number on random graphs $G_{n,p}$*, Combinatorica **7** (1987), no. 1, 121–129. MR MR905159 (88i:05164)

[Sze75] Endre Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245, Collection of articles in memory of Juriĭ Vladimirovič Linnik. MR MR0369312 (51 #5547)

[Sze78] ______, *Regular partitions of graphs*, Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976), Colloq. Internat. CNRS, vol. 260, CNRS, Paris, 1978, pp. 399–401. MR MR540024 (81i:05095)

[Tur41] Pál Turán, *Eine Extremalaufgabe aus der Graphentheorie*, Mat. Fiz. Lapok **48** (1941), 436–452. MR MR0018405 (8,284j)

[Tut47] William T. Tutte, *The factorization of linear graphs*, J. London Math. Soc. **22** (1947), 107–111. MR MR0023048 (9,297d)

[vdW27] Bartel L. van der Waerden, *Beweis einer baudetschen vermutung*, Nieuw Arch. Wiskunde **15** (1927), 212–216.

[vN53] John von Neumann, *A certain zero-sum two-person game equivalent to the optimal assignment problem*, Contributions to the theory of games, vol. 2, Annals of Mathematics Studies, no. 28, Princeton University Press, Princeton, N. J., 1953, pp. 5–12. MR MR0054920 (14,998i)

[vS97] Marilyn vos Savant, *The power of logical thinking*, St. Martin's Griffin, New York, 1997.

[Vu10] Van Vu, *Chernoff bound*, electronic, February 2010, accessed 13 February 2010.

[Wei] Eric W. Weisstein, *Stirling's approximation*, `http://mathworld.wolfram.com/StirlingsApproximation.html`, accessed 27 Jan 2008.

[Wil86] Robin J. Wilson, *An Eulerian trail through Königsberg*, J. Graph Theory **10** (1986), no. 3, 265–275. MR MR856115 (88a:01015)

[Zha] Yufei Zhao, *The number of independent sets in a regular graph*, preprint.