MA 5109: Topics in Graph Theory

Amit Rajaraman

Last updated January 14, 2023

Contents

§0. Notation

We use $[n]$ to represent $\{1, 2, \ldots, n\}$.

For integers a and b, $[a, b]$ means $\{a, a+1, \ldots, b\}$.

A graph G_n is a graph with n vertices.

Given a graph G, $e(G)$ is the number of edges G has.

For a vertex v, denote by $N(v)$ or $\Gamma(v)$ the set of *neighbours* of v – all the vertices that have an edge to v.

For a vertex v, denote by $d_G(v) = |\Gamma(v)|$ the *degree* of v – the number of edges incident on it. If the graph G is clear from context, we write simply $d(V)$.

For $v \in V$ and $K \subseteq V$, $d(v, K)$ is the number of edges

$$\left|\{u \in K : uv \in E\}\right|$$

from v into K.

For vertices v, w, write $d(v, w) = |\Gamma(v) \cap \Gamma(w)|$.
Given a graph $G = (V, E)$, denote by $\delta(G)$ and $\Delta(G)$ the minimum and maximum degree in G respectively. That is,

$$\delta(G) = \min_{v \in V} d(v) \text{ and } \Delta(G) = \max_{v \in V} d(v).$$

We shorten "there exists n_0 such that for all $n > n_0$" to "for all n sufficiently large", which in turn is shortened to "for all $n \gg 0$".

We say that something occurs "with high probability" if the probability of the complement occurring converges to 0 as the input size grows to ∞.

§1. Introduction

1.1. Basic Definitions

Definition 1.1. A (simple undirected) **graph** G is an ordered pair (V, E) where V is a finite set called the *vertex set* and E, called the *edge set*, is a subset of $\binom{V}{2}$, where $\binom{S}{k}$ represents the set of all k-element subsets of S.

We typically represent graphs pictorially, showing vertices as dots and edges as arcs joining the vertices present in the corresponding subset.
A few important graphs are:

- the *null graph* with vertex set V, where $E = \varnothing$.
- the *complete graph* K_n, where $V = [n]$ and $E = \binom{[n]}{2}$.
- the *complete bipartite graph* $K_{m,n}$, where $V = A \cup B$ with $|A| = m$, $|B| = n$, and A, B are disjoint, and $E = \{\{a, b\} : a \in A, b \in B\}$.
- the *path graph* P_{n+1} of length n, where $V = [n+1]$ and $E = \{\{m, m+1\} : m \in [n]\}$.
- the *cycle* of length n, where $V = [n]$ and $E = \{\{l, m\} : l, m \in [n], (m - l) \equiv 1 \pmod{n}\}$.

Now, consider the graph G with vertex set $[4]$ and edge set $\{\{1, 3\}, \{3, 2\}, \{2, 4\}\}$. This graph appears to be the same as the path graph of length 3, but how do we make this correspondence more concrete?
Relabeling vertices doesn't create a "new" graph.

Definition 1.2 (Graph Isomorphism)**.** Two graphs $G = (V, E)$ and $G' = (V', E')$ are said to be **isomorphic** and we write $G \simeq G'$ if there exists a bijection $f : V \to V'$ such that there is an edge between two vertices u and v in G if and only if there is an edge between $f(u)$ and $f(v)$ in G'.

If two graphs are isomorphic, they are identical for our purposes (we only care about graphs up to isomorphism).
We now give a few more definitions that are useful.

Definition 1.3 (Subgraph)**.** Given a graph $G = (V, E)$, a **subgraph** $H = (V', E')$ is a graph such that $V' \subseteq V$ and $E' \subseteq E$. Given $V' \subseteq V$, the subgraph *induced* by V' on G is that with vertex set V' and edge set $\binom{V'}{2} \cap E$.

Definition 1.4 (r-partite Graph)**.** A graph $G = (V, E)$ is said to be r**-partite** if there exists a partition $V_1, V_2, \ldots, V_r$ of V such that for any edge $e = uv \in E$, u and v are in distinct V_i. That is, there are no edges within any of the V_i.
In particular, a 2-partite graph is said to be **bipartite**.

Definition 1.5 (Independent Set)**.** Given a graph $G = (V, E)$, $I \subseteq V$ is said to be **independent** if no two vertices of I are adjacent (the subgraph induced by I is null).
$\alpha(G)$, the *independence number* of G, denotes the size of the largest independent set in G.

Definition 1.6 (Clique)**.** Given a graph $G = (V, E)$, $K \subseteq V$ is said to be a **clique** if any two vertices of K are adjacent (the subgraph induced by I is complete). $\omega(G)$, the *clique number* of G, denotes the size of the largest clique in G.

Definition 1.7 (Complement Graph)**.** Given a graph $G = (V, E)$, the **complement graph** of G is $\bar{G} = (V, \binom{V}{2} \setminus E)$.

Observe that $S \subseteq V$ is independent in G if and only if S is a clique in $\bar{G}$. In particular, $\alpha(G) = \omega(\bar{G})$.

Definition 1.8 (Connectedness)**.** A graph G is said to be **connected** if for any pair of vertices u, v, there is a sequence $u = v_0, v_1, \ldots, v_r = v$ for some r such that $v_{i-1}v_i$ is an edge for each $i \in [r]$.

Definition 1.9 (Girth)**.** The **girth** of a graph G is the smallest k (> 2) for which C_k is isomorphic to a subgraph of G.

If G has no cycles, it is said to have infinite girth.

1.2. K_{r+1}-free graphs

Extremal graph theory is motivated by the following simple problem:

> At most how many edges can a graph G_n have if it contains no triangles?

More precisely, what is

$$\max_{\substack{\text{no subgraph of } G_n \\ \text{is isomorphic to } K_3}} e(G_n)?$$

Clearly, this number is well-defined since a graph on n vertices cannot have more than $\binom{n}{2}$ vertices.
A simple observation is that any complete bipartite graph has no triangles: if there were a triangle, then two vertices would be in the same "part", which contradicts the existence of edges only between the two parts.
As a consequence, for any $1 \leq m \leq n$, it is possible to construct $m \times (n - m)$ edges (with this bound being attained for $K_{m,n-m}$). In particular, it is possible to construct a graph with $\lfloor n^2/4 \rfloor$ edges.

Theorem 1.1 (Mantel's Theorem)**.** If G_n has no triangle, then

$$e(G_n) \leq \left\lfloor \frac{n^2}{4} \right\rfloor.$$

Further, equality is attained iff $G_n \simeq K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$.

Proof. Suppose G_n has no triangles. Saying that G_n has no triangles is equivalent to saying that for distinct adjacent u, v, $\Gamma(u) \cap \Gamma(v) = \varnothing$.

So, $d(u) + d(v) \leq n$. Therefore,

$$\begin{aligned} ne(G_n) &\overset{(1)}{\geq} \sum_{uv \in E} d(u) + d(v) \\ &= \sum_{uv \in E} |\Gamma(u) \cup \Gamma(v)| \\ &= \big|(e, w) : e = uv \in E, w \in \Gamma(u) \cup \Gamma(v)\big| \\ &= \sum_{u \in V} |\{(e, w) : w \in \Gamma(u), e = uv \in E\}| \\ &= \sum_{u \in V} |\{(v, w) : v, w \in \Gamma(u)\}| \\ &\overset{(2)}{=} \sum_{u \in V} d(u)^2 \\ &\overset{(3)}{\geq} \frac{1}{n} \left(\sum_{u \in V} d(u) \right)^2 \\ &\overset{(4)}{=} \frac{4e(G_n)^2}{n}, \end{aligned}$$

where (2) follows from the changing the main thing being summed over to u, the "middle" vertex in the L-like structure, (3) follows from the Cauchy-Schwarz inequality, and (4) follows from the handshaking lemma.

What happens when equality is attained? Let us look at the case where n is even.
(1) is only tight when $d(u) + d(v) = n$ for all edges uv and (3) is only tight when $d(u)$ is a constant (independent of u). This implies that $d(u) = \frac{n}{2}$ for every $u \in V$. Now, if uv is an edge, $\Gamma(u) \cap \Gamma(v) = \varnothing$ implies that $\Gamma(u) \cup \Gamma(v) = V$, and so $G_n = K_{\frac{n}{2},\frac{n}{2}}$.
The case where n is odd is analyzed similarly, with slight nuances in (3) since exact equality is not attained. ■

While the above is one of the early results in extremal graph theory, the subject was only really born due to Turán in the following result.

Theorem 1.2 (Turán's Thoerem)**.** If G_n has no K_{r+1} $(r \geq 2)$, then $e(G_n) \leq t_r(n)$, with equality attained iff $G \simeq T_r(n)$.

The version for $r = 2$ is just a triangle-free graph and is the same as Mantel's Theorem. In the proof of this, we split the vertex set into two parts and dumped all the edges between these parts.
If we want to avoid K_4 $(r = 3)$, then perhaps we could split the vertex set into three parts and dump all the edges between these parts.
In general, we want to partition V of size n into r "almost equal" parts and set only those edges between vertices in distinct parts – such a graph is known as the **Turán graph** $T_r(n)$ and the number of edges $e(T_r(n))$ is the **Turán number** $t_r(n)$.
In particular, when $r \mid n$,

$$t_r(n) = \binom{r}{2} \left(\frac{n}{r}\right)^2 = \frac{n^2}{2}\left(1 - \frac{1}{r}\right).$$

Here, we give three proofs of Turán's Theorem.

Proof of Turán's Theorem. We perform strong induction on $n + r$. We have already proved the result for $r = 2$.
Suppose $e(G_n) \geq t_r(n)$ and G_n is K_{r+1}-free, where $r > 2$. We wish to prove that $G \simeq T_r(n)$.
Since $t_r(n) > t_{r-1}(n)$[1], the inductive hypothesis implies that G has a copy $K \subseteq V$ of K_r. Observe that for $v \notin K$,

[1] To show this, use the fact that any $(r-1)$-partite graph can be thought of as r-partite graph with one of the pieces being empty, and then that the Turán graph has the most edges among r-partite graphs.

$d(v, K) \leq r - 1$ – otherwise, there would be a copy of K_{r+1} in G.
As a result, $e(V \setminus K, K) \leq (r-1)(n-r)$. By the induction hypothesis, $e(V \setminus K, V \setminus K) \leq t_r(n-r)$. Therefore,

$$t_r(n) \leq e(G_n) \leq t_r(n-r) + (r-1)(n-r) + \binom{r}{2}.$$

However, as can be checked manually, $t_r(n-r) + (r-1)(n-r) + \binom{r}{2} = t_r(n)$!
It follows that equality holds everywhere – $e(G_n) = t_r(n)$, $e(V \setminus K) = t_r(n-r)$, and $d(v, K) = r-1$ for all $v \in V \setminus K$.
This graph is then isomorphic to $T_r(n)$ – for each $v \in V \setminus K$, we can put the vertex in K that is not adjacent to v in the same bucket as v. Then, the only edges are those between distinct buckets (Why?), so $G_n \simeq T_r(n)$. ■

Erdős' Proof of Turán's Theorem. Erdős proves a slightly more general claim: given a K_{r+1}-free graph G_n, there exists an r-partite graph H on V such that $d_G(v) \leq d_H(v)$ for all $v \in V$.
It is then a simple task to check that among the r-partite graphs on n vertices, the Turán graph $T_r(n)$ has the most edges.

To prove our claim, we perform induction on r.
The claim is trivial for the base case $r = 1$.
Now, suppose the claim holds for values less than r. Let $v_0 \in V$ such that $d_G(v_0) = \max_{v \in V} d_G(v)$ (the vertex of maximum degree in G) and $W = \Gamma(z)$. Since G is K_{r+1}-free, W is K_r-free. Inductively, there is an $(r-1)$-partite graph H' on W such that for all $v \in W$, $d_{H'}(v) \geq d_W(v)$.
Let $U = V \setminus W$. For each $u \in U$, remove all its edges in G and set its new neighbour set as W.
Our desired graph H is that with these edges along with those in H' and the edges from v_0 to W. That is, the rth part is $U \cup \{v_0\}$ and the remaining $(r-1)$ parts are those formed by H'. The graph is clearly r-partite by definition. What about the degree inequality?

- $d_G(v_0) = d_H(v_0)$ trivially.
- For $u \in U$, $d_H(u) = d_G(v_0) \geq d_G(u)$.
- For $w \in W$,

$$d_H(w) = |U| + 1 + d_{H'}(w) \geq |U| + 1 + d_W(w) \geq d_U(w) + 1 + d_W(w) = d_G(w).$$ ■

(Why does equality imply that the graph is isomorphic to $T_r(n)$?)

Theorem 1.3 (Turán's Theorem, reformulation)**.** If $d = e(G_n)/n$ is the average degree of the vertices of G_n, then G_n has an independent set of size at least $n/(d+1)$.

Proof. Why is this equivalent to Turán's Theorem?
If G_n has no K_{r+1}, then $\alpha(\bar{G}) \leq r$. If $\bar{G_n}$ has average degree d, the above result would imply that $r \geq n/(d+1)$, that is, $d \leq (n/r) - 1$. The total number of edges in G_n is then

$$\binom{n}{2} - \frac{nd}{2} \leq \binom{n}{2} - \frac{n}{2}\left(\frac{n}{r} - 1\right) = \frac{n^2}{2}\left(1 - \frac{1}{r}\right),$$

which gives Turán's bound!

Let us now get to the proof of the above reformulation. First, consider the following algorithm to come up with *some* independent set in G:

1. Order V to get $\{v_1, \ldots, v_n\}$ and initialize $S = \varnothing$.
2. Add v_1 to S.

3. Having processed vertices v_1 through v_i, add v_{i+1} to S iff there is no vertex in S that is adjacent to v_{i+1}.

It is clear that this always produces an independent set, but the size of the independent set depends on the ordering we choose at the beginning.
For a given ordering σ, denote by $\mathcal{A}(\sigma)$ the independent set produced by the algorithm.
How do we choose a "good" ordering?
Enter the probabilistic method. Define the random variable π to be uniformly random on the set of all permutations of V. Then,

$$\begin{aligned}\mathbf{E}[|\mathcal{A}(\pi)|] &= \mathbf{E}\left[\sum_{v\in V} \mathbb{1}_{v\in\mathcal{A}(\pi)}\right] \\ &= \sum_{v\in V} \mathbf{E}\left[\mathbb{1}_{v\in\mathcal{A}(\pi)}\right] \\ &= \sum_{v\in V} \Pr\left[v \in \mathcal{A}(\pi)\right].\end{aligned}$$

Fix some $v \in V$ and permutation σ. What is the probability that $v \in \mathcal{A}(\sigma)$?
If at the time of processing v for the ordering σ, $\Gamma(v)\cap S \neq \varnothing$, then v is not picked. In particular, if v is the first element of $\Gamma(v)\cup\{v\}$ in the ordering σ, then it is definitely chosen by the algorithm. The probability of this occurring is $\frac{1}{d(v)+1}$.
So,

$$\begin{aligned}\mathbf{E}[|\mathcal{A}(\pi)|] &= \sum_{v\in V} \Pr\left[\mathbb{1}_{v\in\mathcal{A}(\pi)}\right] \\ &\geq \sum_{v\in V} \frac{1}{d(v)+1} \\ &\overset{(*)}{\geq} \frac{n^2}{\sum_{v\in V}(d(v)+1)} = \frac{n}{d+1},\end{aligned}$$

where $(*)$ follows from the AM-HM inequality.
Since the expectation of $|\mathcal{A}(\pi)|$ is at least $n/(d+1)$, there must exist some permutation σ such that $|\mathcal{A}(\sigma)| \geq n/(d+1)$, proving the result. ■

1.3. The Zarankiewicz Problem

Turán's Theorem is the primary result that birthed Extremal Graph Theory. To generalize the problem studied in the previous section, define the following.

Definition 1.10 (Extremal Function)**.** Given a graph H, define the **extremal function**

$$\operatorname{ex}(n; H) = \max_{\substack{\text{no subgraph of } G_n \\ \text{is isomorphic to } H}} e(G_n) \tag{1.1}$$

as the maximum number of edges in a graph on n vertices without H as a subgraph.
More generally, if $\mathcal{F}$ is a family of graphs, define

$$\operatorname{ex}(n; \mathcal{F}) = \max_{\substack{\text{no subgraph of } G_n \\ \text{is isomorphic to a } H\in\mathcal{F}}} e(G_n).$$

In particular, given graphs $H_1, \ldots, H_m$, we denote

$$\operatorname{ex}(n; H_1, \ldots, H_m) = \operatorname{ex}(n; \{H_1, \ldots, H_m\}).$$

With this notation, Turán's Theorem then says that $\mathrm{ex}(n; K_{r+1}) = t_r(n)$, with the corresponding maximum in (1.1) being attained iff $G_n \simeq T_r(n)$.

Definition 1.11 (Zarankiewicz Function). Fix $s, t \in \mathbb{N}$ with $t \geq s \geq 2$ and $m, n \in \mathbb{N}$. The **Zarankiewicz function** $z(m, n; s, t)$ is the maximum number of edges in a bipartite graph $G = (A \sqcup B, E)$ such that

- the two components A and B of the graph are of cardinality m and n respectively[2], and
- there exist no $S \subseteq A$, $T \subseteq B$ with $|S| = s$, $|T| = t$, and all edges between S and T present in E.[3]

For ease of writing, we refer to the above described criterion as the *Zarankiewicz condition*.
That is, we "forbid" the subgraph $K_{s,t}$ with the components of size s and t on the side of A and B respectively.

The **Zarankiewicz problem** asks for a closed form representation of $z(m, n; s, t)$. Failing this, for fixed t, it asks for a tight asymptotic bound on $z(n, n; t, t)$ as n grows large.
Perhaps surprisingly, this problem remains unsolved! (as of the time of writing these notes)

Theorem 1.4 (Kővári-Sós-Turán Theorem). For $t \geq s \geq 2$ and $m, n \in \mathbb{N}$,

$$z(m, n; s, t) \leq (s-1)^{1/t}(n - t + 1)m^{1-1/t} + (t-1)m.$$

Proof. Let G be bipartite with vertex set $A \sqcup B$ and satisfy the Zarankiewicz condition.
Fix $a \in A$. By definition, $|\Gamma(a)| = d(a)$. Now, $\binom{d(a)}{t}$ is the number of t-element subsets of $\Gamma(a)$. We may assume that for all $a \in A$, $d(a) \geq t - 1$. Indeed, otherwise, we may add arbitrary edges to a to make its degree $t - 1$; any vertex from A in a subgraph isomorphic to $K_{s,t}$ must have at least degree t so a cannot be part of it.
We have that

$$\begin{aligned}\sum_{x \in A} \binom{d(x)}{t} &= \left|\{(x, T) : x \in A, T \subseteq \Gamma(x), |T| = t\}\right| \\ &= \sum_{\substack{T \subseteq B \\ |T| = t}} |\{x \in A : T \subseteq \Gamma(x)\}|.\end{aligned}$$

If we fix a T, then the number of such x for that T is at most $s - 1$, due to our assumption. As a result,

$$\sum_{x \in A} \binom{d(x)}{t} \leq \binom{n}{t}(s-1).$$

Now, observe that the function

$$f(x) = \frac{x(x-1)\cdots(x-t+1)}{t!}$$

is convex on $[t-1, \infty)$. Using Jensen's inequality together with our assumption that $d(x) \geq t - 1$ for all $x \in A$,

$$\begin{aligned}\binom{n}{t}(s-1) &\geq \sum_{x \in A} \binom{d(x)}{t} \\ &\geq m \binom{\frac{1}{m}\sum_{x \in A} d(x)}{t} \\ &= m \binom{e(G)/m}{t}. \end{aligned} \tag{1.2}$$

[1]by "components" of the bipartite graph we mean that for any edge uv in the graph, $u \in A$ and $v \in B$ or $u \in B$ and $v \in A$.
[2]we make the tacit assumption that $s \leq m$ and $t \leq n$.

Let $d = e(G)/m$, the average degree of the vertices in A. Simplifying the above expression,

$$\frac{s-1}{m} \geq \frac{d(d-1)\cdots(d-t+1)}{n(n-1)\cdots(n-t+1)} \geq \left(\frac{d-t+1}{n-t+1}\right)^t,$$

where the second inequality follows from the fact that $d \leq n$. Therefore,

$$e(G) \leq m\left(\left(\frac{s-1}{m}\right)^{1/t}(n-t+1)+(t-1)\right) = (s-1)^{1/t}(n-t+1)m^{1-1/t}+(t-1)m,$$

completing the proof. ■

Next, let us look at some consequences of the above bound.

1.3.1. The Zarankiewicz problem and the extremal function for complete bipartite graphs

We get a bound on $\mathrm{ex}(n; K_{s,t})$. We claim that for $n \in \mathbb{N}$ and $s, t \geq 2$,

$$\mathrm{ex}(n; K_{s,t}) \leq \frac{1}{2} z(n,n;s,t). \tag{1.3}$$

Indeed, if $G_n = (V, E)$ has no $K_{s,t}$, make a bipartite graph G' that has vertex set $V \times \{0,1\}$ such that uv is an edge in G iff $\{(u,0),(v,1)\}$ is an edge in G'.
G' satisfies the Zarankiewicz condition. If there do exist $S \subseteq V \times \{0\}$ and $T' \subseteq V \times \{1\}$ such that all S-T edges are in G, then $\pi_1(S) \cap \pi_1(T') = \varnothing$ (otherwise, a vertex would be adjacent to itself in G, which is clearly false). This implies that $K_{s,t} \subseteq G$, which is a contradiction.
Since $e(G') = 2e(G)$, the claim follows.

1.3.2. The case where $s = t = 2$

When $s = t = 2$, we get

$$z(m,n;2,2) \leq (n-1)m^{1/2} + m \text{ and } z(n,n;2,2) \leq (n-1)n^{1/2} + n.$$

Therefore,

$$\mathrm{ex}(n; K_{2,2}) \leq \frac{1}{2}(n + (n-1)\sqrt{n}).$$

Note that $K_{2,2} \simeq C_4$. Therefore, a square-free graph on n vertices has $\mathcal{O}(n^{3/2})$ edges.

In fact, this bound is tight! We give an algebraic construction of a suitable graph with no $K_{2,2}$, known as the *Levi graph for the projective plane*.
Let q be a prime power and consider the 3-dimensional vector space $\mathcal{V} = \mathbb{F}_q^3$ (over $\mathbb{F}_q$). Let $\mathcal{P}$ and $\mathcal{L}$ be the set of all 1- and 2-dimensional subspaces of $\mathcal{V}$ respectively.
Define the graph $G = (\mathcal{P} \sqcup \mathcal{L}, E)$ as follows. For $x \in \mathcal{P}$ and $L \in \mathcal{L}$, let x be adjacent to L in G iff $x \subseteq L$.
We claim that G has no $K_{2,2}$. Suppose otherwise and let $x_1, x_2 \in \mathcal{P}$ and $L_1, L_2 \in \mathcal{L}$ such that the x_i are adjacent to the L_j. If $x_1 = \langle u \rangle$ and $x_2 = \langle v \rangle$, then u and v are linearly independent, which implies that $L_i = \langle u, v \rangle$. This contradicts the fact that the L_i are distinct!
What are the cardinalities of $\mathcal{P}$ and $\mathcal{L}$?

- To get a 1-dimensional subspace, we pick a non-zero u and consider $\langle u \rangle$. In $\mathcal{V}$, there are $q^3 - 1$ non-zero u. We must now divide by $q-1$ to account for the fact that linearly dependent vectors generate the same 1-dimensional subspace. Any non-zero u has precisely $q-1$ non-zero multiples. Therefore,

$$|\mathcal{P}| = \frac{q^3-1}{q-1} = q^2 + q + 1.$$

- It turns out that the number of 2-dimensional subspace is equal to the number of 1-dimensional subspaces (more generally, the number of d-dimensional subspaces is equal to the number of $(n-d)$-dimensional subspaces of $\mathbb{F}_q^n$), so
$$|\mathcal{L}| = q^2 + q + 1.$$

How many edges does G have?
Fix $x = \langle u \rangle \in \mathcal{P}$. We wish to determine how many $L \in \mathcal{L}$ are adjacent to x in G. Such an L can be created by choosing $v \notin x$ and letting $L = \langle u, v \rangle$.
The number of choices of v is $q^3 - q$, but each such subspace is repeated for $q^2 - q$ choices of v since the cardinality of $\langle u, v \rangle$ is q^2. Therefore,
$$d(x) = \frac{q^3 - q}{q^2 - q} = q + 1.$$
The total number of edges is
$$|\mathcal{P}|(q+1) = (q+1)(q^2+q+1) = q^3 + 2q^2 + 2q + 1,$$
which is $\mathcal{O}(|\mathcal{P}|^{3/2})$! Therefore, our $\mathcal{O}(n^{3/2})$ bound is tight.
In fact, the Levi graph is optimal in the case where $n = 2(q^2+q+1)$, as seen in Corollary 1.5.

Corollary 1.5. For $n \in \mathbb{N}$,
$$z(n, n; 2, 2) \leq \frac{n(1 + \sqrt{4n-3})}{2}. \tag{1.4}$$
Consequently,
$$\text{ex}(n; C_4) \leq \frac{1}{4} n (1 + \sqrt{4n-3}). \tag{1.5}$$

Proof. Equation (1.5) clearly follows from Equations (1.3) and (1.4), so it suffices to show the first equation. Equation (1.2) in the proof of the Kővári-Sós-Turán Theorem for the case where $s = t = 2$ just says that
$$\binom{n}{2} \geq n \binom{d}{2},$$
where $d = e(G)/m$. That is, $d^2 - d - (n-1) \leq 0$. Then,
$$d \leq \frac{1 + \sqrt{1 + 4(n-1)}}{2} = \frac{1 + \sqrt{4n-3}}{2},$$
which is exactly the bound we want. This bound is tight in the case where $n = 2(q^2+q+1)$, as seen in the Levi graph. ■

Before we move on to the next section, let us build a tiny bit of intuition for why the construction of the Levi graph works.
The projective plane is chosen to ensure that any two distinct points determine a unique line (which holds even in the non-projective plane setting), and any two distinct lines determine a unique point. This corresponds to the absence of $K_{2,2}$ – if it *was* present as a subgraph, then there would be two lines (points) that determine two points (lines), which cannot happen!

1.3.3. The case where $s = t = 3$

We next look at $\text{ex}(n; K_{3,3})$.

The Kővári-Sós-Turán Theorem applied here gives
$$z(n, n; 3, 3) \leq (2)^{1/3}(n-2)n^{2/3} + 2n,$$
which is $\mathcal{O}(n^{5/3})$.

Similar to the Levi graph, we construct an (algebraic) extremal example. Let p be a prime and fix some $r \in \mathbb{F}_p$. Consider the graph G that has vertex set $\mathcal{V} = \mathbb{F}_p^3$ where (x, y, z) is adjacent to (u, v, w) iff

$$(x-u)^2 + (y-v)^2 + (z-w)^2 = r. \tag{1.6}$$

Before moving on to explaining why this works, let us try to impart some intuition. (1.6) resembles the equation of a sphere in $\mathbb{R}^3$. If we take three spheres of the same radius, the points of intersection of all three must lie on two circles, corresponding to the circles of intersection of two pairs of spheres. Since any two circles meet at atmost two points, the absence of $K_{3,3}$ follows.
It may be shown that even over $\mathbb{F}_p$, two "spheres" intersect on a "circle" and any two circles meet at at most 2 points (Check this!).
So, if we have a $K_{3,3}$ in the described graph, we have three spheres (centered at each of the three points) that intersect at three points, which is not possible.

It remains to count the number of edges in this graph. To do so, let us estimate the degree of $(0, 0, 0)$, since all vertices have the same degree (Why?). That is, we want to determine

$$|\{(x, y, z) \in \mathbb{F}_p^3 : x^2 + y^2 + z^2 = r\}|.$$

Letting z be arbitrary, we want to find

$$N(\xi) = |\{(x, y) \in \mathbb{F}_p^2 : x^2 + y^2 = \xi\}|$$

for any arbitrary $\xi \in \mathbb{F}_p$.

Definition 1.12 (Legendre Symbol)**.** For $a \in \mathbb{F}_p$, define the **Legendre symbol**

$$\left(\frac{a}{p}\right) = \begin{cases} 1, & a \in \mathbb{F}_p^\times \text{ is a square,} \\ -1, & a \in \mathbb{F}_p^\times \text{ is not a square,} \\ 0, & a = 0. \end{cases}$$

With the above, it is not too difficult to see that

$$N(\xi) = \sum_{(a,b):a+b=\xi} \left(1 + \left(\frac{a}{p}\right)\right)\left(1 + \left(\frac{b}{p}\right)\right).$$

Let us now compute the above quantity.

- First of all,
$$\sum_{(a,b):a+b=\xi} 1 = p.$$
- Observe that
$$\sum_{a \in \mathbb{F}_p} \left(\frac{a}{p}\right) = 0.$$
Indeed, the number of squares and non-squares in $\mathbb{F}_p^\times$ is the same. As a result,
$$\sum_{(a,b):a+b=\xi} \left(\frac{a}{p}\right) = \sum_{a \in \mathbb{F}_p} \left(\frac{a}{p}\right) = 0.$$

Therefore,

$$N(\xi) = p + \sum_{(a,b):a+b=\xi} \left(\frac{a}{p}\right)\left(\frac{b}{p}\right). \tag{1.7}$$

Notice that the map $\mathbb{F}_p^\times \to \{-1, 1\}$ given by $x \mapsto \left(\frac{x}{p}\right)$ is a group homomorphism.

Lemma 1.6. If $\xi \neq 0$, $N(\xi) = N(1)$.

Proof. Using (1.7) and letting $a' = a/\xi$ and $b' = b/\xi$,

$$\begin{aligned} N(\xi) - p &= \sum_{a+b=\xi} \left(\frac{a}{p}\right)\left(\frac{b}{p}\right) \\ &= \sum_{a'+b'=1} \left(\frac{a'\xi}{p}\right)\left(\frac{b'\xi}{p}\right) \\ &= \sum_{a'+b'=1} \left(\frac{a'}{p}\right)\left(\frac{b'}{p}\right)\left(\frac{\xi^2}{p}\right) \\ &= \sum_{a'+b'=1} \left(\frac{a'}{p}\right)\left(\frac{b'}{p}\right) = N(1) - p. \end{aligned}$$

■

So, we have

$$\begin{aligned} (p-1)(N(1)-p) &= \sum_{\xi \in \mathbb{F}_p^\times} \sum_{(a,b):a+b=\xi} \left(\frac{a}{p}\right)\left(\frac{b}{p}\right) \\ &= \sum_{\xi \in \mathbb{F}_p^\times} \sum_{a \in \mathbb{F}_p} \left(\frac{a}{p}\right)\left(\frac{\xi - a}{p}\right) \\ &= \sum_{a \in \mathbb{F}_p} \left(\left(\frac{a}{p}\right) \sum_{\xi \in \mathbb{F}_p^\times} \left(\frac{\xi - a}{p}\right)\right) \\ &= \sum_{a \in \mathbb{F}_p} \left(\frac{a}{p}\right)\left(0 - \left(\frac{-a}{p}\right)\right) \\ &= -\left(\frac{-1}{p}\right) \sum_{a \in \mathbb{F}_p} \left(\frac{a^2}{p}\right) && \text{(using the group homomorphism property)} \\ &= -(p-1)\left(\frac{-1}{p}\right) \end{aligned}$$

and therefore, $N(1) = p - \left(\frac{-1}{p}\right)$.

If we choose r to be a non-square, then $r - z^2 \neq 0$ for any $z \in \mathbb{F}_p$ and $N(r - z^2) = p - \left(\frac{-1}{p}\right)$. In this case, the degree of $(0,0,0)$ is $\Theta(p^2)$. The size n of the vertex set $\mathcal{V}$ is p^3, so the number of edges is $\Theta(p^5)$ which is $\Theta(n^{5/3})$ and thus, the bound given by the Kővári-Sós-Turán Theorem is (asymptotically) tight.

For $s \leq t$, the Kővári-Sós-Turán Theorem gives

$$\text{ex}(n; K_{s,t}) < C_{s,t} n^{2-1/s}$$

for some $c_{s,t}$ depending only on s and t. However, it is not known if the above bound is tight for most cases.

Lower bounds of the form

$$\text{ex}(n; K_{s,t}) \geq c_{s,t} n^{2-1/t}$$

are known only if $t > (s-1)!$. See [KRS96] for more details.

1.4. P_{k+1}-free graphs

Next, we study $\text{ex}(n; P_{k+1})$.

First of all, note that K_k is P_{k+1}-free. So, if we split the n vertices up into blocks of k vertices and add all edges within each block, the resulting graph will be P_{k+1}-free as well. That is, G_n is a disjoint union of K_ks (and possibly one K_r for $r < k$). For this particular graph G_n,

$$e(G_n) \leq \frac{(k-1)n}{2}.$$

It turns out that we cannot do better than this.

Theorem 1.7. $\text{ex}(n; P_{k+1}) \leq (k-1)n/2$ with equality if and only if G_n is a disjoint union of K_ks.

The above theorem just says that equality is not attained for connected graphs without P_{k+1}.
To prove this, we prove the following (seemingly) more general statement.

Lemma 1.8. Let $G_n = (V, E)$ be connected and suppose $d(v) \geq k$ for all $v \in V$. If $n \geq 2k$, G_n contains a path of length $2k$. Otherwise, G_n contains an n-cycle.

Proof. We prove the result for the case where $n \geq 2k$ first. Consider the longest path $u = v_1, v_2, \ldots, v_r = v$ in G_n and let $U = \{v_i : i \in [r]\}$. We must show that $r \geq 2k$. Suppose otherwise and let $r < 2k$.
First of all, we must have $\Gamma(u) \subseteq U$ – otherwise, the path can be extended by adding another vertex from $\Gamma(u) \setminus U$. Similarly, $\Gamma(v) \subseteq U$.
Next, v_1 and v_r cannot be adjacent. If they are, then we can obtain a longer path by cycling through and choosing some edge from a vertex in U to one outside of U (such a vertex must exist since the graph is connected and has at least $2k$ vertices). More generally, if there exists i such that v_1, v_{i+1} and v_i, v_r are adjacent, then we arrive at a contradiction (for the same reason).
Let $S = \{i : v_i v_r \in E\}$ and $T = \{i : v_1 v_{i+1} \in E\}$. By the above observation, $S \cap T = \varnothing$. However, by our first observation, $|S| = d(v_r) \geq k$ and $|T| = d(v_1) \geq k$. Therefore, $r \geq |S \cup T| \geq 2k$.
The result for $n < 2k$ is shown using nearly the same proof. ∎

Proof of Theorem 1.7. We perform strong induction on n. We may assume that $n > k$. Suppose G_n has no P_{k+1}. If G_n is *not* connected, then it consists of a disjoint union of connected subgraphs. We may then apply the inductive hypothesis to each of these smaller pieces.
So, let G_n be connected. By Lemma 1.8, there is some vertex v such that $d(v) \leq (k-1)/2$ (otherwise, there must be a path of length k). Additionally, observe that G_n does not have any subgraph isomorphic to K_k – using the connectedness assumption gives a path of length k otherwise.
The graph $G \setminus \{x\}$[4] has no P_{k+1} and further, it has no K_k either. Therefore, $G \setminus \{x\}$ is *not* extremal and $e(G \setminus \{x\}) < (k-1)(n-1)/2$. Therefore,

$$\begin{aligned} e(G) &= d(x) + e(G \setminus x) \\ &< \frac{k-1}{2} + \frac{(k-1)(n-1)}{2} = \frac{(k-1)n}{2}. \end{aligned}$$ ∎

[4]This is the subgraph induced by G on the vertex set $V \setminus \{x\}$

§2. Fundamental Results in Extremal Graph Theory

2.1. The Erdős-Stone-Simonovits Theorem

2.1.1. Motivation

Using the folklore result that a graph is bipartite iff it has no odd cycle, we have

$$\mathrm{ex}(n; C_{2k+1}) \geq \left\lfloor \frac{n^2}{4} \right\rfloor.$$

Remark. It in fact turns out that for $n \gg 0$, $\mathrm{ex}(n; C_{2k+1}) = \lfloor n^2/4 \rfloor$, as we shall see later..
Is there any more general relationship between forbidden subgraphs and r-partite graphs?

Definition 2.1 (Chromatic Number)**.** Given a graph G, its **chromatic number** is given by

$$\chi(G) = \min\{r : G \text{ is } r\text{-partite}\}.$$

Alternatively, we can define the above using the following.

Definition 2.2. Given a graph $G = (V, E)$, an r**-coloring** of G is a function $f : V \to [r]$ such that for any $uv \in E$, $f(u) \neq f(v)$.
The chromatic number of a graph is the least r such that it is r-colorable.

Suppose H is an arbitrary graph such that $\chi(H) = r + 1$. Then, no r-partite graph contains H. As a result,

$$\mathrm{ex}(n; H) > t_r(n).$$

Our earlier observation on $\mathrm{ex}(n; C_{2k+1})$ is then just a consequence of the fact that C_{2k+1} is 3-colorable.
Let us give a more concrete example. Suppose we want to find $\mathrm{ex}(n; \text{Petersen})$, where Petersen is the Petersen graph.
It may be checked that Petersen has chromatic number 3. So,

$$\mathrm{ex}(n; \text{Petersen}) > t_2(n) = \left\lfloor \frac{n^2}{4} \right\rfloor.$$

Can we do better?
This is answered by the Erdős-Stone-Simonovits Theorem.

2.1.2. The result

Theorem 2.1 (Erdős-Stone-Simonovits Theorem, Version 1)**.** Let $r \geq 1$ and $0 < \varepsilon < 1/2$. Then, for $n \gg 0$ and any graph G_n, if

$$\delta(G_n) \geq \left(1 - \frac{1}{r} + \varepsilon\right) n, \tag{2.1}$$

there exist pairwise disjoint subsets $V_1, \ldots, V_{r+1}$ of V such that for each i,

$$|V_i| = t \geq \frac{\varepsilon \log n}{2^{r-1}(r-1)!}$$

and the complete $(r+1)$-partite graph on these subsets is contained in G_n.

Theorem 2.2 (Erdős-Stone-Simonovits Theorem, Version 2)**.** Let $r \geq 1$ and $0 < \varepsilon < 1/2$. Then, for $n \gg 0$ and any graph G_n, if

$$e(G_n) \geq \left(1 - \frac{1}{r} + \varepsilon\right)\binom{n}{2}, \tag{2.2}$$

there exist pairwise disjoint subsets $V_1, \ldots, V_{r+1}$ of V such that for each i,

$$|V_i| = t \geq \frac{\varepsilon \log n}{2^{r+1}(r-1)!}$$

and the complete $(r+1)$-partite graph on these subsets is contained in G_n.

Observe the difference in the exponent of 2 in the denominator of t in the two versions.
For example, if $n \geq e^{32/\varepsilon}$ and Equation (2.2) is satisfied, then $T_3(12)$, and thus Petersen, is a subgraph of G_n.
Therefore, (replacing ε with 2ε)

$$\text{ex}(n; \mathsf{Petersen}) \leq \left(\frac{1}{2} + 2\varepsilon\right)\binom{n}{2} \leq \left(\frac{1}{4} + \varepsilon\right)n^2.$$

for all $\varepsilon > 0$.

Corollary 2.3. If $\chi(H) = r + 1$, then for any $\varepsilon > 0$, for $n \gg 0$,

$$\left(1 - \frac{1}{r}\right)\frac{n^2}{2} \leq \text{ex}(n; H) \leq \left(1 - \frac{1}{r} + \varepsilon\right)n^2 - \mathcal{O}(n).$$

This implies that if $\chi(H) = r + 1$, then for $n \gg 0$

$$\left(1 - \frac{1}{r}\right)\frac{n^2}{2} \leq \text{ex}(n; H) \leq \left(1 - \frac{1}{r} + o(1)\right)\frac{n^2}{2}$$

Proof of Erdős-Stone-Simonovits Theorem, Version 1. We first show the result for $r = 1$. Let G_n be a graph such that $\delta(G_n) \geq \varepsilon n$. We want a "large" t such that $K_{t,t}$ is isomorphic to a subgraph G_n.
Suppose that G_n is $K_{t,t}$-free (we shall fix t later). Now, we have that

$$t\binom{n}{t} \overset{(1)}{>} |\{(v, T) : v \in V, T \subseteq \Gamma(v), |T| = t\}| = \sum_{v \in V}\binom{d(v)}{t} \overset{(2)}{\geq} n\binom{\varepsilon n}{t},$$

where (2) follows from the hypothesis on $\delta(G_n)$ and (1) follows due to our assumption that G_n is $K_{t,t}$-free.
Therefore,

$$n\binom{\varepsilon n}{t} < t\binom{n}{t}.$$

So,

$$\begin{aligned}
1 &< \frac{t}{n} \cdot \frac{n(n-1)\cdot(n-t+1)}{\varepsilon n(\varepsilon n - 1)\cdot(\varepsilon n - t + 1)} \\
&< \frac{t}{n} \cdot \left(\frac{n}{\varepsilon n - t + 1}\right)^t \\
&= \frac{t}{n}\frac{1}{\varepsilon^t}\left(\frac{1}{1 - \frac{t-1}{\varepsilon n}}\right)^t.
\end{aligned}$$

We desire a large t such that the above results in a contradiction. Let us look at the last term. Suppose we want

$$\left(1 - \frac{t-1}{\varepsilon n}\right)^t \to 1.$$

Equivalently,

$$t \log\left(1 - \frac{t-1}{\varepsilon n}\right) \to 0$$
$$t\left(\frac{t-1}{\varepsilon n} + \frac{1}{2}\left(\frac{t-1}{2n}\right)^2 + \frac{1}{3}\left(\frac{t-1}{\varepsilon n}\right)^3 + \cdots\right) \to 0.$$

If $\varepsilon n > t^4$, the above holds (Why?).
That is, if $t < (\varepsilon n)^{1/4}$, then for sufficiently large n, t,

$$\left(1 - \frac{t-1}{\varepsilon n}\right)^{-t} < 2$$

and

$$1 < \frac{t}{n} \cdot \frac{1}{\varepsilon^t}\left(1 - \frac{t-1}{\varepsilon n}\right)^{-t} < \frac{2t}{\varepsilon^t n}.$$

If the above is ≤ 1, we shall arrive at a contradiction. That is, if

$$\log(2t) + t\log\left(\frac{1}{\varepsilon}\right) < \log n,$$

then we arrive at a contradiction. In particular,

$$t = \lceil \varepsilon \log n \rceil$$

is a suitable choice (Why?), completing the proof of the theorem for $r = 1$.

Now, let us prove the general case by performing induction on r. Let G_n be a graph with

$$\delta(G_n) \geq \left(1 - \frac{1}{r} + \varepsilon\right) n.$$

Now,

$$1 - \frac{1}{r} + \varepsilon > 1 - \frac{1}{r-1} + \frac{1}{r(r-1)}.$$

Let $\varepsilon' = 1/r(r-1)$. By induction, there are $V_1', \ldots, V_r'$ such that

$$|V_i'| \geq t' = \frac{\log n}{r(r-1)2^{r-2}(r-2)!} = \frac{\log n}{2^{r-2}r!} \tag{2.3}$$

for each i and the complete r-partite graph on these sets is a subgraph of G_n. Let $K = \bigcup V_i'$.
We obviously have $\varepsilon < 1/r$, since the claim is vacuously true otherwise.
We shall find $V_i \subseteq V_i'$ for each $1 \leq i \leq r$ and some $V_{r+1} \subseteq V \setminus K$ such that the complete bipartite graph on $(V_1, \ldots, V_{r+1})$ is the required subgraph of G_n.
Now, define

$$U = \left\{x \in V \setminus K : d(x, K) \geq \left(1 - \frac{1}{r} + \lambda\right)|K|\right\}$$

for some λ we shall fix later. We shall bound $e(K, V \setminus K)$ in two different ways. From the perspective of K,

$$e(K, V \setminus K) \geq |K|\left(\left(1 - \frac{1}{r} + \varepsilon\right) n - |K|\right).$$

From the perspective of $V \setminus K$,

$$\begin{aligned} e(K, V\setminus K) &= e(V\setminus (K \sqcup U), K) + e(U, K) \\ &\le \left(n - |U| - |K|\right)\left(1 - \frac{1}{r} + \lambda\right)|K| + |U||K|. \end{aligned}$$

Putting the two together,

$$\begin{aligned} |K|\left(\left(1 - \frac{1}{r} + \varepsilon\right)n - |K|\right) &\le \left(n - |U| - |K|\right)\left(1 - \frac{1}{r} + \lambda\right)|K| + |U||K| \\ \left(1 - \frac{1}{r} + \varepsilon\right)n - |K| &\le \left(n - |U| - |K|\right)\left(1 - \frac{1}{r} + \lambda\right) + |U|. \end{aligned}$$

Set $\lambda = \varepsilon/2$. We then have

$$\frac{r}{2}\varepsilon n \le \left(1 - \frac{r\varepsilon}{2}\right)(|U| + |K|).$$

So,

$$|U| \ge \frac{r\varepsilon}{2 - r\varepsilon}n - |K| \overset{(*)}{\ge} \frac{\varepsilon}{1-\varepsilon}n - rt',$$

where $(*)$ follows from the fact that $r\varepsilon/(2 - r\varepsilon)$ is increasing in r and $r \ge 2$.
Since t', and thus rt', is of the order of $\log n$, the first term in the expression dominates for sufficiently large n. So, for $n \gg 0$, $|U| \ge \varepsilon n$.

Now,

$$\begin{aligned} \left(1 - \frac{1}{r} + \frac{\varepsilon}{2}\right)|K| &\ge \left(1 - \frac{1}{r} + \frac{\varepsilon}{2}\right)rt' \\ &\ge (r-1)t' + \frac{\varepsilon r}{2}t'. \end{aligned}$$

This implies that each $u \in U$ has at least $(\varepsilon r/2)t'$ neighbours in each V_i'.
We can now use a pigeonhole argument to choose a subset of U whose vertices are all adjacent to some common set of vertices in each V_i'. To do so, consider

$$|\{(u, W_1, \dots, W_r) : W_i \subseteq V_i', |W_i| = (\varepsilon r/2)t', \text{ and } u \text{ is adjacent to all the vertices of each } W_i\}|.$$

By our earlier observation, this must be at least $|U| \ge \varepsilon n$.
On the other hand, it is at most the number of ways of choosing the W_i, which is $\binom{t'}{(\varepsilon r/2)t'}^r$.
In particular, using a pigeonhole argument, there exist $V_1, \dots, V_r$ and a $V_{r+1} \subseteq U$ such that $V_i \subseteq V_i'$ for each $1 \le i \le r$ and for all $u \in V_{r+1}$, $(u, V_1, \dots, V_r)$ is in the set whose cardinality we just considered, and $|V_{r+1}| \ge \varepsilon n / \binom{t'}{(\varepsilon r/2)t'}^r$. Let us now bound this expression.

$$\begin{aligned} |V_{r+1}| &\ge \frac{\varepsilon n}{\binom{t'}{(\varepsilon r/2)t'}^r} \\ &\ge \frac{\varepsilon n}{(2e/\varepsilon r)^{t'\varepsilon r^2/2}} && \text{(see here for the bound used)} \\ &\ge \varepsilon n \cdot \left(\frac{\varepsilon}{e}\right)^{t'\varepsilon r^2/2} && \text{(since } r \ge 2,\ 2/r \le 1). \end{aligned}$$

Setting

$$t = \frac{\varepsilon \log n}{2^{r-1}(r-1)!},$$

we see that $t = \varepsilon r t'/2$.
Keeping in mind that the bound we want is $|V_{r+1}| \geq t$,

$$\begin{aligned}
\log\left(n\left(\frac{\varepsilon}{e}\right)^{t'\varepsilon r^2/2}\right) &\geq \log n + \frac{\varepsilon r^2 t'}{2}\log\left(\frac{\varepsilon}{e}\right) \\
&\geq t'\left(2^{r-2}r! - \log\left(\frac{e}{\varepsilon}\right)\cdot\frac{\varepsilon r^2}{2}\right) && \text{(using the expression for } t' \text{ in (2.3))} \\
&\geq t'\left(2^{r-2}r! - \log(2e)\cdot\frac{r^2}{4}\right) && (-\varepsilon\log(e/\varepsilon) \text{ is decreasing in } \varepsilon) \\
&\geq \log(rt'/2).
\end{aligned}$$

Therefore, $V_{r+1} \geq t$. Since $|V_i| = \varepsilon r t'/2 = t$ by definition, the proof is complete. ■

Whew. Let us now give a simple corollary of the above result.

Porism 2.4. Suppose $H_1, \ldots, H_m$ are graphs. Then, for any $\varepsilon > 0$ and $n \gg 0$,

$$\mathrm{ex}(n; H_1, \ldots, H_m) \leq \left(1 - \frac{1}{r} + \varepsilon\right)\frac{n^2}{2} - \mathcal{O}(n^2),$$

where $r + 1 = \max\{\chi(H_1), \ldots, \chi(H_m)\}$.

Before we move onto the proof of the second version of the Erdős-Stone-Simonovits Theorem, we give a lemma that will assist in its proof.

Lemma 2.5. For $n \gg 0$, if $e(G_n) \geq (c+\varepsilon)\binom{n}{2}$ for some $c > 0$ and $\varepsilon > 0$, then there exists $H \subseteq G_n$ such that

1. $|H| \geq \sqrt{\varepsilon} n$.
2. $\delta(H) \geq c|H|$.

Proof. Assume that G_n itself does not satisfy the conclusions. Then, there exists some $x_n \in V(G_n)$ such that $d(x_n) < cn$.
Let $H_{n-1} = G \setminus \{x_n\}$. If H_{n-1} fails the second conclusion of the theorem, there exists $x_{n-1} \in V(H_{n-1})$ such that $d(x_{n-1}) < c(n-1)$.
Repeating the above, we get a sequence of graphs $G_n = H_n \supsetneq H_{n-1} \supsetneq \cdots \supsetneq H_\ell$, where $V(H_{n-r}) \setminus V(H_{n-r-1}) = \{x_{n-r}\}$ for each r and $\ell \geq \sqrt{\varepsilon} n$. Further, for each i, $d_{H_i}(x_i) < ci$ for $i = \ell, \ldots, n$.
Now,

$$\begin{aligned}
e(H_\ell) &> e(H_n) - \big(cn + c(n-1) + \cdots + c(\ell+1)\big) \\
&= (c+\varepsilon)\binom{n}{2} - c\left(\frac{n(n+1)}{2} - \frac{\ell(\ell+1)}{2}\right) \\
&= (c+\varepsilon)\binom{n}{2} - c\left(\binom{n}{2} + n - \binom{\ell+1}{2}\right) \\
&= \varepsilon\binom{n}{2} - cn + c\binom{\ell+1}{2}.
\end{aligned}$$

Now, the final expression must be at most $\binom{\ell}{2}$ (since H_ℓ has ℓ vertices).
As a result, it would suffice to show that the above expression on taking $\ell = \lfloor \varepsilon n \rfloor$ is greater than $\binom{\ell}{2}$ for sufficiently large n (this implies that the sequence must stop before reaching this ℓ due to one of the graphs satisfying the conclusions). Indeed, this is seen to be true as

$$\begin{aligned}
\varepsilon\binom{n}{2} - cn + c\binom{\ell+1}{2} &\geq \varepsilon\binom{n}{2} + c\left(\frac{(\lfloor\sqrt{\varepsilon}n\rfloor + 1)\lfloor\sqrt{\varepsilon}n\rfloor}{2} - n\right) \\
&\geq \varepsilon\binom{n}{2} && \text{(for sufficiently large } n\text{)} \\
&\geq \binom{\ell}{2},
\end{aligned}$$

completing the proof. ■

2.2. An Introduction to Random Graphs

Definition 2.3 (Erdős-Rényi Model)**.** Fix $0 \leq p \leq 1$. The **Erdős-Rényi random graph model**, denoted $G_{n,p}$ is the random variable which is a graph with vertex set $[n]$, such that for each $\{i,j\} \in \binom{[n]}{2}$, $\{i,j\}$ is an edge with probability p, independently across distinct pairs.

So, for any graph H on vertex set $[n]$,

$$\Pr(G_{n,p} = H) = p^{e(H)}(1-p)^{\binom{n}{2} - e(H)}.$$

Remark. $G_{n,1/2}$ is the uniform distribution on the set of graphs on $[n]$.
A recurring theme in probability theory is that random objects tend to behave very nicely given a large number of samples (along the lines of the laws of large numbers and the central limit theorem).

2.2.1. A motivating extremal problem (bounding $\mathrm{ex}(n; C_{2k})$)

To understand why random graphs are important, let us look at $\mathrm{ex}(n; C_{2k})$.

Theorem 2.6. For any k, there exists a constant c such that for $n \gg 0$, there is a C_k-free G_n with

$$e(G_n) \geq c \cdot n^{1+1/(k-1)}.$$

The above result does not yield anything useful for k odd.

Proof. Consider $G_{n,p}$ for some p we shall fix later. Let $N(G)$ be the number of copies of C_k in a given graph G.
Given a cycle $(v_1, \ldots, v_k)$, observe that the sequences $(v_2, v_3, \ldots, v_k, v_1)$ and $(v_k, v_{k-1}, \ldots, v_1)$ determine the same cycle. That is, performing cyclic shifts of a sequence of vertices or reversing their order around gives the same cycle.
Let $\mathcal{C}$ be the set of all these cycles.[5] By our observation,

$$|\mathcal{C}| = \frac{n(n-1)\cdots(n-k+1)}{2k} = \frac{n!}{2k \cdot k!}.$$

[5] This can be made more formal by taking all length k sequences of $[n]$ consisting of distinct elements and considering the equivalence classes formed by the equivalence relation defined on the previous line.

Now,

$$\begin{aligned}\mathbf{E}[N(G_{n,p})] &= \mathbf{E}\left[\sum_{(v_1,\ldots,v_k)\in\mathcal{C}} \mathbb{1}_{v_1v_2,v_2v_3,\ldots,v_{k-1}v_k,v_kv_1 \text{ are edges}}\right] \\ &= \sum_{(v_1,\ldots,v_k)\in\mathcal{C}} \Pr\left[\mathbb{1}_{v_1v_2,v_2v_3,\ldots,v_{k-1}v_k,v_kv_1 \text{ are edges}}\right] && \text{(linearity of expectation)} \\ &= \sum_{(v_1,\ldots,v_k)\in\mathcal{C}} p^k \\ &= \frac{n!}{2k\cdot k!}p^k.\end{aligned}$$

This quantity is obviously less than $(np)^k/2k$. If p, and thus the expectation is small, we expect to not see many C_ks. On the other hand,

$$\begin{aligned}\mathbf{E}[e(G_{n,p})] &= \mathbf{E}\left[\sum_{\{i,j\}\in\binom{[n]}{2}} \mathbb{1}_{ij \text{ is an edge}}\right] \\ &= \binom{n}{2}p.\end{aligned}$$

Given a graph, if we delete an (arbitrary) edge from each copy of C_k in it, we will be left with no cycles. That is, given any graph G, there is a graph on the same vertex set with $e(G) - N(G)$ edges that is C_k-free. Inspired by this, by the linearity of expectation,

$$\mathbf{E}[e(G_{n,p}) - N(G_{n,p})] \geq \binom{n}{2}p - \frac{(np)^k}{2k}.$$

If we set $p = \left(\frac{k}{2}\right)^{1/(k-1)} n^{-1+1/(k-1)}$, then the above quantity is at least $n(n-1)p/4 \geq c\cdot n^{1+1/(k-1)}$ for an appropriate constant c and $n \gg 0$, completing the proof. ∎

2.2.2. Digression: A coloring result of Erdős

The question we consider in this section is:

> Are there C_3-free graphs with large chromatic number?

More generally,

> Are there graphs with large chromatic number and large girth?

For example, if the girth of a graph is 7, then there cannot be adjacent vertices v, w such that $\Gamma(v)\cap\Gamma(u)$ and $\Gamma(w)\cap\Gamma(u)$ are non-empty for some u distinct from v, w. As a result, we can draw a "2-step tree" rooted at any u, which has $\Gamma(u)$ at the first level and the neighbours (other than u) of vertices of $\Gamma(u)$ at the second.
This seems to suggest some level of sparseness in the graph, due to which there are not too many edges and as a result, the chromatic number is low. However, it turns out that this intuition is not true, as proved by Erdős in [Erd59].

Theorem 2.7. There exist graphs with arbitrarily large girth and chromatic number. That is, given a $g, k \geq 3$, there exists a graph G_n such that $\mathrm{girth}(G) > g$ and $\chi(G) > k$.

Proof. Consider $G_{n,p}$ for some p we fix later. Further assume that $np \geq 1$.
Given a graph G, let $N_i(G)$ (for $3 \leq i \leq g$) be the number of cycles of size i in G. As we saw in the proof of Theorem 2.6,

$$\mathbf{E}[N_i(G_{n,p})] = \frac{n!}{2i \cdot i!} p^i < \frac{(np)^i}{6}.$$

Let $N = \sum_{i=3}^{g} N_i$. Then

$$\begin{aligned}
\mathbf{E}\left[N(G_{n,p})\right] &= \mathbf{E}\left[\sum_{i=3}^{g} N_i(G_{n,p})\right] \\
&< \frac{(np)^3}{6}\left(\frac{(np)^{g-2}-1}{np-1}\right) \\
&< \frac{(np)^g}{3}. && \text{(since } np \geq 1\text{)}
\end{aligned}$$

Using Markov's inequality,

$$\Pr\left[N(G_{n,p}) > \frac{2}{3}(np)^g\right] < \frac{1}{2} \tag{2.4}$$

This takes care of the girth (we want the above probability to be small). On the other hand, we need to make the chromatic number large. Towards this, observe that $\chi(G) \geq n/\alpha(G)$ (Why?). We have

$$\begin{aligned}
\Pr\left[\alpha(G_{n,p}) \geq r\right] &= \Pr\left[\bigcup_{X \subseteq [n]:|X|=r} \{X \text{ is independent}\}\right] \\
&\leq \sum_{X \subseteq [n]:|X|=r} \Pr[X \text{ is independent}] \\
&= \sum_{X \subseteq [n]:|X|=r} (1-p)^{\binom{r}{2}} \\
&\leq \binom{n}{r} \cdot e^{-pr(r-1)/2} \\
&\leq \left(\frac{en}{r}\right)^r \cdot e^{-pr^2/3} \\
&= \left(\frac{e^{1-pr/3} n}{r}\right)^r.
\end{aligned}$$

We shall choose r and p such that with positive probability, $N(G_{n,p}) \leq 2(np)^g/3$ *and* $\alpha(G_{n,p}) < r$. This implies the existence of a graph G_n such that both of the above hold.
We cannot use the tactic of removing edges we did in the earlier proof since that might increase α. Deleting vertices on the other hand works, since this can increase neither α nor N.
If we delete a single vertex from each cycle involved in N, the resulting graph will have girth greater than g. That is, given a graph G_n, there exists a graph with at least $n - N(G_n)$ vertices that has girth greater than g. Denote this corresponding graph as G'_n.

Set $p = n^{1/(g+1)-1}$. In this case, $2(np)^g/3 < n/2$ for $n \gg 0$ and using (2.4),

$$\Pr\left[n - N(G_{n,p}) > \frac{n}{2}\right] \geq \frac{1}{2}. \tag{2.5}$$

Set $r = 4\log n/p = 4n^{1-1/(g+1)} \log n$. For these values of p and r,

$$\Pr\left[\alpha(G_{n,p}) \geq r\right] \xrightarrow{n\to\infty} 0. \tag{2.6}$$

Using (2.5), our construction of G', and (2.7), it is true with with positive probability that

$$|G'_{n,p}| \geq \frac{n}{2}, \operatorname{girth}(G'_{n,p}) > g,$$

and

$$\chi(G'_{n,p}) \geq \frac{n}{\alpha(G'_{n,p})} \geq \frac{n}{r} = \frac{n^{1/(g+1)}}{4\log n} \xrightarrow{n\to\infty} \infty.$$

Therefore, for $g, k \geq 3$, there exists $n \gg 0$ and graph G on n vertices such that $\operatorname{girth}(G) > g$ and $\chi(G) > k$. ■

2.3. Szemerédi's Regularity Lemma

The second of our powerful results in extremal graph theory (after the Erdős-Stone-Simonovits Theorem) is Szemerédi's Regularity Lemma, which says that any sufficiently large graph behaves in some way like a random graph.

2.3.1. Motivation

First, let us give a bound from probability theory that will be useful.

Lemma 2.8 (Chernoff Bound). Suppose $X \sim B(n,p)$, the binomial distribution with parameters n, p. Then for any $t \geq 0$,

$$\Pr\left[X - \mathbf{E}[X] \geq t\right] \leq \exp\left(-\frac{t^2}{2\left(\mathbf{E}[X] + t/3\right)}\right)$$

and

$$\Pr\left[X - \mathbf{E}[X] \leq -t\right] \leq \exp\left(-\frac{t^2}{2\mathbf{E}[X]}\right).$$

Fix disjoint $A, B \subseteq [n]$ and let $|A| = a$, $|B| = b$. Then given a graph G on $[n]$,

$$e(A,B) = \sum_{x\in A, y\in B} \mathbb{1}_{xy\in E(G)}.$$

Fix $0 < p < 1$. Then if $G \sim G_{n,p}$,

$$e(A,B) \sim B\left(ab, p\right).$$

By the Chernoff Bound, for some fixed constant c,

$$\Pr\left[|e(A,B) - pab| > c\left(b\sqrt{pa\log\left(\frac{2n}{b}\right)}\right)\right] \xrightarrow{n\to\infty} 0.$$

In particular, if $a = b = \alpha n$ for some $0 < \alpha < 1/3$, then with high probability,

$$|e(A,B) - pab| = \mathcal{O}\left(b\sqrt{ap\log\left(\frac{2n}{b}\right)}\right)$$

for any sets A, B of sizes a and b respectively.
This seems to say that the actual number of edges between two sets of the given size does not deviate very much from the expected number of edges between the two sets. The expression on the right is of the order of $\mathcal{O}(n\sqrt{n})$, which is asymptotically less than the expectation $pab = \mathcal{O}(n^2)$.

2.3.2. The result

The regularity lemma gives a qualitative version of the above observation. Before we move to the actual result, let us provide some notation.

Definition 2.4 (Density)**.** Given a graph $G = (V, E)$ and $U, W \subseteq V$, the **density** $d(U, W)$ is equal to $e(U, W)/|U||W|$.

Remark. Here, $e(U, W)$ is $\{(u, w) \in U \times W : \{u, w\} \in E\}$. If U and W are disjoint, this is the same as our earlier definition of $e(\cdot, \cdot)$. If they are not disjoint however, edges within the intersection are counted *twice* in our current definition.
This does not matter all that much since we usually apply the regularity lemma on disjoint sets.

Definition 2.5 (ε-regular pair)**.** Suppose $0 < \varepsilon < 1$. A pair of subsets (U, W) is said to be **ε-regular** if for any $A \subseteq U$, $B \subseteq W$ with $|A| \geq \varepsilon|U|$ and $|B| \geq \varepsilon|W|$, we have

$$|d(A, B) - d(U, W)| \leq \varepsilon$$

This corresponds to some sort of uniform behaviour throughout the sets, where subsets behave similarly to their parent sets in terms of density. If (U, W) is ε-regular, all sufficiently large subsets of U, W have roughly the same edge density as (U, W).

Theorem 2.9 (Szemerédi's Regularity Lemma)**.** Given $0 < \varepsilon < 1$, there exists M such that for $n \gg 0$, any graph G_n admits a vertex partition $\mathcal{P} = (V_0, V_1, \ldots, V_k)$, where

- $k \leq M$,
- $|V_0| \leq \varepsilon n$ (V_0 is known as an "exceptional set"),
- all the V_i for $1 \leq i \leq k$ are of equal size, and
- the number of ε-*ir*regular pairs (V_i, V_j) $(1 \leq i, j \leq k)$ is at most εk^2.

Such a partition where the number of ε-regular pairs is at most εk^2 is often referred to as an **ε-regular partition**.

We present the proof of the above, that uses an "energy increment" argument, over a series of lemmas.

Definition 2.6 (Energy)**.** Given a graph G_n with vertex set V, for disjoint $U, W \subseteq V$, define the **energy** of the pair (U, W) by

$$q(U, W) = \frac{|U||W|}{n^2} d^2(U, W).$$

If $\mathcal{U} = \{U_1, \ldots, U_m\}$ and $\mathcal{W} = \{W_1, \ldots, W_\ell\}$ are partitions of U and W respectively, then the energy of the pair $(\mathcal{U}, \mathcal{W})$ is

$$q(\mathcal{U}, \mathcal{W}) = \sum_{\substack{1 \leq i \leq m \\ 1 \leq j \leq \ell}} q(U_i, W_j).$$

For a partition $\mathcal{P}$ of V, let
$$q(\mathcal{P}) = q(\mathcal{P}, \mathcal{P}) = \sum_{U,W \in \mathcal{P}} q(U, W).$$

Lemma 2.10. If U, W are disjoint subsets of V and $\mathcal{U}, \mathcal{W}$ are partitions of U, W respectively, then
$$q(\mathcal{U}, \mathcal{W}) \geq q(U, W).$$

Proof. Independently pick u, w uniformly randomly from V. Define the random variable
$$Z(u, w) = \begin{cases} d(U', W'), & u \in U' \in \mathcal{U}, w \in W' \in \mathcal{W}, \\ 0, & \text{otherwise.} \end{cases}$$

Then,
$$\begin{aligned}
\mathbf{E}[Z]^2 &= \left(\sum_{\substack{U' \in \mathcal{U} \\ W' \in \mathcal{W}}} d(U', W') \left(\frac{|U'||W'|}{|U||W|} \right) \right)^2 \\
&= \left(\frac{1}{|U||W|} \sum_{\substack{U' \in \mathcal{U} \\ W' \in \mathcal{W}}} e(U', W') \right)^2 \\
&= d(U, W)^2 \\
&= \frac{n^2}{|U||W|} q(U, W) \text{ and} \\
\mathbf{E}[Z^2] &= \sum_{\substack{U' \in \mathcal{U} \\ W' \in \mathcal{W}}} d^2(U', W') \left(\frac{|U'||W'|}{|U||W|} \right) \\
&= \frac{n^2}{|U||W|} q(\mathcal{U}, \mathcal{W}).
\end{aligned}$$

Since variance is always non-negative, $\mathbf{E}[Z^2] \geq \mathbf{E}[Z]^2$ and thus,
$$q(\mathcal{U}, \mathcal{W}) \geq q(U, W).$$
■

Partitioning increases energy! Also observe that since the random variable Z is at most 1 (1/2, in fact), so is the energy.
This also implies that if the partition $\mathcal{Q}$ is a refinement of $\mathcal{P}$, then
$$q(\mathcal{Q}) \geq q(\mathcal{P}).$$

Lemma 2.11. If (U, W) is not ε-regular, there is a partition $\mathcal{U} = \{U_1, U_2\}$ and $\mathcal{W} = \{W_1, W_2\}$ of U, W respectively such that
$$q(\mathcal{U}, \mathcal{W}) > q(U, W) + \varepsilon^4 \frac{|U||W|}{n^2}.$$

Proof. By the ε-irregularity, there exist $U_1 \subseteq U$ and $W_1 \subseteq W$ with $|U_1| \geq \varepsilon |U|$, $|W_1| \geq \varepsilon |W|$, and
$$|d(U_1, W_1) - d(U, W)| > \varepsilon.$$

Consider the partitions $\mathcal{U} = \{U_1, U \setminus U_1\}$ and $\mathcal{W} = \{W_1, W \setminus W_1\}$. With the same Z as in the proof of the previous lemma,

$$\begin{aligned}\mathbf{E}[Z^2] - \mathbf{E}[Z]^2 &= \mathbf{E}[(Z - \mathbf{E}[Z])^2] \\ &\geq \mathbf{E}[(Z - \mathbf{E}[Z])^2 \mathbb{1}_{u \in U', w \in W'}] \\ &= \varepsilon^4.\end{aligned}$$

The claim follows. ■

Lemma 2.12. Suppose $0 < \varepsilon < 1/4$ and let $\mathcal{P} = (V_0, V_1, \dots, V_k)$ be a partition of V such that

- $\mathcal{P} \setminus \{V_0\}$ is not ε-regular (there are at least εk^2 irregular pairs),
- $\mathcal{P} \setminus \{V_0\}$ is equitable[6], and
- $|V_0| \leq \varepsilon n$.

Then, there is a refinement $\mathcal{Q} = (V_0', V_1', \dots, V_\ell')$ of $\mathcal{P}$ such that

- $\ell \leq k4^k$,
- $\mathcal{Q} \setminus \{V_0'\}$ is equitable,
- $|V_0'| \leq |V_0| + n/2^k$, and
- $q(\mathcal{Q}) \geq q(\mathcal{P}) + \varepsilon^5/2$.

Proof. Let $|V_i| = t$ for all $1 \leq i \leq k$.
Suppose (V_i, V_j) is an ε-irregular pair in $\mathcal{P}$ and let $(V_{i,1}, V_{i,2})$ and $(V_{j,1}, V_{j,2})$ be the partitions of V_i, V_j respectively described in Lemma 2.11. If $\mathcal{Q}_1$ is this particular refinement

$$\mathcal{P} \cup \{V_{i,1}, V_{i,2}, V_{j,1}, V_{j,2}\} \setminus \{V_i, V_j\}$$

of $\mathcal{P}$, then

$$q(\mathcal{Q}_1) \geq q(\mathcal{P}) + \varepsilon^4 \frac{|V_i||V_j|}{n^2}$$

Let us similarly produce refinements corresponding to all irregular pairs of $\mathcal{P}$. Let $\mathcal{Q}'$ be the "common" refinement of all these partitions $(\mathcal{Q}_i)$. That is, for each $v \in V$, v is placed in the subset $\bigcap_{W \in \mathcal{Q}_i : v \in W} W$.
We then have

$$\begin{aligned}q(\mathcal{Q}') &\geq q(\mathcal{P}) + \varepsilon^4 \cdot \frac{t^2}{n^2}(\varepsilon k^2) \\ &= q(\mathcal{P}) + \varepsilon^5 \frac{(tk)^2}{n^2} \\ &\geq q(\mathcal{P}) + \frac{\varepsilon^5}{2}. \qquad (tk = n - |V_0| \geq (1-\varepsilon)n \geq 3n/4 \text{ and } 9/16 \geq 1/2) \end{aligned} \tag{2.7}$$

To make the partition $\mathcal{Q}'$ equitable, create the partition $\mathcal{Q}$ using it as follows.
Suppose we partition V_i into (V_{ij}), where V_{ij} is formed due to the irregularity of (V_i, V_j). Partition each of these parts of $\mathcal{Q}'$ into sets of size $b := \lfloor t/4^k \rfloor$. Whatever residual part cannot be cut out in this manner, we merge with V_0.
Since partitioning can only increase energy, this operation will only strengthen (2.7), if anything.

[6] All the blocks of the partition are of equal size.

Since the size of any block of $\mathcal{Q}$ is b, the number of blocks is at most $n/b = k4^k$.
Finally, what is the size of $|V_0'|$?
$\mathcal{Q}'$ has at most $k \cdot 2^{k-1}$ parts (there are at most $(k-1)$ 2-part partitions of each block of $\mathcal{P}$, and together they give a partition of size at most 2^{k-1}). So,

$$\begin{aligned} |V_0'| &\leq |V_0| + |\mathcal{Q}'|b \\ &\leq |V_0| + k2^{k-1}\frac{t}{4^k} \\ &\leq |V_0| + \frac{n}{2^k}, \end{aligned}$$

as desired. ■

Observe that with the above lemma, Szemerédi's Regularity Lemma follows without too much difficulty.
Begin with a partition $\mathcal{P}_0$ with k_0 parts, where $2^{k_0} \geq 2/\varepsilon$.
Given $\mathcal{P}_k$, let $\mathcal{P}_{k+1}$ be the partition defined by Lemma 2.12. Since the energy of any partition is bounded above by 1, this process must terminate after at most $2/\varepsilon^5$ steps.
Further, since the size of the refined partition is bounded between quantities dependent solely on the old partition, the size of the final partition after termination of the above process is at most some quantity dependent only on ε (this quantity might be massive[7], but that is besides the point).

2.4. Some Corollaries of Szemerédi's Regularity Lemma

In this section, we cover several interesting corollaries of Szemerédi's Regularity Lemma.

Before we begin however, how do we process the regularity lemma? We typically begin with a "cleaning" of the graph as follows.

- Given $\varepsilon > 0$, get a partition as described in the lemma.
- First, delete all edges between irregular pairs. This loses at most $(\varepsilon k^2)t^2 < \varepsilon n^2$ edges.
- Delete all edges between "sparse" pairs, where we say that a pair (V_i, V_j) is sparse if $d(V_i, V_j) < \varepsilon$ (say). This loses at most $k^2(\varepsilon t^2) < (\varepsilon/2)n^2$ edges.
- Delete all edges inside the exceptional part V_0. This loses at most $(\varepsilon^2/2)n^2$ edges.

All of these deletions cost at most $2\varepsilon n^2$ edges. In the remaining graph, *all* pairs (V_i, V_j) for $1 \leq i, j \leq k$ are ε-regular. Further, if $e(V_i, V_j) > 0$, then $d(V_i, V_j) \geq \varepsilon$.

2.4.1. The graph counting and removal lemmas

Theorem 2.13 (Triangle Counting Lemma)**.** Suppose V_1, V_2, V_3 forms a pairwise ε-regular partition of the vertex set of graph G_n, and that $d(V_i, V_j) \geq d$ (for some $d \geq 2\varepsilon$). Then, there are at least

$$\varepsilon^2(1-2\varepsilon)(d-\varepsilon)|V_1||V_2||V_3|$$

triangles xyz with $x \in V_1$, $y \in V_2$, $z \in V_3$.

[7] Unbelievably so. For example, $\varepsilon = 1/8$ gives a bound of the order of $4 \uparrow\uparrow 2^{15}$ (using Knuth's up-arrow notation). It further turns out that such a massive tetration-type bound is necessary, as proved in [Gow97].

Proof. Let
$$X_1 = \{x \in V_1 : d(x, V_2) \leq \varepsilon |V_2|\}.$$
We claim that $|X_1| < \varepsilon|V_1|$. Suppose otherwise. Then $e(X_1, V_2)/|X_1||V_2| < \varepsilon$ and further, ε-regularity implies that $d(X_1, V_2) \geq (d - \varepsilon)$, which leads to a contradiction.
Similarly, $X_2 := \{x \in V_1 : d(x, V_3) \leq \varepsilon|V_3|\}$ has size less than $\varepsilon|V_1|$ too.
For all $x \in V_1 \setminus (X_1 \cup X_2)$, $d(x, V_2) \geq \varepsilon|V_2|$ and $d(x, V_3) \geq \varepsilon|V_3|$. Fix such an x and let $X_i' = \Gamma(x) \cap V_i$ (for $i = 2, 3$). By the ε-regularity of (V_2, V_3), $d(V_2, V_3) - d(X_2', X_3') \leq \varepsilon$ so $d(X_2', X_3') \geq d - \varepsilon$. As a result,
$$e(X_2', X_3') \geq (d - \varepsilon)|X_2'||X_3'| \geq (d - \varepsilon)\varepsilon^2|V_2||V_3|.$$
Therefore, the number of triangles of the desired form is at least
$$\underbrace{(1 - 2\varepsilon)|V_1|}_{\substack{\text{the number of} \\ \text{such } x}} \cdot \underbrace{(d - \varepsilon)\varepsilon^2|V_2||V_3|}_{\substack{\text{the number of triangles} \\ \text{corresponding to each } x}},$$
completing the proof. ■

Along similar lines is the following result.

Theorem 2.14 (Triangle Removal Lemma)**.** Given $\varepsilon > 0$, there exists δ (depending only on ε) such that for $n \gg 0$, any graph G_n with at most δn^3 triangles can be made triangle-free by deleting at most εn^2 edges.

Proof. Start with an $(\varepsilon/4)$ regular partition of G_n using Szemerédi's Regularity Lemma and the cleaning process from earlier. Delete all edges within each of the V_i. This costs at most $k(t^2/2) < (kt)^2/2k \leq \varepsilon n^2/4$ edges.
So, we have lost $\leq \varepsilon n^2$ edges in all. If there is a triangle remaining in the graph, it must come from a triple (V_i, V_j, V_k) with all three pairs being $(\varepsilon/4)$-regular and density at least $\varepsilon/2$. We can then use the triangle counting lemma to conclude that there are at least $t^3(\varepsilon/4)^3(1 - \varepsilon/2) > n^3(\varepsilon/8)^3/M(\varepsilon)^3$. Letting δ to be the $1/6$ this quantity, we are done. ■

Let (V_1, V_2), (V_2, V_3), and (V_3, V_1) all be ε-regular pairs with densities $d_{12}, d_{23}, d_{13} > 2\varepsilon$. A random graph on $V_1 \sqcup V_2 \sqcup V_3$ with edge probabilities d_{ij} between V_i and V_j has an expected number of triangles of $|V_1||V_2||V_3|d_{12}d_{23}d_{31}$. Ideally, we would have a result that the number of triangles is indeed to this quantity (so the graph behaves almost randomly, in a sense similar to that in Szemerédi's Regularity Lemma).

Theorem 2.15 (Graph Counting Lemma)**.** Let G be a graph on n vertices and H a graph on $[k]$. Let $V_1, \ldots, V_k \subseteq V(G)$ such that (V_i, V_j) is ε-regular whenever $ij \in E(H)$. Then,
$$\left|\{(v_1, \ldots, v_k) : v_i \in V_i \text{ and } \{v_1, \ldots, v_k\} \text{ form a copy of } H \text{ in } G\}\right|$$
is within $\varepsilon e(H)|V_1||V_2| \cdots |V_k|$ of
$$\prod_{i=1}^{k} |V_i| \prod_{ij \in E(H)} d(V_i, V_j),$$
assuming that $\prod_{ij \in E(H)} d(V_i, V_j) > \varepsilon e(H)$.

Proof. We shall prove this by inducting on $e(H)$. If $e(H) = 0$, the result is trivial.
Let us rephrase the problem probabilistically. Pick $v_i \in V_i$ independently and uniformly. Then, we wish to prove that
$$\left|\Pr\left[v_i v_j \in E(G) \text{ for all } ij \in E(H)\right] - \prod_{ij \in E(H)} d(V_i, V_j)\right| \leq \varepsilon e(H). \tag{2.8}$$

Suppose $\{1,2\} \in E(H)$. It suffices to show that

$$\left| \Pr\left[v_i v_j \in E(G) \text{ for all } ij \in E(H)\right] - d(V_1, V_2) \Pr\left[v_i v_j \in E(G) \text{ for all } E(H) \setminus \{1,2\}\right] \right| \leq \varepsilon. \tag{2.9}$$

Indeed, by induction,

$$\left| \Pr\left[v_i v_j \in E(G) \text{ for all } E(H) \setminus \{1,2\}\right] - \prod_{\substack{ij \in H \\ ij \neq \{1,2\}}} d(V_i, V_j) \right| < \varepsilon(e(H) - 1), \tag{2.10}$$

and (2.8) follows from (2.9) and (2.10) on using the union bound. We shall prove that (2.9) holds when we condition on the choices v_i for $i > 2$, and thus generally. Let

$$A_j = \{v_j \in V_j : \{v_j, v_i\} \in E(G) \text{ when } \{j,i\}\} \in E(H).$$

for $j = 1, 2$. Equation (2.9) is then equivalent to

$$\left| \frac{e(A_1, A_2)}{|V_1||V_2|} - d(V_1, V_2) \frac{|A_1||A_2|}{|V_1||V_2|} \right| \leq \varepsilon. \tag{2.11}$$

We claim that (2.11) holds for all A_1, A_2. If $|A_1| \geq \varepsilon |V_1|$ and $|A_2| \geq \varepsilon |V_2|$, then ε-regularity implies that $|d(A_1, A_2) - d(V_1, V_2)| \leq \varepsilon$, that is,

$$\left| \frac{e(A_1, A_2)}{|A_1||A_2|} - d(V_1, V_2) \right| \leq \varepsilon.$$

So,

$$\frac{|V_1||V_2|}{|A_1||A_2|} \left| \frac{e(A_1, A_2)}{|V_1||V_2|} - d(V_1, V_2) \frac{|A_1||A_2|}{|V_1||V_2|} d(V_1, V_2) \right| \leq \varepsilon,$$

and (2.11) easily follows. If $|A_1| < \varepsilon |V_1|$, then the above follows immediately anyway, since

$$\begin{aligned} \left| \frac{e(A_1, A_2)}{|V_1||V_2|} - d(V_1, V_2) \frac{|A_1||A_2|}{|V_1||V_2|} \right| &= \frac{|A_1||A_2|}{|V_1||V_2|} \left| \frac{e(A_1, A_2)}{|A_1||A_2|} - d(V_1, V_2) \right| \\ &= \frac{|A_1||A_2|}{|V_1||V_2|} \left| d(A_1, A_2) - d(V_1, V_2) \right| \\ &\leq \frac{|A_1|}{|V_1|} < \varepsilon, \end{aligned}$$

so we are done. ∎

The generalized version of the triangle removal lemma is the following.

Theorem 2.16 (Graph Removal Lemma). Given $\varepsilon > 0$ and any graph H, there exists δ (depending only on ε) such that for $n \gg 0$, any graph G_n with at most $\delta n^{V(H)}$ subgraphs isomorphic to H can be made H-free by deleting at most εn^2 edges.

2.4.2. Roth's Theorem and Corners

Next, we describe Roth's Theorem. The result deals with a conjecture of Erdős and Turán:

> Given $\varepsilon > 0$ and $r \in \mathbb{N}$, there exists N_0 such that for all $N \geq N_0$, the following holds. If $A \subseteq [N]$ with $|A| \geq \varepsilon N$, then A contains an arithmetic progression of length r.

Roth proved in the early 50s that the conjecture holds for $r = 3$. Szemerédi proved in the early 70s that it holds for all r. In fact, Szemerédi had originally come up with the regularity lemma to prove this result (this is why it is a *lemma*).

Lemma 2.17. Suppose that every edge of G_n is in exactly one triangle. Then, $e(G_n) = o(n^2)$.

How is this related to Roth's Theorem? Given $N \gg 0$ and $A \subseteq [N]$, suppose A is 3-AP free. Let $M = 2N + 1$ and construct a 3-partite graph G whose three parts X, Y, and Z are copies of $\mathbb{Z}/M$.
For $x \in X$, $y \in Y$, and $z \in Z$, keep an edge from x to y iff $y - x \in A$, y to z iff $z - y \in A$, and z to x iff $(z - x)/2 \in A$.
The final part is well-defined since 2 is invertible in $\mathbb{Z}/M$ (Why?).
Observe that if xyz is a triangle with $a = y - x$, $b = z - y$, and $2c = z - x$, then $a + b = 2c$, so a, c, b are in AP.
Since a, b, and c are all in A, then by our assumption we must have $a = b = c$, $y - x = z - y = (z - x)/2$, and so x, y, z must be in AP in $\mathbb{Z}/M$.
This implies that any edge must be in a unique triangle.
By Lemma 2.17, $e(G) = o((3M)^2) = o(N^2)$. On the other hand, $e(G) = 3M|A|$. So, $3M|A| = o(N^2)$ and $|A| = o(N)$.

Proof of Lemma 2.17. The number of triangles in G_n is exactly $e(G_n)/3 = o(n^3)$. By the Triangle Removal Lemma, G_n can be made triangle-free by removing $o(n^2)$ edges. However, we must remove at least $e(G_n)/3$ edges to make the graph triangle-free, so $e(G_n) = o(n^2)$. ■

Inspired by this problem, a natural question to ask is: what is the maximum sized $A \subseteq [N]$ that is 3-AP free?
Picking A greedily gives $O(\sqrt{n})$ elements.

It is possible to do better, as shown by *Behrend's construction*.
The basic idea is that on a sphere, the midpoint of any two points does not lie on the sphere.
Consider the set $S = [k]^d \subseteq \mathbb{R}^d$ for some k, d. Observe that $\|x\|^2 \le k^2 d$ for any point in S. By the pigeonhole principle, there is a (centered) sphere with at least k^{d-2}/d of these points.
To get a subset A from these points, project $x = (a_1, \dots, a_d)$ to $\sum_{j=1}^{d} a_j (2k+1)^{j-1}$, thus converting it to an integer in base $(2k + 1)$. Observe that if $x_1 \mapsto m_1$ and $x_2 \mapsto m_2$, then $(x_1 + x_2)/2 \mapsto (m_1 + m_2)/2$. As a result, if X is a subset of S on a single sphere, its projection by this map gives a 3-AP free subset.
The maximum element by this projection is $k \sum_{i=0}^{d-1} (2k + 1)^i = ((2k + 1)^d - 1)/2$. Set $(2k + 1)^d = 2N + 1$. By the argument from before, there is a 3-AP free set of size k^{d-2}/d.
This achieves a bound of

$$\frac{N}{e^{c\sqrt{\log n}}}$$

for some constant c. This is $\Omega(N^{1-\delta})$ for any (fixed) $\delta > 0$!

Definition 2.7 (Corner)**.** An $A \subseteq [N]^2$ is said to have a **corner** if the points (x, y), $(x + d, y)$, and $(x, y + d)$ are in A for some $d > 0$ and $x, y \in [N]$.

Theorem 2.18 (No Corners Theorem)**.** Suppose $A \subseteq [N]^2$ has no corners. Then $|A| = o(N^2)$.

Proof. First, let us get rid of the $d > 0$ clause.
Denote by $A + A$ the set $\{a + b : a, b \in A\} \subseteq [2N]^2$ (the Minkowski sum of the two sets) and by $x + A$ $\{x + a : a \in A\}$.
By the pigeonhole principle, there exists a $z \in [2N]^2$ such that $z = a + b$ for at least $|A|^2/4N^2$ pairs $(a, b) \in A^2$.
Consider $A' = A \cap (z - A)$ for such a z. Then

$$|A'| \ge \frac{|A|^2}{4N^2}$$

by the definition of z. It is not too difficult to see that $A' = (z - A')$. As a result, there is a correspondence between corners with positive d and negative d in A'. Further, if A is corner-free, so is A'. By the cardinality bound on A', it suffices to show that $|A'| = o(N^2)$.
So, let us drop the $d > 0$ condition on corners and work with a corner-free A.

We construct a 3-partite graph G. Let the three parts of the vertex set V be the set H of horizontal lines in $[N]^2$, the set V of vertical lines in $[N]^2$, and the set D of lines with slope -1 in $[2N]^2$.
Given ℓ, ℓ' in distinct parts of V, let ℓ and ℓ' be adjacent in G iff $\ell \cap \ell' \in A$. Triangles in this graph correspond to either corners or three lines all passing through a single point (that is in A). However, there are no corners in A, so each edge is in a unique triangle. The result then follows from Lemma 2.17. ∎

2.4.3. A weaker version of the Erdős-Stone-Simonovits Theorem

Next, we shall prove a slightly weaker version of the Erdős-Stone-Simonovits Theorem, Version 2 using Szemerédi's Regularity Lemma.

Theorem 2.19 (Erdős-Stone-Simonovits Theorem, Version 3)**.** Fix $t \in \mathbb{N}$ and $\varepsilon > 0$. For $n \gg 0$, if G_n is $K_{r+1}(t)$-free,[8] then

$$e(G_n) \leq \left(1 - \frac{1}{r}\right)\frac{n^2}{2} + \varepsilon n^2.$$

This is weaker than the original result we saw since here, t is fixed and so cannot grow to the $\log n$ estimate we gave earlier.

Proof. Suppose $e(G_n) > (1 - 1/r)n^2/2 + \varepsilon n^2$ edges.
Let $\delta > 0$ which we shall fix later. By Szemerédi's regularity lemma, we get a partition $(V_0, V_1, \ldots, V_k)$ such that $|V_0| \leq \delta n$, $|V_i| = \ell$ (for some ℓ) for $1 \leq i \leq k$, and at most δk^2 of the pairs are δ-irregular.
'Clean' the graph by deleting all edges within the V_i (losing $\leq \ell^2 k/2$ edges), edges adjacent to V_0 (losing $\leq \delta n^2$ edges), edges in irregular pairs (losing $\leq \delta k^2 \ell^2$ edges), and edges between pairs of density less than δ (losing $\leq \ell\delta(k^2/2)$ edges).
The number of deleted edges is thus at most $4\delta n^2$. As a result, the graph after purification still has $> (1 - 1/r)n^2/2 + (\varepsilon - 4\delta)n^2$ edges.

Let H be a graph with vertex set $\{V_i : 1 \leq i \leq k\}$, and $V_iV_j \in E(H)$ iff (V_i, V_j) is $(\varepsilon/8)$-regular with density at least $\varepsilon/4$.
By the pigeonhole principle, since $e(V_i, V_j) \leq \ell^2$

$$e(H) \geq \left(1 - \frac{1}{r} + \varepsilon\right)\frac{n^2}{2\ell^2} \geq \left(1 - \frac{1}{r} + \varepsilon\right)\frac{k^2}{2}.$$

By Turán's Theorem, H contains K_{r+1}. Say $(V_1, \ldots, V_{r+1})$ form this K_{r+1}. All (V_i, V_j) are $(\varepsilon/8)$-regular with density at least $\varepsilon/4$.
We now use the Graph Counting Lemma. Take $F = K_{r+1}(t)$. The number of copies of F in $(V_1, \ldots, V_{r+1})$ is at least

$$\prod_{i=1}^{r+1} |V_i|^t \left((\varepsilon/4)^{t^2\binom{r+1}{2}} - \delta t^2 \binom{r+1}{2}\right).$$

So, if

$$\delta \leq \frac{(\varepsilon/4)^{t^2\binom{r+1}{2}}}{2t^2\binom{r+1}{2}},$$

then there are several copies of K_{r+1} in G, completing the proof. ∎

[8] This is the complete $(r+1)$-partite graph with t vertices in each part, that is, $T_{r+1}(t(r+1))$.

2.4.4. An application in computer science

Now, we give a delightful application from computer science.
Suppose we are given a graph G_n and we wish to check (algorithmically) whether G has any triangles. An easy way to do this is to iterate through all triples and check if any of them form a triangle, which takes $\mathcal{O}(n^3)$ time.
In this context, we give a paradigm due to Shafi Goldwasser and Madhu Sudan. We want to come up with an algorithm that takes a *constant* number of verifications, and still returns a reasonable output. It is clearly impossible to return the answer correctly all the time. The output we desire is such that with high probability,

- if G_n has no triangles, our output is correct.
- if G_n has triangles and we output otherwise, then G_n must be "close" to triangle-free.

Such a randomized constant time algorithm is known as a **property tester**. The question we wish to answer is: which graph properties admit such testers?
For starters, does triangle-freeness admit a property tester? It turns out that it *does*! The algorithm is in fact very simple.

- Pick a random triple (x, y, z) and check if it forms a triangle.
- If yes, return "not triangle-free".
- If no, return "ε-close to triangle-free", where at most εn^2 edges need to be deleted to remove all triangles.

Why is this a tester?
By the Triangle Removal Lemma, if G_n is ε-close to being triangle-free, then G_n has at most δn^3 triangles for some δ. That is,

$$\Pr[xyz \text{ is not a triangle}] \geq 1 - \frac{\delta n^3}{\binom{n}{3}} \geq 1 - 6\delta,$$

so our algorithm is indeed a property tester.
We shall return to property testing in more detail later.

2.4.5. The Frankl-Pach Theorem

Theorem 2.20 (Frankl-Pach Theorem)**.** Suppose $s, r \in \mathbb{N}$ and $0 < c < 1$. Then, for $n \gg 0$,

$$\text{ex}(n; K_{r+1}, K_{s,\lceil cn \rceil}) \leq \left(c^{1/s} \left(1 - \frac{1}{r}\right)^{1/s} + o(1) \right) n^2.$$

Observe that one of the graphs we are forbidding depends on n itself.

Proof. Let $\varepsilon > 0$ and G be a graph on n vertices that is K_{r+1}- and $K_{s,\lceil cn \rceil}$-free. Using Szemerédi's regularity lemma and the cleaning process, obtain a graph G' such that

- $V(G') = V_1 \sqcup \cdots \sqcup V_k$ where $k = O_\delta(1)$,
- $|V_i| = \ell$ for all i and $V(G') \geq (1 - \delta)n$,
- All pairs (V_i, V_j) are δ-regular,
- if $e(V_i, V_j) > 0$, $d(V_i, V_j) \geq \eta$, and
- $e(G) - e(G') \leq \varepsilon n^2$

for some $\delta, \eta > 0$ depending only on ε. Henceforth, we operate only in G'.

Construct a graph H on vertex set $[k]$ with $ij \in E(H)$ iff $d(V_i, V_j) \geq \eta$. By the Graph Counting Lemma, if H contains K_{r+1}, so does G' and thus G. Therefore, H is K_{r+1}-free and by Turán's Theorem, $e(H) \leq (1 - 1/r)k^2/2$.
Since $n \gg 0$ and $k = O_\delta(1)$, if $K_{s,\lceil cn \rceil}$ appears in G, one of the V_is must have at least s vertices for $n \gg 0$ (by a simple pigeonhole argument).
So, let us bound $e(G)$ given that $K_{s,\lceil cn \rceil}$ is forbidden and the size s part of the bipartite graph does not come from a single V_i.
Similar to what we did in the Kővári-Sós-Turán Theorem, let us count pairs (x, S) where $|S| = s$, $S \subseteq V_i$ for some i, and x is adjacent to every $y \in S$. We call such a structure an (i, s)-*claw*. We have that

$$\text{number of } (i,s)\text{-claws } = \sum_{x \in V'} \sum_{i=1}^{k} \binom{|N(x) \cap V_i|}{s}.$$

On the other hand, by our assumption regarding $K_{s,\lceil cn \rceil}$-freeness,

$$\text{number of } (i,s)\text{-claws } \leq (\lceil cn \rceil - 1) \sum_{i=1}^{k} \binom{|V_i|}{s} \leq cnk \binom{\ell}{s}.$$

Therefore,

$$\sum_{x \in V'} \sum_{i=1}^{k} \binom{|N(x) \cap V_i|}{s} \leq cnk \binom{\ell}{s}.$$

Define $\mathcal{P} = \{(x, i) : |N(x) \cap V_i| > 0\}$. Applying Jensen's inequality in a manner similar to that in proof of the Kővári-Sós-Turán Theorem,

$$\begin{aligned}
\sum_{(x,i) \in \mathcal{P}} \binom{|N(x) \cap V_i|}{s} &\geq |\mathcal{P}| \binom{(1/|\mathcal{P}|) \sum_{(x,i) \in \mathcal{P}} |N(x) \cap V_i|}{s} \\
&= |\mathcal{P}| \binom{2e(G')/|\mathcal{P}|}{s} \\
&= |\mathcal{P}| \binom{u}{s},
\end{aligned}$$

where $u = 2e(G')/|\mathcal{P}|$. Then,

$$\begin{aligned}
cnk \frac{\ell^s}{s!} &\geq cnk \binom{\ell}{s} \\
&\geq \sum_{x \in V'} \sum_{i=1}^{k} \binom{|N(x) \cap V_i|}{s} \\
&= \sum_{(x,i) \in \mathcal{P}} \binom{|N(x) \cap V_i|}{s} \\
&\geq |\mathcal{P}| \binom{u}{s} \\
&\geq |\mathcal{P}| \frac{(u - s + 1)^s}{s!}.
\end{aligned}$$

Simplifying,

$$e(G') \leq \frac{u|\mathcal{P}|}{2} \leq \frac{1}{2} \left((cnk)^{1/s} \ell |\mathcal{P}|^{1-1/s} + |\mathcal{P}|(s-1) \right). \tag{2.12}$$

To bound $|\mathcal{P}|$ from above, observe that

$$|\mathcal{P}| \leq e(H) \cdot 2\ell$$

since each edge in H gives at most 2ℓ choices for x (if the edge is ij, x is in V_i or V_j). By K_{r+1}-freeness, $e(H) \leq (1-1/r)k^2/2$ so

$$|\mathcal{P}| \leq \left(1 - \frac{1}{r}\right) k^2 \ell.$$

Substituting the above back in (2.12),

$$\begin{aligned}
e(G') &\leq \frac{1}{2}\left((cnk)^{1/s}\ell|\mathcal{P}|^{1-1/s} + |\mathcal{P}|(s-1)\right) \\
&\leq \frac{1}{2}\left((cn)^{1/s}\left(1-\frac{1}{r}\right)^{1/s}(k\ell)^{2-1/s} + \left(1-\frac{1}{r}\right)k^2\ell(s-1)\right) \\
&\leq \frac{1}{2}\left(c^{1/s}\left(1-\frac{1}{r}\right)^{1/s} n^2 + \left(1-\frac{1}{r}\right)k(s-1)n\right) \\
&= \left(c^{1/s}\left(1-\frac{1}{r}\right)^{1/s} + o(1)\right)\frac{n^2}{2}.
\end{aligned}$$

The result follows since $e(G) - e(G') \leq \varepsilon n^2$. ■

It turns out that the bound given by the Frankl-Pach Theorem is asymptotically tight!

2.4.6. A look at $\mathrm{ex}(n; C_{2k+1})$

As we saw back in Corollary 2.3,

$$\frac{n^2}{4} \leq \mathrm{ex}(n; C_{2k+1}) \leq \left(\frac{1}{4} + o(1)\right) n^2.$$

Recall that if $e(G_n) \geq n^2/4 - t$ and G_n is K_3-free, then G_n can be made bipartite by deleting atmost t edges.

Lemma 2.21. Given $\varepsilon > 0$, there exists $\delta > 0$ such that the following holds for $n \gg 0$. If G_n is C_{2k+1}-free and has $\geq (1/4 - \delta)n^2$ edges, it can be made bipartite by deleting at most εn^2 edges.

Proof. First, we shall try to make $\delta(G_n)$ 'large'. If G_n has a vertex of degree $< (1/2 - 2\sqrt{\delta})n$, delete this vertex to get a graph G_{n-1}. Perform this process repeatedly to finally arrive at some G_ℓ such that

$$\begin{aligned}
e(G_\ell) &> \left(\frac{1}{4} - \delta\right) n^2 - \left(\frac{1}{2} - 2\sqrt{\delta}\right)(n + (n-1) + \cdots + (\ell+1)) \\
&= \left(\frac{1}{4} - \delta\right) n^2 - \left(\frac{1}{2} - 2\sqrt{\delta}\right)\left(\binom{n+1}{2} - \binom{\ell+1}{2}\right) \\
&= \left(\frac{1}{4} - \delta\right) n^2 - \left(\frac{1}{4} - \sqrt{\delta}\right)\left(n^2 + n - \ell^2 - \ell\right) \\
&\geq \left(\frac{1}{4} - \delta\right)\ell^2 + (\sqrt{\delta} - \delta)n^2 - \left(\frac{1}{4} - \delta\right) n.
\end{aligned}$$

If the quantity above is at least $(1/4 + \eta)\,\ell^2$ for some constant $\eta > 0$, we have

$$\ell < \sqrt{\frac{\sqrt{\delta} - \delta}{\eta + \delta}} n.$$

This implies that for sufficiently large ℓ (by virtue of sufficiently large n), G_ℓ has C_{2k+1} which yields a contradiction. Therefore, setting η appropriately, this process must terminate with $\ell \geq (1 - 4\sqrt{\delta})n$.

Therefore, suppose $G' = (V', E')$ is a subgraph of G_n such that $|V'| = n' \geq (1 - 4\sqrt{\delta})n$, $\delta(G') \geq (1/2 - 2\sqrt{\delta})n'$, and G' is C_{2k+1}-free. We have $e(G') \geq (1/4 - \sqrt{\delta})n'^2$.
Recall that $\text{ex}(n; C_{2k})$ is $o(n^2)$. As a result, it follows that if $n \gg 0$, G' contains a copy of C_{2k}, say $v_1 v_2 \cdots v_{2k} v_1$. Letting $W = V' \setminus \{v_1, \ldots, v_k\}$, we must have that $A_1 = N(v_1) \cap W$ and $A_2 = N(v_2) \cap W$ are disjoint. Since G' is C_{2k+1}-free, the subgraph induced on A_1 (or A_2) is P_{2k}-free. By Theorem 1.7, $e(A_1)$ (or $e(A_2)$) is at most kn'. To summarize,

- $|A_1|, |A_2| \geq (1/2 - 2\sqrt{\delta})n' - 2k \geq (1/2 - \sqrt{\delta})n'$.
- $e(A_1), e(A_2) \leq kn'$.
- $A_1 \cap A_2 = \varnothing$.

Delete all edges that are not between A_1 and A_2 (namely those within the A_is and those that touch at least one vertex not in $A_1 \cup A_2$). This deletes at most

$$\underbrace{2kn'}_{e(A_i)} + \underbrace{2\sqrt{\delta}n'^2}_{|V' \setminus (A_1 \cup A_2)| \geq 2\sqrt{\delta}n'}.$$

The resulting graph is clearly bipartite. Thus, from G_n, we have deleted at most, say, $5\sqrt{\delta}n^2$ edges, so setting δ less than say $\varepsilon^2/100$ works. ■

Theorem 2.22. For $n \gg 0$,

$$\text{ex}(n; C_{2k+1}) = \left\lfloor \frac{n^2}{4} \right\rfloor.$$

Proof. Let G_n be C_{2k+1}-free and have the maximum possible number of edges. Since the complete bipartite graph is C_{2k+1}-free, $e(G_n) \geq \lfloor n^2/4 \rfloor$.
Let $\varepsilon > 0$. For starters, let us assume that $\delta(G_n) \geq (1/2 - 2\sqrt{\varepsilon})n$. By Lemma 2.21, G_n can be made bipartite by deleting at most εn^2 edges (due to the assumption on $\delta(G_n)$ and the previous proof). Let (A, B) be the parts of such a bipartite graph. Further assume that $e(A, B)$ is the maximum possible (we choose that bipartite subgraph with the most edges). We claim that $|A|$ and $|B|$ are at least $(1/2 - \varepsilon)n$. Indeed, otherwise, since $|A| + |B| = n$,

$$e(A, B) \leq |A||B| < \left(\frac{1}{2} - \sqrt{\varepsilon}\right)\left(\frac{1}{2} + \varepsilon\right)n^2 = \left(\frac{1}{4} - \varepsilon\right)n^2.$$

In this case, $e(G_n) \leq e(A, B) + \varepsilon n^2 < \lfloor n^2/4 \rfloor$, proving the claim.

Since $e(A, B)$ is the maximum possible, it follows that for any $a \in A$, $d(a, B) \leq d(a, A)$ (otherwise we can switch a from A to B).
Next, we claim that no vertex a of A is such that $d(a, A) \geq 2\sqrt{\varepsilon}n$. Suppose otherwise. We then have $d(a, B) \geq 2\sqrt{\varepsilon}n$ as well. Let $A_1 = A \cap \Gamma(a)$ and $B_1 = B \cap \Gamma(a)$. Consider the bipartite subgraph on (A_1, B_1). This subgraph is P_{2k}-free. By Theorem 1.7,

$$e(A_1, B_1) \leq 2kn.$$

In particular, the number of missing edges between A_1 and B_1 is at least $4\varepsilon n^2 - 2kn > \varepsilon n^2 \geq e(A) + e(B)$ for $n \gg 0$. As a result,

$$\begin{aligned} e(G_n) &= e(A, B) + e(A) + e(B) \\ &\leq e(A, B) + (\text{number of missing pairs between } A_1 \text{ and } B_1) \\ &< |A||B| \leq n^2/4, \end{aligned}$$

which gives a contradiction.

Let $a, a' \in A$ be adjacent. We have

$$\begin{aligned}|(\Gamma(a) \cap B) \cap (\Gamma(a') \cap B)| &\geq d(a, B) + d(a', B) - |B| \\ &> d(a) - 2\sqrt{\varepsilon}n + d(a') - 2\sqrt{\varepsilon}n - \left(\frac{1}{2} + \sqrt{\varepsilon}\right)n \\ &\geq 2\delta(G_n) - 4\sqrt{\varepsilon}n - \left(\frac{1}{2} + \sqrt{\varepsilon}\right)n \\ &\geq \left(\frac{1}{2} - 9\sqrt{\varepsilon}\right)n.\end{aligned}$$

Let $B' = \Gamma(a) \cap \Gamma(a') \cap B$ and $A' = A \setminus \{a, a'\}$ and consider the bipartite subgraph between these two parts. Since the previous subgraph was C_{2k+1}-free, there cannot be a P_{2k-1} between two vertices in B' (or A'). It follows that there does not exist a path of length $2k$ between *any* two vertices of $A' \cup B'$. So,

$$e(A', B') \leq 2kn.$$

Estimating the number of edges again,

$$\begin{aligned}e(G_n) &\leq e(A', B') + e(A \setminus A', V) + e(B \setminus B', V) \\ &= 2kn + \underbrace{2n}_{|A \setminus A'| = 2} + \underbrace{10\sqrt{\varepsilon}n}_{\substack{|B| \leq (1/2 + \sqrt{\varepsilon})n \\ |B'| \geq (1/2 - 9\sqrt{\varepsilon})n}} \\ &< \frac{n^2}{4} \text{ for } n \gg 0,\end{aligned}$$

giving a contradiction again.

Therefore, it now suffices to show why we can make our initial assumption that $\delta(G_n) \geq (1/2 - 2\sqrt{\varepsilon})n$. As in the proof of the previous lemma, repeatedly delete any vertex that has degree less than $(1/2 - 2\sqrt{\varepsilon})n$. We must stop at some G_ℓ with $\ell > (1 - 4\sqrt{\varepsilon})n$. Further, G_ℓ must have greater than $\ell^2/4$ edges and is C_{2k+1}-free, so we arrive at a contradiction anyway due to our proof taking the assumption at the beginning. ■

2.5. Pseudorandomness

Szemerédi's regularity lemma, in layman terms, says that any dense graph consists 'mostly' of a 'small' number of 'random-like' bipartite graphs. The meaning of the first two words within the quotes should be clear; they just say that $|V_0|$ is small and that k is bounded. The question then is: what does it mean to say that a graph is random-like? We have a single graph, not a distribution.

2.5.1. Notions of randomness

Let us first list a couple of features that feel like something a random-like graph should have. Suppose that G_n has $(p + o(1))\binom{n}{2}$ edges, where p is a fixed constant.

1. Disc (*Discrepancy*). For all $X, Y \subseteq V(G)$,
$$\left|e(X, Y) - p|X||Y|\right| = o(n^2).$$
2. Disc′. For all $X \subseteq V(G)$,
$$\left|e(X) - p\binom{|X|}{2}\right| = o(n^2).$$
3. Count. For a fixed H, the number of labelled copies of H in G is $(1 + o(1))p^{e(H)}n^{|V(H)|}$ (Why choose this number?).

4. Codegree. We have
$$\sum_{x,y\in V} |d(x,y) - p^2 n| = o(n^3).$$
5. Eigen. If $\lambda_1 \geq \cdots \geq \lambda_n$ are the eigenvalues of the adjacency matrix of G_n, then
$$\lambda_1 = pn(1 + o(1))$$
and
$$\max_{i>1} |\lambda_i| = \max\{|\lambda_2|, |\lambda_n|\} = o(n).$$

Observe that there are already asymptotics at play here since $o(\cdot)$ is not very meaningful when we are dealing with a single graph.

While the reader is likely familiar with what an adjacency matrix is, we define it here for the sake of completeness.

Definition 2.8 (Adjacency Matrix)**.** Given a graph $G = (V, E)$, the *adjacency matrix* A of G is a matrix whose rows and columns are indexed by V, and for $v, w \in V$
$$A_{vw} = \begin{cases} 1, & vw \in E, \\ 0, & vw \notin E. \end{cases}$$

The eigenvalues of A are sometimes referred to as the eigenvalues of G.

While the first four properties above might make sense based on our discussion thus far, the last seems to come out of nowhere.
To motivate Eigen, we give the following result.

Theorem 2.23 (Expander Mixing Lemma)**.** Suppose $G = (V, E)$ is a d-regular graph on n vertices and $\lambda = \max\{|\lambda_2|, |\lambda_n|\}$, where $\lambda_1 \geq \cdots \geq \lambda_n$ are the eigenvalues of G. Then for any $X, Y \subseteq V$,
$$\left| e(X,Y) - \frac{d}{n}|X||Y| \right| \leq \lambda\sqrt{|X||Y|}.$$

Proof. Observe that
$$e(X,Y) = \mathbb{1}_X^\top A \mathbb{1}_Y,$$
where $\mathbb{1}_X$ is the 'indicator vector' of X, a vector indiced by the vertices of G, with 1 at the positions in X and 0 elsewhere.
Let A be the adjacency matrix of G. Since A is real symmetric, we can apply the spectral theorem to write $A = \sum_i \lambda_i v_i v_i^\top$, where v_i is an eigenvector for λ_i.
Observe that d is an eigenvalue of A with corresponding eigenvector $\mathbf{1}$, the all 1s vector. In fact, d is the largest eigenvalue – if λ_1 is the largest eigenvalue and x is an eigenvector with $x_v = \max_u |x_u|$,
$$|\lambda_1 x_v| = |(Ax)_v| = \left| \sum_{u\in\Gamma(v)} x_u \right| \leq d(v)|x_v| \leq \Delta(G)|x_v|.$$

Letting J be the all 1s matrix of appropriate size,
$$\begin{aligned} \left| e(X,Y) - \frac{d}{n}|X||Y| \right| &= \left| \mathbb{1}_X^\top A \mathbb{1}_Y - \frac{d}{n}\mathbb{1}_X J \mathbb{1}_Y \right| \\ &= \left| \mathbb{1}_X^\top \left(A - \frac{d}{n}J \right) \mathbb{1}_Y \right| \\ &\leq \left\| \mathbb{1}_X^\top \right\| \left\| A - \frac{d}{n}J \right\| \left\| \mathbb{1}_Y \right\|. \end{aligned}$$

We claim that $\|A-(d/n)J\| \leq \lambda$. Since A is symmetric with all row sums as d, it commutes with J, and so they can be simultaneously diagonalized. Further, n is an eigenvalue of J and d is an eigenvalue of A with eigenvector $\mathbf{1}$. ■

Next, the following result says that all five notions of randomness we came up with are equivalent!

Theorem 2.24 (Chung-Graham-Wilson Theorem)**.** Disc, Disc$'$, Count, Codegree, and Eigen are equivalent.

Proof. To prove this we shall show that all of our 5 are equivalent to the following.
C_4-Count. The number of labelled C_4s is $(p^4+o(1))n^4$ (this is just Count for $H = C_4$ specifically).

- Disc to Disc$'$. Setting $X = Y$ then dividing by 2 gives the result directly.
- Disc$'$ to Disc. Let $X, Y \subseteq V$. We have

$$e(X,Y) = e(X \cup Y) + e(X \cap Y) - e(X \setminus Y) - e(Y \setminus X)$$

  Applying Disc$'$ to each of these 4 sets gives the results near directly:

$$\begin{aligned} e(X,Y) &= p\binom{|X\cup Y|}{2} + p\binom{|X\cap Y|}{2} - p\binom{|X\setminus Y|}{2} - p\binom{|Y\setminus X|}{2} + o(n^2) \\ &= p|X||Y| + o(n^2). \end{aligned}$$

- Disc to Count. This is just the Graph Counting Lemma (we only use the Disc part of regularity in the proof).
- Count to C_4-Count. This is clear from the definition.
- C_4-Count to Codegree. Now,

$$\begin{aligned} \sum_{x,y\in V} d(x,y) &= \sum_{x\in V} d(x)^2 \\ &\geq \frac{1}{n}\left(\sum_{x\in V} d(x)\right)^2 \\ &= \frac{4e(G)^2}{n} = (1+o(1))p^2n^3. \end{aligned}$$

  Also note that

$$\sum_{x,y\in V} d(x,y)^2 = \text{number of labelled } C_4\text{s} + o(n^4) = (1+o(1))p^4n^4.$$

  So,

$$\begin{aligned} \sum_{x,y\in V} |d(x,y)-p^2n| &\leq n\left(\sum_{x,y\in V}(d(x,y)-p^2n)^2\right)^{1/2} \\ &= n\left(\sum_{x,y\in V} d(x,y)^2 - 2p^2n\sum_{x,y\in V} d(x,y) + p^4n^4\right)^{1/2} \\ &\leq n\left((1+o(1))p^4n^4 - 2p^2n(1+o(1))p^2n^3 + p^4n^2\right)^{1/2} \\ &= n(o(1)n^4)^{1/2} \\ &= o(n^3). \end{aligned}$$

- Codegree to Disc. We have

$$\begin{aligned}
\sum_{x \in V} |d(x) - pn| &\leq \sqrt{n} \left(\sum_{x \in V} (d(x) - pn)^2 \right)^{1/2} \\
&= \sqrt{n} \left(\sum_{x \in V} d^2(x) - 2pn \sum_{x \in V} d(x) + p^2 n^3 \right) \\
&= \sqrt{n} \left(\sum_{x,y \in V} d(x,y) - 2p^2 n^3 + p^2 n^3 o(n) \right) \\
&= \sqrt{n} \left(\sum_{x,y \in V} \left(d(x,y) - p^2 n \right) \right) \\
&\leq \sqrt{n} \left(\sum_{x,y \in V} \left| d(x,y) - p^2 n \right| \right) = o(n^2).
\end{aligned}$$

For $X, Y \subseteq V$,

$$\begin{aligned}
\left| e(X,Y) - p|X||Y| \right| &\leq \sum_{x \in X} \left| d(x,y) - p|Y| \right| \\
&\leq \sqrt{|X|} \left(\sum_{x \in X} (d(x,Y) - p|Y|)^2 \right)^{1/2} \\
&\leq \sqrt{|X|} \left(\sum_{v \in V} (d(v,Y) - p|Y|)^2 \right)^{1/2} \\
&= \sqrt{|X|} \left(\sum_{v \in V} d(v,Y)^2 - 2p|Y| \sum_{v \in V} d(v,Y) + p^2 |Y|^2 n \right)^{1/2} \\
&= \sqrt{|X|} \left(\sum_{y,y' \in Y} d(y,y')^2 - 2p|Y| \sum_{y \in Y} d(y) + p^2 |Y|^2 n \right)^{1/2} \\
&= \sqrt{|X|}
\end{aligned}$$

- Eigen to C_4-Count. Note that the number of labelled C_4s in G equals the number of closed walks of length 4 minus $O(n^3)$. Now, observe that given two vertices v, w, the number of walks from v to w of length 4 is just $(A^4)_{vw}$ (Why?). In particular, the number of closed walks from v to v is $(A^4)_{vv}$. Recalling that the trace of a matrix is the sum of its eigenvalues,

$$\begin{aligned}
\text{number of } C_4\text{s} &= \operatorname{Tr}(A^4) - O(n^3) \\
&\leq \lambda_1^4 + \sum_{i>1} \lambda_i^4 \\
&\leq (1 + o(1))(pn)^4 + \left(\max_{i>1} \lambda_i^2 \right) \sum_{i=1}^{n} \lambda_i^2 \\
&= (1 + o(1))(pn)^4 + o(n^2) \cdot 2e(G) \\
&= (1 + o(1))(pn)^4 + o(n^4) = (1 + o(1)) p^4 n^4.
\end{aligned}$$

- C_4-Count to Eigen. The maximum eigenvalue of A is just

$$\lambda_1 = \sup_{\|x\|_2=1} x^\top Ax.$$

Taking $x = (1/\sqrt{n})\mathbf{1}$, we have

$$\lambda_1 \geq \frac{2e(G)}{n} = (p - o(1))n.$$

For the remaining eigenvalues,

$$\begin{aligned}\max_{i\neq 1} |\lambda_i|^4 &\leq \operatorname{Tr}(A^4) - \lambda_1^4 \\ &\leq \underbrace{(p^4 + o(1))n^4}_{\text{from } C_4-\text{Count}} - (p - o(1))^4 n^4 \\ &= o(n^4),\end{aligned}$$

completing the proof.

■

Remark.
- The proof requires $p = \Omega(1)$, in particular to show that Disc implies Count.
- Count does not hold in sparse regimes. In fact, Noga Alon in the 90s gave a graph with $p = \Omega(n^{-1/3})$ which is Eigen-pseudorandom, but triangle-free.

2.5.2. Strongly regular graphs

Do we have any 'known' families of pseudorandom graphs?

Definition 2.9. Suppose q is a prime power such that $q \equiv 1 \pmod 4$. Consider the **Paley graph** P_q with vertex set $V(P_q) = \mathbb{F}_q$ and xy is an edge iff $x - y$ is a quadratic residue[9].

Let us see a couple of properties of the Paley graph.

Definition 2.10. A graph G is said to be an (n, d, λ)**-graph** if it is a regular graph of degree d over n vertices with $\max_{i\neq 1} |\lambda_i| \leq \lambda$, where $\lambda_1 \geq \cdots \geq \lambda_n$ are the eigenvalues of the graph.

Definition 2.11. A graph $G = (V, E)$ is a (n, d, λ, μ)**-strongly regular graph** if

- $|V| = n$,
- G is regular with degree d,
- Any two adjacent vertices have exactly λ common neighbours, and
- Any two non-adjacent vertices have exactly μ common neighbours.

If a graph is (n, d, λ, μ)-strongly regular for some n, d, λ, μ, it is just said to be strongly regular.

---

[9] $x - y = a^2$ for some $a \in \mathbb{F}_q^\times$.

For example, a complete graph on n vertices is $(n, n-1, n-1, 0)$-strongly regular and C_5 is $(5, 2, 0, 1)$-strongly regular.
Slightly more complicatedly, the Petersen graph is $(10, 3, 0, 1)$-strongly regular.

Observe that a graph is (n, d, λ, μ)-strongly regular iff its adjacency matrix A is such that

$$(A^2)_{vw} = \begin{cases} d, & i = j, \\ \lambda, & v, w \text{ are adjacent}, \\ \mu, & v, w \text{ are not adjacent}. \end{cases}$$

More succinctly,

$$A^2 = dI + \lambda A + \mu(J - I - A).$$

We see that

$$\begin{aligned} A^3 &= dA + \lambda A^2 + \mu dJ - \mu(A + A^2) \\ 0 &= A^3 + (\mu - \lambda)A^2 - dA - \mu dJ. \end{aligned}$$

The graph is strongly regular iff its adjacency matrix satisfies a cubic with integer coefficients and leading coefficient 1!
This further implies that a strongly regular graph has only 3 eigenvalues, with one of them being d.

Returning to the Paley graph, it is not too difficult to check that it is $(q, (q-1)/2, (q-5)/4, (q-1)/4)$-strongly regular.
Consequently, it is a $\left(q, (q-1)/2, (\sqrt{q}+1)/2\right)$-graph.
Observe that any strongly regular graph has eigenvalues (other than λ_1) bounded in a similar fashion.

2.5.3. The sparse regime

Let us restrict ourselves to (n, d, λ)-graphs. A popular family of (n, d, λ)-graphs comes from the following.

Definition 2.12. Suppose Γ is a finite group and $S \subseteq \Gamma \setminus \{1\}$ is such that $S = S^{-1}$. We define the **Cayley graph** $\mathrm{Cay}(\Gamma, S)$ as that on vertex set Γ with $gh \in E(G)$ iff $g^{-1}h \in S$.

It is easy to see that $\mathrm{Cay}(\Gamma, S)$ is $|S|$-regular. In fact, P_q is just a Cayley graph with $\Gamma = \mathbb{F}_q$ (under addition) and S is the set of quadratic residues in $\mathbb{F}_q$ other than 1.

Theorem 2.25 (Conlon-Zhao, 2017)**.** For Cayley graphs, Disc and Eigen are equivalent.

Proof. First, note that Eigen implies Disc with the same constant for both. That is, if

$$\max_{i \neq 1} |\lambda_i| \leq \varepsilon n.$$

then

$$\left|e(X, Y) - p|X||Y|\right| \leq \varepsilon n^2$$

by the Expander Mixing Lemma.
To show the other side, we require a result of Grothendieck's:

There exists an absolute constant $K > 0$ such that for all real matrices $A = (a_{ij})$ and a Hilbert space $\mathcal{H}$,

$$\sup_{\substack{x_i, y_j \in \mathcal{H} \\ \|x\|, \|y\| = 1}} \sum_{ij} a_{ij} \langle x_i, y_j \rangle \leq K \sup_{\{u,v\} \in \{-1,1\}^n} \sum_{i,j} a_{ij} u_i v_j.$$

The best possible K is unknown, but we do know that $K < 1.79$ is possible. We skip the proof of this.
Suppose that Disc holds with a constant of ε and let A be the adjacency matrix of $G = \text{Cay}(\Gamma, S)$. Let $u, v \in \{-1, 1\}^n$. Denote by u^+ and u^- the positive and negative parts of u respectively. Letting $U^+ \subseteq V$ be the support of u^+ (the subset of vertices with non-zero components) then using Grothendieck's lemma,

$$(u^\pm)^\top \left(A - \frac{d}{n}J\right) v^\pm = e(U^\pm, V^\pm) - \frac{d}{n}|U^\pm||V^\pm|,$$

where the sign over the Us (resp. Vs) are the same as that over u (resp. v). Using Disc,

$$u^\top \left(A - \frac{d}{n}J\right) v \leq 4(\varepsilon d)n.$$

Now, since

$$\max_{i \neq 1} |\lambda_i| = \sup_{\|x\|=\|y\|=1} x^\top \left(A - \frac{d}{n}J\right) y,$$

it suffices to show that the quantity on the right is 'small'.
Let $B = A - (d/n)J$. We have

$$\begin{aligned}
x^\top B y &= \sum_{g,h \in \Gamma} x_g B_{g,h} y_h \\
&= \frac{1}{n} \sum_{s,g,h \in \Gamma} x_{sg} B_{sg,sh} y_{sh} \\
&= \frac{1}{n} \sum_{s,g,h \in \Gamma} x_{sg} B_{g,h} y_{sh} && (\Gamma \text{ is an automorphism group of } G) \\
&= \frac{1}{n} \sum_{g,h \in \Gamma} \left(B_{g,h} \sum_{s \in \Gamma} x_{sg} y_{sh} \right) \\
&= \frac{1}{n} \sum_{g,h \in \Gamma} B_{g,h} \langle x^{(g)}, y^{(h)} \rangle && \text{where } x^{(g)}_k = x_{kg} \\
&\leq \frac{1}{n} \sup_{\|x\|=\|y\|=1} \sum_{g,h \in \Gamma} B_{g,h} \langle x^{(g)}, y^{(h)} \rangle && \left(\left\|x^{(g)}\right\| = \|x\| = 1 \text{ and } \left\|y^{(h)}\right\| = \|y\| = 1\right) \\
&\leq \frac{1}{n} \cdot K \cdot (4\varepsilon d n) && (\text{using Grothendieck's result and the above result}) \\
&< 8\varepsilon d,
\end{aligned}$$

using the fact that $\left\|x^{(g)}\right\|^2 = \|x\|^2$, completing the proof. ■

Theorem 2.26. If G is d-regular, then $\lambda = \max\{|\lambda_2(G)|, |\lambda_n(G)|\}$ is at least $\sqrt{d}(1 - o_n(1))$.

Proof. Let A be the adjacency matrix of G. Observe that

$$\begin{aligned}
nd &= \text{Tr}(A^2) \\
&= \sum_i \lambda_i^2 \\
&\leq d^2 + (n-1)\lambda^2,
\end{aligned}$$

so

$$\lambda \geq \sqrt{d}(1 - o_n(1)).$$

■

However, we can get a better bound.

Theorem 2.27 (Alon-Boppana bound)**.** If G is a connected d-regular graph, then $\lambda(G) = \max\{|\lambda_2(G)|, |\lambda_n(G)|\}$ is at least $2\sqrt{d-1}(1 - o_n(1))$.

Proof. To prove the result, we shall get a vector $x \perp \mathbf{1}$ such that

$$\frac{x^\top Ax}{x^\top x} \geq 2\sqrt{d-1}(1 - o_n(1)).$$

Fix a vertex v and let $r \in \mathbb{N}$ (which we shall fix later). For $u \in V(G)$, consider the vector x defined by

$$x_u = \begin{cases} (d-1)^{-i/2}, & d(u,v) = i \leq r-1, \\ 0, & \text{otherwise,} \end{cases}$$

where $d(u, v)$ is the distance between the vertices u and v. Let $V_i = \{u : d(u, v) = i\}$. We have

$$x^\top x = \sum_{i=0}^{r-1} \frac{|V_i|}{(d-1)^i}$$

and

$$\begin{aligned}
x^\top Ax &= \sum_{u \in V} x_u \sum_{w \in \Gamma(u)} x_w \\
&\geq \sum_{i=0}^{r-2} |V_i| \cdot \frac{1}{(d-1)^{i/2}} \left(\frac{1}{(d-1)^{(i-1)/2}} + \frac{(d-1)}{(d-1)^{(i+1)/2}} \right) + \frac{|V_{r-1}|}{(d-1)^{(r-1)/2}} \cdot \frac{1}{(d-1)^{r/2}} \\
&\qquad \text{(if } u \in V_i\text{, then } \Gamma(u) \subseteq V_{i-1} \cup V_{i+1}\text{, and } u \text{ must have a neighbour in } V_{i-1}) \\
&\geq \sum_{i=0}^{r-2} |V_i| \cdot \frac{2\sqrt{d-1}}{(d-1)^i} + \frac{|V_{r-1}|(d-1)}{(d-1)^{(2r-1)/2}} \\
&\geq 2\sqrt{d-1} \left(\sum_{i=0}^{r-2} \frac{|V_i|}{(d-1)^i} + \frac{1}{2} \frac{|V_{r-1}|}{(d-1)^r} \right) \\
&\geq 2\sqrt{d-1} \left(1 - \frac{1}{2r} \right) \sum_{i=0}^{r-2} \frac{|V_i|}{(d-1)^i}. \qquad (|V_i|/(d-1)^i \leq |V_{i-1}|/(d-1)^{i-1})
\end{aligned}$$

So,

$$\frac{x^\top Ax}{x^\top x} \geq 2\sqrt{d-1} \left(1 - \frac{1}{2r} \right).$$

However, x is not orthogonal to $\mathbf{1}$, so this does not directly yield a bound on $\lambda(G)$.
Now note that if $n > 1 + (d-1) + (d-1)^2 + \cdots + (d-1)^{k-1}$, hen there must be two vertices that are at a distance of at least k. Equivalently,

$$n > \frac{(d-1)^k - 1}{d-2},$$

which gives a lower bound on k that is $\Omega(\log n)$.
Let v_1, v_2 be two farthest vertices in G, and let r such that $d(v_1, v_2) \geq 2r$. Let the vector x_1 be that as defined above with $v = v_1$ with

$$\frac{x_1^\top Ax_1}{x_1^\top x_1} \geq 2\sqrt{d-1} \left(1 - \frac{1}{2r} \right).$$

Similarly get x_2 using v_2. Note that x_1 and x_2 must be orthogonal. Let c be a constant such that $y = (x_1 + cx_2)$ is orthogonal to **1** (Why does such a c exist?). We then have

$$y^\top y = \|x_1\|^2 + c^2 \|x_2\|^2$$

and

$$y^\top Ay = x_1^\top Ax_1 + c^2 x_2^\top Ax_2 \geq 2\sqrt{d-1}\left(1 - \frac{1}{2r}\right)(\|x_1\|^2 + c^2 \|x_2\|^2),$$

completing the proof. ■

This leads to the following definition, which we shall not get into.

Definition 2.13. An (n, d, λ)-graph is called a **Ramanujan graph** if $\lambda \leq 2\sqrt{d-1}$.

This notion is due to Lubotzky, Phillips, and Sarnak, who constructed what are called "LPS graphs", that are Cayley graphs on the group $\mathrm{PSL}(2, \mathbb{F}_q)$, that are $(q+1)$-regular and Ramanujan.

Alon conjectured (mayhaps surprisingly) that random regular graphs are Ramanujan with high probability. This was resolved by Joel Friedman in the early 2000s.
Questions on the existence of Ramanujan graphs (and explicit constructions) are largely open, and an active area of research in combinatorics and theoretical computer science.
For $\alpha > 0$, one can also define (n, d, α)-expanders, which are (n, d, λ)-graphs with $\lambda \leq \alpha d$.

§3. Matching Theory

Definition 3.1. Given a graph G, a **matching** $\mathcal{M}$ in G is a collection of pairwise disjoint edges in G (no two of the edges have a common vertex).

For example, $\{\{1,6\},\{2,5\},\{3,4\}\}$ is a matching in K_6.

Definition 3.2. Given a graph G, the **matching number** $\nu(G)$ of G is the size of a maximum matching in G.

Given a graph, how do we compute a maximum matching, or simpler yet, $\nu(G)$?

3.1. The Bipartite Setting

For this section, suppose that $G = (X, Y, E)$ is bipartite, where all edges in E are between X and Y. Assume $|X| \leq |Y|$. It is clear that $\nu(G) \leq |X|$.

The question is: under what conditions on G is $\nu(G) = |X|$?
This (the bipartite scenario) is sometimes referred to as a *system of distinct representatives* (SDR). That is, G has an SDR if $\nu(G) = |X|$.

3.1.1. Hall's Marriage Theorem

An obvious necessary condition is that given any $S \subseteq X$,

$$|\Gamma(S)| \geq |S|.$$

Otherwise, there cannot exist a matching since two vertices in S would be forced to map to the same element in Y.
This is known as *Hall's condition*.
It in fact turns out that Hall's condition is sufficient for the existence of a matching too!

Theorem 3.1 (Hall's Marriage Theorem)**.** Suppose $G = (X, Y, E)$ is bipartite with $|X| \leq |Y|$. Then $\nu(G) = |X|$ if and only if it satisfies Hall's condition.

Much like Turán's theorem, this has a large number of proofs.

Proof 1 of Hall's Marriage Theorem. We begin with a proof by induction over $|X|$. If $|X| = 1$, the theorem obviously holds.
Make the stronger assumption that for all $S \subseteq X$ of size $|X|-1$, $|\Gamma(S)| > |S|$, and $|\Gamma(X)| \geq |X|$. Pick $x \in X$ and pair it with an arbitrary neighbour $y \in Y$. Using the inductive hypothesis on the subgraph induced on $(X \setminus \{x\}) \cup (Y \setminus \{y\})$, we get a matching on G.
Now, let $S \subseteq X$ with $|\Gamma(S)| = |S| = |X| - 1$. Using the inductive hypothesis on the subgraph induced on $S \cup \Gamma(S)$, together with the edge from the (single) element in $X \setminus S$ to any element in $\Gamma(X) \setminus \Gamma(S)$, we get a matching on G. ■

An alternate way to prove this is by doing the first part of the proof with all non-empty $S \subseteq X$ (instead of only those of size $|X| - 1$). In the case where we have some non-empty $S \subseteq X$ with $|\Gamma(S)| = |S|$, we can use the inductive hypothesis twice to get two matchings $\mathcal{M}_1$ and $\mathcal{M}_2$ on the subgraphs induced by $S \cup \Gamma(S)$ and $(X \setminus S) \cup (Y \setminus \Gamma(S))$, then take the union of the two matchings to get a matching $\mathcal{M}$ on the original graph G.

A nice consequence of the above proof is that it allows us to extend Hall's Theorem to an infinite setting as well.

Definition 3.3. Given a matching $\mathcal{M}$ on a graph $G = (V, E)$, call a $v \in V$ **unsaturated** if no edge of $\mathcal{M}$ is incident on v.

For a matching $\mathcal{M}$, $x \in X$, $y \in Y$ such that x, y are unsaturated (with respect to $\mathcal{M}$), a path P from x to y is called $\mathcal{M}$-**augmenting** if every alternate edge in it is a $\mathcal{M}$-edge.
Observe that if $\mathcal{M}$ admits an augmenting path, then $\mathcal{M}$ is not maximal. Indeed, we can 'flip' the edges in the path to show this. That is, given an augmenting path $\mathcal{P}$, where every $y_i x_i$ is in $\mathcal{M}$, it is not too difficult to show that $\mathcal{M} \triangle \mathcal{P}$ is a matching that has strictly more edges.

Proof 2 of Hall's Marriage Theorem. Let $\mathcal{M}$ be a maximum matching of a bipartite graph $G = (X, Y, E)$ that satisfies Hall's condition. Suppose there exists $x_0 \in X$ such that x_0 is unsaturated by $\mathcal{M}$. Let y_1 be a neighbour of x_0 (such a y_1 exists by Hall's condition).
If y_1 is unsaturated by $\mathcal{M}$, then $x_0 y_1$ is a $\mathcal{M}$-augmenting path, contradicting its maximality. So, suppose $x_1 y_1 \in \mathcal{M}$.
In general, if we have $y_1, \ldots, y_r, x_1, \ldots, x_r$, we can pick $y_{r+1} \in \Gamma(\{x_0, \ldots, x_r\})$ (distinct from all the y_i for $1 \leq i \leq r$).
If y_{r+1} is unsaturated, we have a $\mathcal{M}$-augmenting path from x_0 to y_{r+1}, contradicting the maximality of $\mathcal{M}$.
If it is not, we can jump back to the $x_{r+1} \in X$ (distinct from the x_i for $1 \leq i \leq r$) such that $y_{r+1} x_{r+1} \in E$, then continue the process. This must terminate at some point since the number of vertices in Y incident on some element of $\mathcal{M}$ is at most $|X \setminus \{x_0\}| < |Y|$, completing the proof. ∎

3.1.2. Some applications

Next, let us look at a few applications of Hall's Theorem.

Definition 3.4 (Perfect Matching)**.** A matching $\mathcal{M}$ in a graph $G = (V, E)$ is said to be a **perfect matching** if $|\mathcal{M}| = |V|/2$, that is, every vertex in V is saturated by $\mathcal{M}$.

Theorem 3.2. If a bipartite graph $G = (X, Y, E)$ is $d(> 0)$-regular, it has a perfect matching.

Proof. It suffices to check that Hall's condition holds. Let $S \subseteq X$. We clearly have $e(S, \Gamma(S)) = d|S|$. However, the total number of edges coming into $\Gamma(S)$ must include all the edges coming from S, that is, $S \subseteq \Gamma(\Gamma(S))$. So,

$$\begin{aligned} d|\Gamma(S)| &= e(\Gamma(S), \Gamma(\Gamma(S))) \\ &\geq e(\Gamma(S), S) \\ &= d|S|, \end{aligned}$$

completing the proof. ∎

For the next result, consider the following definition.

Definition 3.5. A matrix $A \in \mathbb{R}^{n \times n}$ is said to be **doubly stochastic**[10] if for all i, j,

$$0 \leq a_{ij} \leq 1 \text{ and } \sum_k a_{ik} = \sum_k a_{kj} = 1.$$

[10] These are interesting to study since they arise as the transition matrices of a certain class of discrete Markov chains. The stationary distribution of a discrete Markov chain with a doubly stochastic transition matrix is the uniform distribution.

It is not too difficult to see that any permutation matrix is doubly stochastic.
Suppose A_1, A_2 are doubly stochastic and $0 \leq \lambda \leq 1$. Then observe that $\lambda A_1 + (1-\lambda)A_2$ is doubly stochastic too (Why?).
As a result, a convex combination of doubly stochastic matrices is doubly stochastic.

Theorem 3.3 (Birkhoff-von Neumann Theorem)**.** A matrix is doubly stochastic iff it is a convex combination of permutation matrices.

Proof. We want to show that if A is doubly stochastic, then $A = \sum_{i=1}^{k} \lambda_i P_i$, where the P_i are permutation matrices, $\sum_i \lambda_i = 1$, and $\lambda_i \geq 0$.
The idea is something like performing induction on k. If A is doubly stochastic, P is a permutation matrix, and $0 < \lambda < 1$ such that every entry of $B = A - \lambda P$ is non-negative, then every row sum and column sum of B is equal to $1 - \lambda$. So, if we can find such λ and P with λ chosen 'maximally' (in the sense that if we increase it any more, some entry will become negative), we can ensure that B has more zero elements than A. This then allows us to induct on the number of non-zero entries in A.
The base case is when A has exactly n non-zero entries and in this case, A is equal to a permutation matrix (Why?).
Given the earlier claim about the existence of λ and P, the result follows-near directly on applying the inductive hypothesis to $(1/(1-\lambda))B$.
To prove the claim, consider the bipartite graph $G = (X, Y, E)$ with X and Y being the rows and columns of B respectively, and $xy \in E(G)$ iff $B_{xy} > 0$. If the graph has a perfect matching, then there exist λ, P of the required form – set P as the permutation matrix that has 1s at the edges corresponding to the matching, and let λ be equal to

$$\min\{b_{xy} : xy \text{ is an edge in the matching}\}.$$

To show that the graph admits a perfect matching, we check Hall's condition. For all $S \subseteq X$, we want to show that $|N(S)| \geq |S|$. Now, by definition, the submatrix corresponding to the rows indexed by S and columns indexed by $Y \setminus S$ is then the zero submatrix. Using the row sum condition,

$$\sum_{\substack{x \in S \\ y \in N(s)}} b_{xy} = \alpha|S|.$$

Since B is non-negative, this sum is at most

$$\sum_{\substack{x \in X \\ y \in N(S)}} b_{xy} = \alpha|N(S)|$$

by the column sum condition. The claim follows. ∎

Now, let us look for a moment at matchings in graphs in general (that need not be bipartite).

Theorem 3.4. A matching $\mathcal{M}$ in a graph G is maximum iff it admits no $\mathcal{M}$-augmenting path.

Proof. The proof we gave earlier to show that admitting a $\mathcal{M}$-augmenting implies non-maximality works out even in the non-bipartite scenario.
For the converse, suppose that $\mathcal{M}$ is a matching that admits no $\mathcal{M}$-augmenting path, and that $\mathcal{N}$ is a matching with $|\mathcal{N}| > |\mathcal{M}|$. Consider the subgraph H of G that retains only those edges in $\mathcal{M}$ or $\mathcal{N}$, but not both. Note that for any vertex of G, $d_H(v) \leq 2$.
It is not too difficult to show that H is a disjoint union of cycles and paths. Since the $\mathcal{M}$ edges and $\mathcal{N}$ edges are individually pairwise disjoint, there cannot be an odd cycle in H. Since $\mathcal{N} > \mathcal{M}$, there must be a path that contains more $\mathcal{N}$ edges than $\mathcal{M}$ edges, and this is just an augmenting path(!), thus completing the proof. ∎

The above proposition allows us to find maximum matchings using a polynomial time algorithm.
We give such an algorithm in the bipartite scenario.

Definition 3.6. A **directed graph** is a pair $D = (V, E)$, where V is the vertex set of D, and $E \subseteq V \times V$ is the set of *directed edges* or *arcs*. An edge $e = (u, v)$ is said to be directed from u to v.

We often abuse notation and denote an edge (u, v) by $\overrightarrow{uv}$ or uv.
A **loop** is an edge of the form (v, v).
We can define several things here similar to in ordinary simple graphs. For example, a *directed path* is a sequence $v_1, \ldots, v_n$ of vertices such that (v_i, v_{i+1}) is an edge for all valid i.

Theorem 3.5. Suppose $G = (X, Y, E)$ is bipartite with $|X| \leq |Y|$. There exists a polynomial time algorithm to find a maximum matching in G.

Proof. The idea of the proof is to construct augmenting paths. Given an augmenting path P for a matching $\mathcal{M}$, $\mathcal{M} \triangle P$ is a matching with strictly more edges. If we begin with an arbitrary matching and perform this operation repeatedly, we must terminate within $|X|/2$ iterations.
Let $\mathcal{M}$ be a matching. Our objective is to either produce an augmenting path, or show that there exists no $\mathcal{M}$-augmenting path. Let $X_{\mathcal{M}}$ and $Y_{\mathcal{M}}$ be the vertices in X and Y respectively saturated by $\mathcal{M}$, and let $X_U = X \setminus X_{\mathcal{M}}$ and $Y_U = Y \setminus Y_{\mathcal{M}}$. Notice that an augmenting path (if it exists) is uniquely determined by its sequence in $X_{\mathcal{M}}$ (since the edges from $X_{\mathcal{M}}$ to $Y_{\mathcal{M}}$ are fixed).
Now, define the directed graph $D_{\mathcal{M}} = (X, \mathcal{E})$ with $(x_1, x_2) \in \mathcal{E}$ iff there exists $y \in Y_{\mathcal{M}}$ such that $x_2 y_2 \in \mathcal{M}$ and $x_1 y_2 \in E(G)$.
Let $X^* = N(Y_U)$. Observe that G has a $\mathcal{M}$-augmenting path iff $D_{\mathcal{M}}$ has a path from X_U to X^*. This is easily resolved by just finding the set of vertices 'reachable' from X_U in $D_{\mathcal{M}}$. This is possible to do efficiently using the *breadth-first-search* (BFS) algorithm, which we shall detail next. ■

The BFS algorithm is as follows.

Algorithm 1: BFS algorithm

Input: A directed graph $D = (V, E)$ and a set $X \subseteq V$.
Output: The set of vertices in D reachable from X

- $S \leftarrow X$
- Queue $q \leftarrow X$ `// in any arbitrary order`
- **while** $q \neq \varnothing$ **do**
    - $v \leftarrow$ pop(q)
    - **for** $u \in N(v)$ **do**
        - **if** $u \notin S$ **then**
            - $S \leftarrow S \cup \{u\}$
            - push(q, u)
- **return** S

If $X = \{v\}$, then this algorithm in fact finds a shortest path from v to every reachable vertex from it. The correctness of the algorithm follows by a simple induction on the distance from v to u.

It is not too difficult to see that the BFS algorithm runs in $O(|V| + |E|)$, where m is the number of edges in the graph, since each edge is checked at most twice. So, BFS runs linearly in the input size.

3.2. Flow

3.2.1. Flows and Cuts

Definition 3.7 (Network)**.** A **network** is a tuple (D, s, t, C), such that

- D is a directed graph,
- s and t are (distinct) specified vertices known as the **source** and **sink** respectively, and
- $C : E(D) \to \mathbb{R}_{\geq 0}$ is known as the **capacity function**.

Given an edge e, $C(e)$ is referred to as the capacity of the edge e.

This has a very simple-to-understand physical meaning. We can think of s as a source of water and t as a reservoir, with the intermediate edges being pipes that have a limit on how much they can carry and the intermediate nodes being junctions that transmit water. The main question we wish to answer then is: what is the maximum rate of water flow that can be sent from the source to the sink without violating any capacity constraints?
More concretely,

Definition 3.8 (Flow)**.** Given a network $\mathcal{N} = (D, s, t, C)$, a **flow** is a function $f : E(D) \to \mathbb{R}_{\geq 0}$ such that

- For all $e \in E(D)$, $f(e) \leq C(e)$,
- For every $v \neq s, t$,
$$\sum_{v \in V : uv \in E} f(uv) = \sum_{u \in V : vu \in E} f(vu).$$
This is known as **Kirchhoff's Law**.

Given a flow f, we define the value $\mathrm{Val}(f)$ of the flow by

$$\mathrm{Val}(f) = \sum_{u \in V : su \in E} f(su) - \sum_{u \in V : us \in E} f(us).$$

Given $A, B \subseteq V$, denote
$$f(A, B) = \sum_{\substack{e = uv \in E \\ u \in A, v \in B}} f(e) - \sum_{\substack{e = vu \in E \\ u \in A, v \in B}} f(e).$$

Definition 3.9 (Cut)**.** Given a network (D, s, t, C), a **cut** is the ordered pair $(S, \overline{S})$ for some $S \subseteq V$ with $s \in S$ and $t \notin S$. We further define the **capacity** of this cut by

$$C(S, \overline{S}) = \sum_{\substack{e = uv \in E \\ u \in S, v \in \overline{S}}} f(e).$$

Since Kirchhoff's law is satisfied, we get the following.

Theorem 3.6. For all $S \subseteq V$ such that $s \in S$ and $t \notin S$,

$$\mathrm{Val}(f) = f(S, \overline{S}).$$

Note that the claim follows by definition for $S = \{s\}$.

Proof. For $v \in V$ and $e \in E$, let

$$\phi(v,e) = \begin{cases} 1, & e = vu \text{ for some } u \in V, \\ -1, & e = uv \text{ for some } u \in V, \\ 0, & e \text{ is not incident on } v. \end{cases}$$

Note that Kirchhoff's law then just says that for any $v \in V \setminus \{s,t\}$,

$$\sum_{e \in E} \phi(v,e) f(e) = 0.$$

Denote by $E(S, \overline{S})$ the set of all edges between S and $\overline{S}$ (in either direction).
We have

$$\begin{aligned} f(S,\overline{S}) &= \sum_{\substack{e=uv \in E \\ u \in S, v \in \overline{S}}} f(e) - \sum_{\substack{e=vu \in E \\ u \in S, v \in \overline{S}}} f(e) \\ &= \sum_{e \in E(S,\overline{S})} \sum_{v \in S} f(e)\phi(v,e) \\ &= \sum_{v \in S} \sum_{e \in E(S,\overline{S})} f(e)\phi(v,e) \\ &= \sum_{v \in S} \sum_{e \in E} f(e)\phi(v,e) - sum_{v \in S} \sum_{\substack{e \in E(S,S): \\ \phi(v,e)=1}} f(e) + sum_{v \in S} \sum_{\substack{e \in E(S,S): \\ \phi(v,e)=-1}} f(e) \\ &= \sum_{v \in S} \sum_{e \in E} f(e)\phi(v,e), \end{aligned}$$

where the last step follows since for any edge uv in $E(S,S)$, $\phi(u,e) = 1$, $\phi(v,e) = -1$, and $\phi(w,e) = 0$ for all other vertices w.
By Kirchhoff's law,

$$\sum_{v \in S} \sum_{e \in E} f(e)\phi(v,e) = \sum_{e \in E} f(e)\phi(s,e) = \mathrm{Val}(f),$$

completing the proof. ∎

Corollary 3.7. For a flow f,

$$\mathrm{Val}(f) = \sum_{u \in V : ut \in E} f(ut) - \sum_{u \in V : tu \in E} f(tu).$$

The above is sometimes taken as an alternate definition for the value of a flow.

3.2.2. The Max-Flow Min-Cut Theorem

Corollary 3.8. For a network $\mathcal{N}$, a flow f, and any cut $(S, \overline{S})$,

$$\mathrm{Val}(f) \leq C(S, \overline{S}).$$

Observe that for any network $\mathcal{N}$, there exists at least one flow (namely the zero flow), and the flow value is bounded above.

Definition 3.10. Given a network $\mathcal{N}$, a **maximum flow** is a flow such that $\mathrm{Val}(f)$ is the maximum among all possible flows.

Theorem 3.9. Any network $\mathcal{N}$ on a finite directed graph has a maximum flow.

Proof. It is clear that

$$\sup\{\mathrm{Val}(f) : f \text{ is a flow on } \mathcal{N}\}$$

is well-defined and finite since for any flow f

$$\mathrm{Val}(f) \leq \min_{(S,\overline{S}) \text{ is a cut}} C(S,\overline{S}). \tag{3.1}$$

It is easy to show that the supremum is attained by a flow by considering a sequence of flows that converge (in value) to the supremum flow value, considering an appropriate subsequence such that the flow in each edge converges (this is possible since the set $[0, C(e)]$ is compact for each edge e, so we can apply the Bolzano-Weierstrass Theorem), and considering the flow whose value in each edge is equal to the limit of the sequence of flow values in this edge. Since the capacity constraint is a weak one and not strong, this is indeed a flow that satisfies the capacity constraints. ■

It in fact turns out that the bound in (3.1) is attained!

Theorem 3.10 (Ford-Fulkerson Theorem)**.** For a network $\mathcal{N} = (D, s, t, C)$ (where D is finite),

$$\max_{\substack{f \text{ is a flow} \\ \text{on } \mathcal{N}}} \mathrm{Val}(f) = \min_{(S,\overline{S}) \text{ is a cut}} C(S,\overline{S}).$$

The above is also referred to as the *max-flow min-cut* theorem.

Proof. (3.1) shows that the term on the right is at least the quantity on the right. Therefore, it suffices to show that if f is a max-flow, there exists a cut $(S,\overline{S})$ such that $\mathrm{Val}(f) = C(S,\overline{S})$.
Now, suppose there is a sequence $s = v_0 v_1 v_2 \cdots v_k v_{k+1} = t$ of vertices such that

- for each i, $v_i v_{i+1} \in E$ or $v_{i+1} v_i \in E$. In the former case, we call it a *forward edge* and in the latter, a *backward edge*.
- if $v_i v_{i+1}$ is a forward edge, then $f(v_i v_{i+1}) < C(v_i, v_{i+1})$.
- if $v_{i+1} v_i$ is a backward edge, then $f(v_{i+1}, v_i) > 0$.

We call such a sequence an f**-augmenting path**.
Suppose for a moment that such a path exists. Then, let

$$\delta = \min\left(\{C(e) - f(e) : e \text{ is a forward edge in } P\} \cup \{f(e) : e \text{ is a backward edge in } P\}\right)$$

By definition, $\delta > 0$. Define a new function $f' : E \to \mathbb{R}$ by

$$f'(e) = \begin{cases} f(e), & e \notin P, \\ f(e) + \delta, & e \in P \text{ is a forward edge}, \\ f(e) - \delta, & e \in P \text{ is a forward edge}. \end{cases}$$

It is not too difficult to check that f' is a flow as well. Further,

$$\mathrm{Val}(f') = \mathrm{Val}(f) + \delta > \mathrm{Val}(f).$$

Therefore, if f is a maximum flow, there is no f-augmenting path.
Now, let f be a maximum flow. For each vertex v, we say that a path from s to v is f-augmenting if the earlier definition holds, except that the final vertex is v instead of t.
Let

$$S = \{v : \text{ there is an } f\text{-augmenting path from } s \text{ to } v\}.$$

Clearly, $s \in S$. Since f is a maximal flow, $t \notin S$. Therefore, $(S, \overline{S})$ is a cut.
Suppose $e = uv$ is an edge with $u \in S$ and $v \notin S$. If $f(e) < C(e)$, then $v \in S$ as well due to the existence of a path $su_1 \cdots u_k uv$ (since u is f-augmenting), leading to a contradiction. So, $f(e) = C(e)$.
On the other hand, if $e = vu$ is an edge with $u \in S$ and $v \notin S$ and $f(e) > 0$, $v \in S$, again leading to a contradiction. So, $f(e) = 0$.
However, we then have

$$\text{Val}(f) = f(S, \overline{S}) = C(S, \overline{S}),$$

completing the proof. ■

The above proof is easily modified to an algorithm in the case where C is integral.

Corollary 3.11. Suppose $\mathcal{N}$ is an integral network. That is, $C(e) \in \mathbb{N}_0$ for all $e \in E(D)$. Then, there is an integral maximum flow f on $\mathcal{N}$ (a flow which has maximum value and only takes integral values).

Proof. Begin with the zero flow f_0. Given f_k, choose an f_k-augmenting path, and increase the value of the flow by δ to get f_{k+1}. If there exists no f_k-augmenting path, then we must have attained the maximum flow possible since we must have saturated some cut. Further, due to the integral nature of the network and the flow at each step, δ is a positive integer, so this process terminates after a finite number of time steps. ■

3.2.3. Applications

Here, we give two applications of flow.

Consider a bipartite graph $G = (X, Y, E)$. Define the network $\mathcal{N} = (D, s, t, C)$ as follows:

- s and t are 'new' vertices (not in $X \cup Y$),
- $D = (X \cup Y, E_D)$, where

$$E_D = \underbrace{\{\overrightarrow{uv} : uv \in E\}}_{E_1} \cup \underbrace{\{\overrightarrow{su} : u \in X\}}_{E_2} \cup \underbrace{\{\overrightarrow{vt} : v \in Y\}}_{E_3},$$

- the capacity function is defined by

$$C(e) = \begin{cases} 1, & e \in E_1 \\ \infty, & \text{otherwise.} \end{cases}$$

  If ∞ is problematic to think about, we can just choose a quantity that is greater than $C(X, Y)$ (no flow can have value greater than this anyway).

If there is a matching $\mathcal{M}$ of size m on G, then there is a flow $f_{\mathcal{M}}$ of value m that is 1 on all edges in $\mathcal{M}$, and the flow in each edge of $E_2 \cup E_3$ is chosen such that Kirchhoff's law is satisfied.
Conversely, if f is a maximum integral flow of value m, we can get a matching $\mathcal{M}$ with m edges by considering exactly those edges in E_1 that have flow 1 in them.
Therefore, to find a maximum matching on a bipartite graph, we may merely find a maximum flow in the corresponding network $\mathcal{N}$.

For the second application, we give a result of Schrijver's.

Suppose A is a $m \times n$ $\{0,1\}$-matrix that has k 1s in each row and r 1s in each column (this implies that $mk = nr$). Let $k' \leq k, r' \leq r$ such that $\alpha := k'/k = r'/r$.
Then, one can change some of the 1s in A to 0s such that in the resulting array A', there are k' 1s in each row and r' 1s in each column.

To prove this, assign a bipartite graph $G = (X, Y, E)$ to A as follows. X is the rows of A, Y is the columns of A, and $xy \in E(G)$ (with $x \in X$, $y \in Y$) iff $A_{xy} = 1$. We have that $d(x) = k$ and $d(y) = r$ for all $x \in X, y \in Y$.
Craft a network $\mathcal{N}$ using G exactly as we did above in the first application, with

$$C(e) = \begin{cases} 1, & e = xy \text{ for } x \in X, y \in Y, \\ k, & e = sx \text{ for } x \in X, \\ r, & e = yt \text{ for } y \in Y. \end{cases}$$

Clearly, the flow f that saturates the capacity of every single edge is a maximum flow. Consider another capacity function C' by

$$C'(e) = \begin{cases} 1, & e = xy \text{ for } x \in X, y \in Y, \\ k', & e = sx \text{ for } x \in X, \\ r', & e = yt \text{ for } y \in Y. \end{cases}$$

Note that $f' = \alpha f$ is a maximum flow in $\mathcal{N}' = (D, s, t, C')$. Since $\mathcal{N}'$ is integral, there is an integral max flow f'' in it. Note that for each e of the form sx, $f'(e) = k' = C'(e)$, so we must have $f''(e) = k'$ as well. Similarly, $f''(e) = r'$ for every edge of the form yt. Changing a 1 in A to a 0 iff $f''(e) = 0$ gives the required.

3.2.4. Algorithms for finding max-flow

Now, how do we actually determine a max-flow?
A basic idea comes from the proof of Corollary 3.11.
To do so, let us rephrase the definition of an augmenting path into terms closer to that of Dijkstra's algorithm. Given a network $\mathcal{N} = (D, s, t, C)$ and a flow f on it, define the auxiliary *residual* graph D_f on the vertex set of D such that $e = (u, v) \in E(D_f)$ iff

- uv is an edge in D and $c_f(e) = C(e) - f(e) > 0$ or
- vu is an edge in D and $c_f(e) = f(e) > 0$.

If we find a path in this residual graph, then we get an augmenting path to the flow f (if one exists), and the corresponding value of δ by which we augment the flow is just equal to the maximum value of $c_f(e)$ along edges e in this path.

Algorithm 2: Ford and Fulkerson's Algorithm

for $e \in E(G)$ **do**
    $f(e) \leftarrow 0$
while there exists an s-t path P in G_f **do**
    $c_f(P) \leftarrow \min_{e \in P} c_f(e)$
    foreach $e = uv \in P$ **do**
        if e is a forward edge **then**
            $f(uv) \leftarrow f(uv) + c_f(P)$
        else
            $f(vu) \leftarrow f(vu) - c_f(P)$
return f

There are two important things to note:

- The choice of the path at each step affects how many augmentations we perform.
  To illustrate this, consider the graph on $\{s, u, v, t\}$ with edges su, sv, uv, ut, vt and capacities $2^{10}, 2^{10}, 1, 2^{10}, 2^{10}$ respectively. If we pick the augmenting paths sut and svt in the beginning, we immediately arrive at the max-flow value. However, if we pick the path from s to t containing uv at each step, we take a very large amount of time (2^{11} time steps) to converge since the flow only increases by 1 each augmentation.

- In the case where capacities are irrational and we choose an arbitrary path at each step, we need not even converge to the max-flow value. If the capacities are integral (or rational), this process terminates as seen in the proof of Corollary 3.11.

This algorithm (for an integral network) takes $O(F_{\text{max}} E)$, where F_{max} is the value of a max-flow in the graph.

Henceforth, we abuse notation and denote by E (resp. V) the size of the set E (resp. V).

The first of the two points above leads to the **Edmonds-Karp algorithm**. It is nearly the same as the Ford and Fulkerson algorithm, except that at each step we choose the shortest path from s to t in G_f.
Further, it resolves the second issue as well, with the algorithm working even in the irrational scenario.

Theorem 3.12. The Edmonds-Karp algorithm for finding a max-flow runs in $O(VE^2)$.

Note in particular that the runtime is independent of the value of the max-flow.

Proof. Since each augmentation takes $O(E)$ time, it suffices to show that the number of augmentations is $O(VE)$.
First, we claim that the distance $\delta_f(s, v)$ between s, v in the residual graph G_f monotonically increases for all $v \in V$. That is, if f' is obtained by augmenting f using a shortest path in the residual graph, $\delta_f(s, v) \le \delta_{f'}(s, v)$.
For now, suppose the claim is true. Suppose we have a flow f and corresponding residual graph G_f. If P is a shortest s-t augmenting path, define the *critical capacity* $c_f(P)$ of P as in Algorithm 2. Also call an edge e *critical* if $c_f(P) = c_f(e)$. We shall show that every edge $e = uv$ in D becomes critical in an implementation of Edmonds-Karp at most $|V|/2$ times. This implies that there are at most $|E||V|/2$ augmentations, thus completing the proof.
Suppose that in some augmentation, $e = uv$ is critical for the first time. Then, $\delta_f(s, v) = \delta_f(s, u) + 1$ since we are using a shortest path. Further, after the augmentation, the edge uv disappears. To become critical again, it has to reappear, and this only happens after vu appears in a subsequent augmenting path for a flow f'. But by our earlier claim,

$$\delta_{f'}(s, u) = \delta_{f'}(s, v) + 1 \ge \delta_f(s, v) + 1 = \delta_f(s, u) + 2,$$

thus proving our second claim since $\delta_{f'}(s, u) \le |V| - 1$, so any edge becomes critical at most $(|V| - 1)/2$ times.

It remains to prove our first claim. Suppose it is not true. Then, there is a first instance and a vertex v such that f is augmented to f' but $\delta_{f'}(s, v) < \delta_f(s, v)$. Pick such a v with minimal $\delta_{f'}(s, v)$.
Let P be a shortest path in $G_{f'}$ from s to v, and let the penultimate vertex in it (before v) be u. Then, $\delta_{f'}(s, v) = \delta_{f'}(s, u) + 1$. By the choice of v, $\delta_{f'}(s, u) \ge \delta_f(s, u)$.
Now, suppose that $uv \in E(G_f)$. Then,

$$\delta_f(s, v) \le \delta_f(s, u) + 1 \le \delta_{f'}(s, u) + 1 = \delta_{f'}(s, v),$$

contradicting our assumption. Therefore, $uv \notin E(G_f)$. However, $uv \in E(G_{f'})$. This implies that an s-t augmenting path P in G_f includes the edge vu. Since Edmonds-Karp uses shortest paths, this implies that $\delta_f(s, u) = \delta_f(s, v) + 1$, so

$$\delta_f(s, v) = \delta_f(s, u) - 1 \le \delta_{f'}(s, u) - 1 = \delta_{f'}(s, v) - 2,$$

contradicting our assumption once more and concluding the proof. ■

Next, we give a more recent result on finding perfect matchings in regular bipartite graphs. Let $G = (X, Y, E)$ be d-regular with $|V| = n = 2|X|$. Theorem 3.2 says that G has a perfect matching. The Edmonds-Karp algorithm

together with the first application in Section 3.2.3 gives a $O(n^3 d)$ algorithm.
We give a far more efficient *randomized* algorithm due to [GKK11] with expected runtime $O(n \log n)$.
First, what are randomized algorithms? They utilize sources of randomness, and either

- the output is a random variable (that need not always be correct),
- the output is definitely correct, but the runtime is a random variable,

or some combination of the two.
The algorithm we describe is of the second variety.
Create the corresponding network for G, adding source and sink vertices s and t. Beginning with the 0 flow, we find an augmenting path by performing the following on the residual graph. Beginning from vertex s, if we are at vertex u, then choose an arbitrary outneighbour v of u and move to it. This algorithm gives a flow of value n,[11] and thus a perfect matching in the graph, with expected runtime $O(n \log n)$.

Remark. The expected runtime for the algorithm is independent of the degree d of the graph. While the *expected* time is of the order of $O(n \log n)$, it is possible for the algorithm to take much longer to terminate. One way to deal with this is that if the algorithm does not terminate in, say, $\lceil 10n \log n \rceil$ steps, restart the algorithm.

Now, let us give a proof of the runtime.

Proof. Suppose we have found a flow of value $k < n$. We shall determine the expected number of steps to reach a flow of size $k+1$.
Suppose without loss of generality that our size k matching $\mathcal{M}$ has edges from $X_{\mathcal{M}}$ to $Y_{\mathcal{M}}$, each of size k. Let $X_U = X \setminus X_{\mathcal{M}}$ and $Y_U = Y \setminus Y_{\mathcal{M}}$.
Observe that the only edges from Y to X in the residual graph come from the $\mathcal{M}$-edges.
Also note that the time to reach the size $k+1$ matching is just the length of the augmenting path (up to some multiplicative constant). Further, the length of an augmenting path is

$$\leq 2 + 1 + 2 \times (\text{the number of backward edges in the walk}).$$

Therefore, it suffices to bound the expected number of backward edges in a random walk from s to t in this residual graph.
For each $v \in X_{\mathcal{M}} \cup Y_{\mathcal{M}}$, denote by $\mathcal{M}(v)$ the vertex in $X_{\mathcal{M}} \cup Y_{\mathcal{M}}$ that it has a $\mathcal{M}$-edge to.
Denote by $b(v)$ the expected number of backward edges in a random walk starting at some v and terminating at t.
We wish to determine $b(s)$.
We have

$$b(s) = \frac{1}{n-k} \sum_{x \in X_U} b(x).$$

For each $x \in X_U$,

$$b(x) = \frac{1}{d} \sum_{y \in N(x)} b(y).$$

For each $y \in Y_U$, $b(y) = 0$ and for each $y \in Y_{\mathcal{M}}$,

$$b(y) = 1 + b(\mathcal{M}(y)).$$

Finally, for $x \in X_{\mathcal{M}}$,

$$b(x) = \frac{1}{d-1} \sum_{y : xy \in E \setminus \mathcal{M}} b(y).$$

[11] If any loops are formed during the creation of this augmenting path, delete them afterwards.

So, if $x \in X_{\mathcal{M}}$,

$$\begin{aligned}
d \cdot b(x) &= \sum_{y: xy \in E \setminus \mathcal{M}} b(y) + b(x) \\
&= \left(\sum_{y: xy \in E \setminus \mathcal{M}} b(y) \right) + b(\mathcal{M}(x)) - 1 \\
&= \left(\sum_{y: xy \in E} b(y) \right) - 1
\end{aligned}$$

and if $x \in X_U$,

$$d \cdot b(x) = \sum_{xy \in E} b(y)$$

Summing over all $x \in X$,

$$\begin{aligned}
d \sum_{x \in X} b(x) &= \sum_{xy \in E} b(y) - |X_{\mathcal{M}}| \\
&= \sum_{xy \in E} b(y) - k \\
&= d \sum_{y \in Y} b(y) - k \\
&= d \sum_{y \in Y_{\mathcal{M}}} b(y) - k \\
&= \left(d \sum_{y \in Y_{\mathcal{M}}} (1 + b(\mathcal{M}(y))) \right) - k \\
&= (d-1)k + d \sum_{x \in X_{\mathcal{M}}} b(x) \\
d \sum_{x \in X_U} b(x) &= (d-1)k.
\end{aligned}$$

Therefore,

$$b(s) = \frac{1}{n-k} \sum_{x \in X_U} b(x) = \frac{k(d-1)}{d(n-k)} < \frac{n}{n-k} - 1$$

and expected runtime of the algorithm is

$$O\left(\sum_{k=0}^{n-1} \left(\frac{n}{n-k} - 1 \right) \right) = O(n \log n).$$

■

3.2.5. Baranyai's Theorem

In this section, we give another application of flow due to Zsolt Baranyai. The problem is as follows.

> Suppose $n > k \geq 2$ are integers. Can we partition $\binom{[n]}{k}$ into subcollections $\mathcal{A}_1, \ldots, \mathcal{A}_M$ such that each of the sets within each A_i is itself a partition of $[n]$.

For example, consider the scenario where $k = 2$. If n is odd, the above is clearly impossible. So, let $n = 2m$. The problem is then essentially equivalent to partitioning $E(K_{2m})$ into perfect matchings. This isn't too hard to do, and answers the problem affirmatively in this case.
Note that in genereal, we must have $k \mid n$.

Theorem 3.13 (Baranyai's Theorem)**.** If $k \mid n$, there exists a partition of $\binom{[n]}{k}$ into subfamilies $\mathcal{A}_1, \dots, \mathcal{A}_M$ with $|\mathcal{A}_i| = m$, and the sets of each $\mathcal{A}_i$ partition $[n]$.

Above, we have

$$M = \frac{\binom{n}{k}}{n/k} = \binom{n-1}{k-1} \text{ and } m = \frac{n}{k}.$$

Proof. We prove this by induction. Suppose we have families $\mathcal{A}_1, \mathcal{A}_M$, each of size m satisfying the required. Suppose that n is deleted from every set that it appears in. It is easy to retrieve the original $\{\mathcal{A}_i\}$ since exactly one set from each $\mathcal{A}_i$ has size $k-1$ now, namely that from which n was deleted.
Now, suppose that both n and $n-1$ are deleted from every set they appear in. Now, it is slightly more non-trivial. How would we retrieve $\mathcal{A}_i$ (or another valid solution)?

- Any set from the $\mathcal{A}_i$ of size $k-2$ gets both $n-1$ and n.
- Each set of size $k-1$ gets either n or $n-1$. The question is: which of the size $k-1$ sets do we assign $n, n-1$ to? Note that each set of size $k-1$ appears exactly twice on scanning across all the $\mathcal{A}_i$. So, we just assign n to one of these two subsets and $n-1$ to the other. Why is this a valid solution?

Suppose now that we delete all occurrences of $\ell+1, \ell+2, \dots, n$ for some ℓ. The resulting tampered collection is such that

1. $|\mathcal{A}_i| = M$.
2. Each $\mathcal{A}_i$ consists of sets that constitute a partition of $[\ell]$.
3. Each set in a $\mathcal{A}_i$ has size at most k, and there may be multiple instances of $\varnothing$.

These tampered collections $\{\mathcal{A}_i\}$ are also such that for any $S \subseteq [\ell]$, S appears $\binom{n-|S|}{k-\ell}$ times. We wish to choose a $B_i \in \mathcal{A}_i$ for each $i \in [M]$ as the set into which $\ell+1$ should be added, and the above conditions are satisfied for the new collection $\{\mathcal{A}_i^*\}$ of sets, where $\mathcal{A}_i^* = \mathcal{A}_i \setminus \{B_i\} \cup \{B_i \cup \{\ell+1\}\}$. That is, each susbet $T \subseteq [\ell+1]$ appears exactly $\binom{n-\ell-1}{k-|T|}$ times across the sets of $\{\mathcal{A}_i^*\}$.
We can then apply the inductive hypothesis to build up to the original collection of partitions of $[n]$. If we are able to do this, then since the hypothesis is satisfied for the case where $\ell = 0$ (so $\mathcal{A}_i$ just has m copies of $\varnothing$), the proof is complete.
This is where graph theory enters the picture. Consider the bipartite graph (X, Y, E), where

$$X = \{\mathcal{A}_i\}_{i=1}^M, Y = \{S \in 2^{[\ell]} : |S| \le k\},$$

and $\mathcal{A}_i$ is adjacent to $S \subseteq [\ell]$ if $S \in \mathcal{A}_i$.
Note that the degree of any set S is equal to $\binom{n-\ell}{k-|S|}$. The idea of choosing one set for each $\mathcal{A}_i$ lends itself to a flow analogue. Put in a source s and sink t, similar to what we've done in the past, and get a directed graph G. To get a network $\mathcal{N}$, for any $\mathcal{A}_i \in X$ and $S \in Y$, we set the capacity as

$$C(s, \mathcal{A}_i) = 1 \text{ and } C(\mathcal{A}_i, S) = 1.$$

If we have $T = S \cup \{\ell+1\}$, we want T to appear $\binom{n-\ell-1}{k-|S|-1}$ times. So, we set

$$C(S,t) = \binom{n-\ell-1}{k-|S|-1} = \frac{k-|S|}{n-\ell}\binom{n-\ell}{k-|S|}.$$

Consider the following flow f. Set $f(S, \mathcal{A}_i) = 1$ for each i, $f(S,t) = C(S,t)$ for each S, and to make it a flow,

$$f(\mathcal{A}_i, S) = \frac{k-|S|}{n-\ell}.$$

It is clear Kirchhoff's law is satisfied at the subsets S, and at the $\mathcal{A}_i$, we have

$$\begin{aligned}\sum_S f(\mathcal{A}_i, S) &= \frac{k}{n-\ell}\cdot\frac{n}{k} - \frac{1}{n-\ell}\sum_{S\in\mathcal{A}_i}|S| \\ &= \frac{n}{n-\ell} - \frac{\ell}{n-\ell} = 1.\text{(since } \mathcal{A}_i \text{ forms a partition)}\end{aligned}$$

Further, this is a max-flow! By Corollary 3.11, there must exist an integral max-flow g on this graph, such that $g(s, \mathcal{A}_i) = 1$, $g(S,t) = \binom{n-\ell-1}{k-|S|-1}$, and $g(\mathcal{A}_i, S) \in \{0,1\}$.
If for each $\mathcal{A}_i, S$ such that $g(\mathcal{A}_i, S) = 1$, we add in $\ell+1$.
To conclude, we need to verify that for each $T \subseteq [\ell+1]$, T appears exactly $\binom{n-\ell-1}{k-|T|}$. If $\ell+1 \in T$, then this follows by design. Otherwise, T appeared $\binom{n-\ell}{k-|T|}$ times originally. Since $\binom{n-\ell-1}{k-|T|-1}$ of these copies were altered to include $\ell+1$, the number of copies of T is now exactly

$$\binom{n-\ell}{k-|T|} - \binom{n-\ell-1}{k-|T|-1} = \binom{n-\ell-1}{k-|T|},$$

completing the proof. ■

3.2.6. Menger's Theorem

Next, we discuss a result regarding connectivity.

Definition 3.11. A graph G is said to be k**-connected** if on the deletion of any $< k$ vertices from G, and any vertices u, v, there exists a path from u to v in the remaining graph. A 1-connected graph is said to be **connected**.
The **connectivity** $\kappa(G)$ of G is the largest integer r such that G is r-connected.
Given vertices u and v, a set of **independent** u-v **paths** is a collection of paths from u to v such that any vertex other than u and v is in at most one of these paths.
Given vertices u and v, a set of **edge-independent** u-v **paths** is a collection of paths from u to v such that any edge is in at most one of these paths.

Menger's Theorem deals with the last of the above definitions.

Theorem 3.14 (Menger's Theorem)**.** Suppose s and t are distinct vertices in a graph G. The minimum number of vertices whose deletion removes all paths from s and t is equal to the maximum number of independent s-t paths.

It is clear that the minimum number of vertices that needs to be deleted to remove all paths is at least the maximum number of independent s-t paths (otherwise, one of these paths would necessarily survive).
It remains to show the other direction. The min-max nature of the above result is somewhat reminiscent of flow, and this indeed gives a quite simple proof. Before moving to this however, let us give a more elementary proof that does not invoke flow.

Proof 1 of Menger's Theorem. Suppose that the other direction does not hold. Consider a graph with k as the minimum number of vertices needed to be deleted to destroy all s-t paths, but the maximum number of independent s-t paths is $< k$. Choose k to be the minimal possible number such that the previous statement is true (for some graph and some vertices s, t). Further, let G be a graph having the minimum number of edges satisfying the statement for this particular k.

Suppose that there exists x such that sx and xt are in $E(G)$. Then, observe that in $G \setminus \{x\}$, there must exist at most $k-2$ independent s-t independent paths. Also, in $G \setminus \{x\}$, we must delete at least $k-1$ vertices to kill all s-t paths.

In particular, $G \setminus \{x\}$ is a counterexample with $k-1$, contradicting the minimality of k. As a result, no vertex is adjacent to both s and t.

Let $W \subseteq V(G)$ be a minimal s-t separator. That is, W is the minimal set such that $G \setminus W$ has no s-t paths. By definition, $|W| = k$. Suppose that C_s and C_t are the connected components in $G \setminus W$ containing s and t respectively. Consider the graph G' with vertex set $(V(G) \setminus C_s) \cup \{s'\}$ (for some new vertex s'), with edges $s'w$ for all $w \in W$. Observe that since W is a minimal separator, there are at least $|W|$ edges in G between C_s and W. Therefore, if $|C_s| > 1$, G' has (strictly) fewer edges than G. In particular, there are k independent s'-t paths in G'. Let these paths P_i be of the form $s'w_i \cdots t$ for $i \in [k]$, where $w_i \in W$.
These describe independent paths from w_i to t for each $1 \leq i \leq k$. Similarly, if $|C_t| > 1$, one can conclude that there are disjoint paths P_i'' from s to w_i in G, for $i \in [k]$.
Concatenating the paths from s to w_i and w_i to t, we get k independent paths from s to t, contradicting our assumption.

Therefore, we may assume that if W is any s-t separator of minimum size, either s or t is adjacent to every $w \in W$. Let $sx_1 \cdots x_\ell t$ be a shortest s-t path. Note that $\ell \geq 2$ by our first observation. Consider the graph $H = G \setminus \{x_1, x_2\}$. Since H has fewer edges than G, there is a set U of size $k-1$ such that U separates s and t (in H). Let $W_1 = U \cup \{x_1\}$ and $W_2 = U \cup \{x_2\}$. It is easy to see that $|W_1| = |W_2| = k$, and both these sets separate s and t.
Since t is not adjacent to x_1, s is adjacent to all vertices in W_1. Similarly, since s is not adjacent to x_2, t is adjacent to all vertices in W_2. However, if $k > 1$, this implies that $|U| > 0$ and every vertex of U is adjacent to both s and t, leading to a contradiction. For $k \leq 1$, the result is obvious, completing the proof. ■

The above is quite tedious, and the following flow argument is *far* simpler.

Before moving to it, we need a vertex variant of the Ford and Fulkerson theorem, wherein we have capacities on vertices instead of edges. Everything remains the same, except that the capacity constraint now is that the total incoming (or equivalently, outgoing) flow at any vertex is at most the capacity of that vertex.
A cut now is a subset S of $V \setminus \{s, t\}$ such that there is no positive valued flow from s to t on $D \setminus S$. The capacity of a cut is just the sum of capacities of vertices in the cut.

These networks are equivalent to our usual networks, which is easily seen by constructing a graph as follows. Given a network (with vertex capacities), we first split each vertex v into two vertices v^+ and v^-, replacing each edge uv with u^+v^- and also adding the edge v^-v^+ for each edge. We have arbitrarily large capacity for any edges of the former type, and the capacity on the edge $C(x^-, x^+)$ is $C(x)$. The new source and sink vertices are s^- and t^+ respectively. It is not too difficult to show that any flow in the original (vertex capacity) network is equivalent to a flow in the new (edge capacity) network. Further, the value of the max-flows and min-cuts are the same in both networks.

Proof 2 of Menger's Theorem. Given the graph G with specified vertices s and t, replace each edge uv with two directed edges uv and vu to get a directed graph. Set the capacity of each vertex other than s and t as 1 (and that of s and t arbitrarily large). In this case, the max-flow value is equal to the number of independent paths, and the min-cut capacity is equal to the size of a minimum s-t separator, completing the proof. ■

A similar proof also leads to the following edge-version of Menger's Theorem.

Theorem 3.15 (Menger's Theorem)**.** Suppose s and t are distinct vertices in a graph G. The minimum number of edges whose deletion removes all paths from s and t is equal to the maximum number of edge-independent s-t paths.

We also get the following corollary quite simply from Menger's.

Corollary 3.16. If G is k-connected, then there exist k independent paths between any distinct vertices s and t.

3.2.7. Dilworth's Theorem

To conclude our discussion about flows, we discuss a result related to posets.

Definition 3.12 (Poset)**.** A (finite) **poset** is a pair $(\mathcal{P}, <)$, where $\mathcal{P}$ is a (finite) set and $<$ is a partial order on P.
A **chain** in a poset is a sequence $(x_1, \ldots, x_r)$ of elements in P such that $x_i < x_{i+1}$ for all $i \in [r-1]$.
An **anti-chain** A in a poset is a set $\{y_1, \ldots, y_r\}$ of elements in P such that for $i \neq j$, $y_i \not< y_j$ and $y_j \not< y_j$. That is, no two elements of A are comparable.

Observe that if A is an antichain and C is a chain in $\mathcal{P}$, then $|A \cap C| \leq 1$ (Why?).
Consequently, if $C_1, \ldots, C_m$ is a partition of $\mathcal{P}$ into chains, then for any antichain A in $\mathcal{P}$, $|A| \leq m$. In particular,

$$\text{size of largest anti-chain} \leq \text{size of smallest chain decomposition.}$$

Theorem 3.17 (Dilworth's Theorem)**.** In a poset $\mathcal{P}$, the size of a largest antichain is equal to the size of a smallest chain decomposition.

Proof. We prove this by induction on $|\mathcal{P}|$.
If $|\mathcal{P}| = 0$, the claim is trivial.
Suppose the size of a largest antichain in $\mathcal{P}$ is m. Let C be a maximal chain in $\mathcal{P}$. That is, for any $x \notin C$, $\{x\} \cup C$ is not a chain.
Consider the poset $\mathcal{P}' = \mathcal{P} \setminus C$. If the size of a largest antichain in $\mathcal{P}'$ is $m-1$, then by induction, $\mathcal{P}' = C_1 \sqcup \cdots \sqcup C_{m-1}$ for chains (C_i), so $\mathcal{P} = C \sqcup C_1 \sqcup \cdots \sqcup C_{m-1}$ and we are done.
Therefore, we may assume that the size of a largest antichain A in $\mathcal{P}'$ is also m. Let $A = \{a_1, \ldots, a_m\}$.
Consider the *lower shadow*

$$\mathcal{S}^- = \{x \in \mathcal{P} : x \leq a_i \text{ for some } i\}$$

and the *upper shadow*

$$\mathcal{S}^+ = \{x \in \mathcal{P} : x \geq a_i \text{ for some } i\}$$

Since A is a largest antichain, $\mathcal{P} = \mathcal{S}^- \cup \mathcal{S}^+$ and $A = \mathcal{S}^- \cap \mathcal{S}^+$.
Note that both of these shadows are strict subsets of $\mathcal{P}$. Using the inductive hypothesis, let

$$\mathcal{S}^- = \bigcup_{i=1}^{m} C_i \text{ and } \mathcal{S}^- = \bigcup_{i=1}^{m} C_i'$$

for chains (C_i) and (C_i'), where $a_i \in C_i, C_i'$. If we show that a_i is the maximum element of C_i and minimum element of C_i', we are done.
Suppose instead that a_i is not maximal C_i. Then, there is some $x \in C_i$ such that $a_i < x$. Since $x \in \mathcal{S}^-$, $x < a_j$ (why can't it be equal?), which would imply that $a_i < a_j$, giving a contradiction. The other direction holds similarly simply too, completing the proof. ■

Dilworth's Theorem turns out to be a surprisingly useful result. For example, one may prove Hall's Marriage Theorem using Dilworth's as follows.

Proof. Let $G = (X, Y, E)$ be a bipartite graph satisfying Hall's condition with $|X| \leq |Y|$. Define the poset with $\mathcal{P} = X \sqcup Y$, and $y < x$ iff $xy \in E(G)$ (where $y \in Y$ and $x \in X$).
Note that X and Y are antichains, so the size of a largest antichain is at least $|Y|$. Suppose $S \subseteq X$ and $T \subseteq Y$ are such that $S \sqcup T$ is an antichain. In particular, $S \cup T$ is independent in G. However, since $|N(S)| \geq |S|$, we must have $N(S) \subseteq (Y \setminus T)$. Consequently, $|S| + |T| \leq |N(S)| + |T| \leq |Y|$. Therefore, Y is a maximum antichain.
Using Dilworth's, there exists a chain decomposition of the poset into $|Y|$ chains. Note that each chain either corresponds to an edge in G or is a singleton in Y. In particular, every element of X is in an edge (and the edges are disjoint), completing the proof. ■

§4. Ramsey Theory

The namesake of Ramsey theory is the British logician Frank Plumpton Ramsey. An infinite form of one of the results was published in a paper on mathematical logic, and was later rediscovered by Erdős and Szekeres.

4.1. Introduction and the Erdős-Szkeres Theorem

To begin, we give the following folklore proposition.

> Among any 6 people, eithere there are 3 who are mutual acquaintances of each other or 3 who are mutual non-acquaintances of each other (being an acquaintance is a symmetric relation).

This was in fact also discovered by a Hungarian sociologist later. The proof is very simple and just boils down to showing that a graph on 6 vertices has either a size 3 clique or a size 3 independent set.
Equivalently, if we colour each edge of K_6 blue or red, there is a monochromatic triangle.
This can be proved as follows. Pick any vertex v. By the pigeonhole principle, three of its neighbours u_1, u_2, u_3 are such that vu_1, vu_2, vu_3 are of the same colour, say red. If one of the edges u_iu_j is red as well, we are done. If not, $u_1u_2u_3$ is a monochromatic blue triangle, so we are done.
It is also not difficult to see that 6 is tight (there exists a red-blue colouring of the edges of K_5 without a monochromatic triangle).

This leads to the following more general question.

> Suppose $s, t \in \mathbb{N}$ at least 2. What is the minimum N (if one exists) such that if each edge of K_N are coloured red or blue, there is either a red K_s or a blue K_t.

Theorem 4.1. Given $s, t \in \mathbb{N}$ at least 2, there exists a quantity $R(s, t) \in \mathbb{N}$ such that for all $n \geq R(s, t)$, any red-blue colouring of $E(K_n)$ admits a red K_s or a blue K_t.

It is obvious that $R(s, t) = R(t, s)$.

Proof. We prove this by induction on (s, t).
If $s = 2$, it is easy to see that $R(2, t) = t$ (similarly, $R(s, 2) = s$).
Let v be an arbitrary vertex. Since v has degree $n - 1$, either it has R_1 red neighbours or R_2 blue neighbours for any R_1, R_2 such that $R_1 + R_2 = n$ (we shall fix R_1 and R_2 later). Suppose x has R_1 red neighbours. If $R_1 \geq R(s-1, t)$, then we are done by induction. Similarly, if $R_2 \geq R(s, t-1)$, we are done. So, set $R = R(s, t-1) + R(s-1, t)$. ■

Corollary 4.2. We have
$$R(s, t) \leq \binom{s+t-2}{s-1}.$$

Proof. Letting $N(s,t) = \binom{s+t-2}{s-1}$, note that $R(2,t) = N(2,t)$. It may be checked that $N(s,t) = N(s-1,t) + N(s,t-1)$, and we are done by the preceding theorem. ■

More generally, we have the following definition.

Definition 4.1. For integers $s_1, \ldots, s_r$, the **Ramsey number** $R(s_1, s_2, \ldots, s_r)$ is the minimum n such that for every r-colouring of $E(K_n)$ using $[r]$, there is an i-monochromatic clique of size s_i for some i.

We have the following.

Theorem 4.3. For $r \in \mathbb{N}$ $s_1, \ldots, s_r \in \mathbb{N}$, $R(s_1, \ldots, s_r)$ is well-defined and moreover,

$$R(s_1, \ldots, s_r) \leq R(s_1 - 1, s_2, \ldots, s_r) + R(s_1, s_2 - 1, s_3, \ldots, s_r) + \cdots + R(s_1, \ldots, s_{r-1}, s_r - 1).$$

The proof of the above is a straightforward generalization of the induction in Theorem 4.1.
Another generalization is that to a *hyper*graph, in which edges are formed not by pairs of vertices, but instead sets of vertices. The complete r-hypergraph on n vertices K_n^r is that where the edge set is $\binom{[n]}{r}$. Then, if we colour each edge of this graph red or blue

$$R^{(r)}(s,t) = \min\{n : \text{there is a red } K_s^r \text{ or a blue } K_t^r\}.$$

Of course, we can get a generalization combining both the hypergraph aspect and the multicolour aspect.

Theorem 4.4. $R^{(r)}(s,t)$ is well-defined and further,

$$R^{(r)}(s,t) \leq R^{(r-1)}(R^{(r)}(s-1,t), R^{(r)}(s,t-1)) + 1.$$

The proof of the above is not too difficult.
Consider the diagonal case where $s = t$. We showed above that $R(3,3) = 6$. It may also be shown that $R(4,4) = 18$. Beeyond, this we do not know any diagonal Ramsey numbers. By Corollary 4.2,

$$R(s,s) \leq \binom{2s-2}{s-1}.$$

Recall Stirling's formula

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n.$$

It may be checked that

$$\binom{2s-2}{s-1} \sim c \cdot \frac{4^n}{\sqrt{n}}$$

for some constant $c > 0$.
Finding the exact Ramsey number is incredibly difficult. In the words of Erdős,

> Suppose aliens invade the earth and threaten to obliterate it in a year's time unless human beings can find the Ramsey number for red five and blue five. We could marshal the world's best minds and fastest computers, and within a year we could probably calculate the value. If the aliens demanded the Ramsey number for red six and blue six, however, we would have no choice but to launch a preemptive attack.

This begs two questions:

- Is there a better upper bound for $R(s,s)$ than this exponential one?
- What about a lower bound on $R(s,s)$?

The second of the above questions asks us to give a red-blue colouring of $E(K_n)$ such that there is no monochromatic K_s.
The best known bound for a long time was quadratic, and Turán believed that the Ramsey number itself might be quadratic.

Blowing this bound out of the water however, Erdős gave the first (documented) example of the probabilistic method.

Theorem 4.5. We have

$$R(s,s) > 2^{s/2}.$$

Proof. Fix some n and colour each edge of K_n blue or red with probability $1/2$ each. For a fixed $S \subseteq V(K_n)$ of size s,

$$\Pr[S \text{ is monochromatic}] = \frac{2}{2^{\binom{s}{2}}}.$$

So,

$$\begin{aligned}\Pr[\text{there is some monochromatic } K_s] &= \Pr\left[\bigcup_{|S|=s} \{S \text{ is monochromatic}\}\right] \\ &\le \binom{n}{s} \cdot \frac{2}{2^{\binom{s}{2}}}.\end{aligned}$$

So, if

$$\frac{\binom{n}{s}}{2^{\binom{s}{2}}} < \frac{1}{2},$$

n is a lower bound for $R(s,s)$. Simplifying the above, we have

$$\begin{aligned}\frac{\binom{n}{s}}{2^{\binom{s}{2}}} &\le \frac{n^s}{s!2^{s(s-1)/2}} \\ &\le \left(\frac{n}{2^{(s-1)/2}}\right)^s \cdot \frac{1}{s!}.\end{aligned}$$

If $n = 2^{s/2}$, the above is $(\sqrt{2})^s/s! < 1/2$ if $s \ge 3$, completing the proof. ■

Further note that as s grows, since $(\sqrt{2})^s = o(s!)$, if $n = 2^{s/2}$, the probability of there not existing a monochromatic K_s in a random colouring of $E(K_n)$ goes to 0.
So, we now have

$$2^{s/2} < R(s,s) < 4^s.$$

For a very long time, 4^s was the best known upper bound. In 2008, David Conlon showed in [Con09] that

$$R(s+1,s+1) \le \binom{2s}{s} \cdot \frac{1}{s^{O(\log s/\log\log s)}}.$$

In 2020, Ashwin Sah improved this in [Sah20] to

$$R(s+1,s+1) \le \binom{2s}{s} \cdot \frac{1}{s^{O(\log s)}}.$$

There is also a geometric motivation for Ramsey Theory.
A configuration of points on the plane is said to be in general position if no three of them are collinear. Esther Klein (later Esther Szekeres) noticed the following.

1. Given any 5 points in general position, 4 of them are the vertices of a convex quadrilateral.
2. Given any 9 points in general position, 5 of them are the vertices of a convex pentagon.

This leads to the following question.

> Given $n \in \mathbb{N}$, is there a (finite) $N(n)$ such that any $N(n)$ points in general position have the vertices of a convex n-gon?

Theorem 4.6 (Erdős-Szekeres Theorem)**.** Given any n, there exists a finite $\mathrm{ES}(n)$ such that any $\mathrm{ES}(n)$ vertices in general position admit the vertices of a convex n-gon. Furthermore,

$$\mathrm{ES}(n) \leq \binom{2n-4}{n-2} + 1.$$

The reader might notice that the above is quite similar to the earlier Ramsey number bound!

Proof 1 showing finiteness in the Erdős-Szekeres Theorem, due to M. Tarsi. This proof hinges on the observation that if a (finite) set X of points is such that every 4 points of X form a convex quadrilateral, then all the points of X form a convex polygon. We do not prove this.
Now, we claim that $N(n) = R^{(4)}(n,5)$ works as an upper bound to $\mathrm{ES}(n)$. Let X be our set of N points (for $N > N(n)$) in general position. Colour each 4-tuple of the points of the set X red or blue depending on whether or not the points form a convex quadrilateral. Since $R^{(4)}(n,5)$ is finite, either there is a set $Y \subseteq X$ of n points such that all the 4-tuples of Y are red, or there is a size 5 set $Z \subseteq X$ such that all 4-tuples of Z are blue.
In the former case, the vertices of Y form a convex n-gon by the observation in the first paragraph. Further, the latter case cannot occur, due to the observation of Klein's we had given earlier. ■

This bound is *terrible*.

Proof 2 due to Erdős-Szekeres. Suppose that the points in general position are $p_i = (x_i, y_i)$ for $1 \leq i \leq N$ and also that $x_1 < x_2 < \ldots < x_N$ (the second may be assumed by rotating the plane appropriately).
Call a set $C = \{p_{i_1}, \ldots, p_{i_k}\}$ a k**-cup** if $i_1 < \cdots < i_k$ and the slope of the segments $p_{i_j}p_{i_{j+1}}$ is non-decreasing. Similarly, call a set $C = \{p_{i_1}, \ldots, p_{i_k}\}$ a k**-cap** if $i_1 < \cdots < i_k$ and the slope of the segments $p_{i_j}p_{i_{j+1}}$ is non-increasing.
Clearly, a k-cup or k-cap form a convex k-gon.

The main result proved by Erdős-Szekeres is that for $N > \phi(k,\ell)$, any set of the assumed form above admits either a k-cup or an ℓ-cap, where

$$\phi(k,\ell) = \binom{k+\ell-4}{k-2}.$$

We prove this by induction on $k+\ell$. If $k=2$ or $\ell=2$, the result is trivial. Let X be a set of size $\phi(k,\ell)+1$, and suppose that X contains neither a k-cup nor an ℓ-cap. Let L be the set of last points of $(k-1)$-cups. In particular, $X \setminus L$ has neither $(k-1)$-cups nor ℓ-caps. Consequently, $|X \setminus L| \leq \phi(k-1,\ell)$, so

$$L \geq 1 + \phi(k,\ell) - \phi(k-1,\ell) = \phi(k,\ell-1) + 1.$$

Therefore, L contains either a k-cup or an $(\ell-1)$-cap. In the former case, we are done. So, suppose that L contains an $(\ell-1)$-cap $\{p_{i_1}, \ldots, p_{i_{\ell-1}}\}$. p_{i_1} is the last point of a $(k-1)$-cup, so let the previous point in this cup be q.
If the slope of qp_{i_1} is greater than the slope of $p_{i_1}p_{i_2}$, then $qp_{i_1}p_{i_2}\cdots p_{i_{\ell-1}}$ forms an ℓ-cap. Otherwise, the earlier $(k-1)$-cup together with p_{i_1} forms a k-cup, completing the proof. ■

The bound of $\phi(k,\ell)+1$ is tight for the k-cup ℓ-cap problem. That is, there are point configurations of size $\phi(k,\ell)$ points that have neither k-cups nor ℓ-caps.
This is proved by induction on $k+\ell$. If $k=2$ or $\ell=2$, then $\phi(k,\ell)=1$. Our counterexample set S will be of the form

$$S = \{(i, y_i) : 1 \le i \le \phi(k,\ell)\}.$$

If $k=2$ or $\ell=2$, this is $S=\{(1,0)\}$.
Let Y and Z be the counterexample sets for $\phi(k-1,\ell)$ and $\phi(k,\ell-1)$ respectively. Let $Y^{(\varepsilon)} = \{(i, \varepsilon y_i) : 1 \le i \le \phi(k-1,\ell)\}$ and $Z^{(\varepsilon)} = \{(\phi(k-1,\ell)+i, y+\varepsilon y_i) : 1 \le i \le \phi(k,\ell-1)\}$. Pick $\varepsilon > 0$ small enough and y large enough that any line through two points of $Y(\varepsilon)$ lies below the set $Z^{(\varepsilon)}$ and any line through two points of $Z^{(\varepsilon)}$ lies above the set $Y^{(\varepsilon)}$.
$S = Y^{(\varepsilon)} \cup Z^{(\varepsilon)}$ provides a counterexample. For example, if S has a k-cup, there must be at least two points from $Z^{(\varepsilon)}$ (since there is no $(k-1)$ cup in $Y^{(\varepsilon)}$). However, the line joining these two points is above $Y^{\{}(\varepsilon)\}$ so cannot be part of a k-cup.

4.2. Infinite Ramsey Theory

In this section, we present an infinite version of Ramsey's Theorem. This was the original result given by Ramsey.

Theorem 4.7. Suppose V is an infinite set. Let $c : V^r \to [k]$. Then, there exists an infinite set $A \subseteq V$ such that for all $(v_1, \dots, v_r) \in A^r$, $c(v_1, \dots, v_r)$ is the same. That is, A is monochromatic.

If V is finite, $r=2$, $c=2$, and $c(v_1,v_2)=c(v_2,v_1)$, we are back to the usual Ramsey's Theorem.

Proof. We perform induction on r. The result is trivial for $r=1$ by the pigeonhole principle. Let $A_0 = V$ and $x_1 \in A_0$. Consider the induced k-colouring c^* on $(A_0 \setminus \{x_1\})^{r-1}$, where

$$c^*(v_1, \dots, v_{r-1}) = c(x_1, v_1, \dots, v_{r-1}).$$

By induction, there exists an infinite $A_1 \subseteq A_0$ such that every $(r-1)$-tuple of A_1 gets the same colour α_1.
Pick $x_2 \in A_1$, and consider the induced colouring of $(r-1)$ tuples of $A_1 \setminus \{x_2\}$ to get A_2 such that all $(r-1)$-tuples of A_2 get the colour α_2.
This leads to a sequence $A_0 \supseteq A_1 \supseteq A_2 \supseteq \cdots$, where for any $(r-1)$-tuple $(v_1, \dots, v_{r-1})$ from A_j, $(x_j, v_1, \dots, v_{r-1})$ is coloured α_j. However, $\alpha_i \in [k]$, so some colour α occurs infinitely many times in the (α_i) Let $i_1 < i_2 < \cdots$ with $\alpha_{i_j} = \alpha$ for all j.
Then, the set $\{x_{i_1}, x_{i_2}, \dots\}$ does the job. ■

The infinite version of Ramsey's theorem implies the finite one. For this reduction, we prove the following.

Theorem 4.8. Suppose $\mathcal{H} = (V, E)$ is a hypergraph and V is an arbitrary set with all edge sizes finite. Let $k \ge 2$ be an integer and suppose that for all finite $W \subseteq V$, there is a colouring $\chi_W : W \to [k]$ such that no edge $e \subseteq W$ is monochromatic (under χ_W). Then, there exists $\chi : V \to [k]$ such that no edge of $\mathcal{H}$ is monochromatic under χ.

Proof 1. k-colorings are precisely the elements of $[k]^V$. Give $[k]$ the discrete topology (all sets are open/closed), and give $[k]^V$ the product topology. By Tychonoff's Theorem, $[k]^V$ is compact (since $[k]$ is compact). For finite $W \subseteq V$, let

$$\mathcal{H}_W = \{f \in [k]^V : \text{no } e \subseteq W \text{ is monochromatic under } f\}.$$

Observe that $\mathcal{H}_W \neq \varnothing$ (since $\chi_W \in \mathcal{H}_W$) and further, $\mathcal{H}_W$ is closed in $[k]^V$ (since W is finite). If $W_1, W_2, \dots, W_\ell$ are finite, then

$$\mathcal{H}_{W_1} \cap \mathcal{H}_{W_2} \cap \cdots \cap \mathcal{H}_{W_\ell} \supseteq \mathcal{H}_{W_1 \cup \cdots \cup W_\ell}.$$

Since the W_i are finite, so is $W_1 \cup \cdots \cup W_\ell$. It follows that $\mathcal{H}_{W_1} \cap \cdots \cap \mathcal{H}_{W_\ell}$ is non-empty.
In particular, the collection $\{\mathcal{H}_W\}_{W \text{ finite}}$ has the closed intersection property. This implies that $\bigcap_{|W|<\infty} \mathcal{H}_W$ is non-empty. Any χ belonging to this set does the job, completing the proof. ■

Let us give another proof in a specific case.

Proof 2 when $V = \mathbb{N}$. Suppose that for each $n \in \mathbb{N}$, we have $\chi_n : [n] \to [k]$ such that no edge $e \subseteq [n]$ is monochromatic with respect to χ_n. We shall define $\chi^* : \mathbb{N} \to [k]$ such that no edge is monochromatic with respect to χ^*.
Let $A_0 = \mathbb{N}$. There exists some infinite set $A_1 \subseteq A_0$ such that $\chi_n(1) = \alpha_1$ for all $n \in A_1$. Set $\chi^*(1) = \alpha_1$. In general, there exists some infinite $A_j \subseteq A_{j-1}$ such that $\chi_n(j) = \alpha_j$ for all $n \in A_j$. Set $\chi^*(j) = \alpha_j$. Consider an edge $e = \{i_1, \dots, i_r\}$ and suppose $i_1 < i_2 < \cdots < i_r$. Since χ^* agrees with χ_n on $[i_r]$ for some $n \geq i_r$ and χ_j has no monochromatic edges, e is not monochromatic under χ^* either. ■

The contrapositive of this result implies the finite Ramsey result.

Ramsey's proof shows that given an infinite set A and $\binom{A}{r}$ is $[k]$-colored, there is an infinite (countable!) subset $B \subseteq A$ such that $\binom{B}{r}$ is monochromatic. Can we guarantee that $|B| = |A|$? It turns out that we cannot, due to a proof by Sierpinski in 1932.
Let $A = \mathbb{R}$ and $k = r = 2$. It is known (assuming the axiom of choice) that any non-empty set has a well-order $<^*$ (every non-empty subset has a least element).
Colour xy red if $(x < y \iff x <^* y)$ and blue otherwise. We claim that under this colouring, any monochromatic infinite set is countable. The colouring is such that for a red set A, the two orders $<$ and $<^*$ agree on A. That is, $<$ is a well-order on A. So, let a_0 be the least element of A, and in general, a_{i+1} be the least element of $A \setminus \{a_0, a_1, \dots, a_i\}$. Pick a rational $q_i \in (a_i, a_{i+1})$. Since the rationals are countable, a red monochromatic set is countable. If A is blue, then $<^*$ and $>$ agree on A so the same reasoning applies, thus proving that any monochromatic set is countable.

References

[Con09] David Conlon. A new upper bound for diagonal ramsey numbers. *Annals of Mathematics*, 170(2):941–960, 2009.

[Erd59] P. Erdös. Graph theory and probability. *Canadian Journal of Mathematics*, 11:34–38, 1959.

[GKK11] Ashish Goel, Michael Kapralov, and Sanjeev Khanna. Perfect matchings in $O(n \log n)$ time in regular bipartite graphs, 2011.

[Gow97] W. T. Gowers. Lower bounds of tower type for Szemerédi's uniformity lemma. *Geometric & Functional Analysis GAFA*, 7(2):322–337, May 1997.

[KRS96] János Kollár, Lajos Rónyai, and Tibor Szabó. Norm-graphs and bipartite Turán numbers. *Combinatorica*, 16(3):399–406, 1996.

[Sah20] Ashwin Sah. Diagonal ramsey via effective quasirandomness. *arXiv: Combinatorics*, 2020.