Friendly Flag Algebras

Yves Jäckle
Version 1

Contents

1 Densities in extremal combinatorics

1.1 Extremal graph theory

We start with a typical introduction to extremal graph theory.

1.1.1 Triangle-free graphs and Mantel's theorem

Intuitively, if a graph has many edges, it should have a triangle: the more edges you add to the graph, the harder it gets to add them between vertices that don't already have a common neighbour. Let's ask for precision: *what is the most edges an n-vertex graph can have, so as to contain no triangle ?*

The answer is:

Mantel's theorem:

A triangle-free graph on n vertices can have at most $\left\lfloor \frac{n^2}{4} \right\rfloor$ edges.

As was suggested just before, a triangle-free graph has the following property: the neighbourhood of a vertex is an independent set, in the sense that there are no edges between distinct neighbours, as such edges would close a triangle. So for any vertex v, all edges of the graph must have at least one of their endpoints in $V \backslash N(v)$.

Since our goal is to bound the number of edges, we will count them by their endpoints (like in the handshake lemma), which we now have information on. In the sum $\sum_{u \in V \backslash N(v)} \deg(u)$, all edges will be accounted for at least once, as they have an endpoint in $V \backslash N(v)$ and will be counted by a corresponding $\deg(u) = |\delta(u)| = |N(u)|$, with only the edges with both endpoints in $V \backslash N(v)$ that will be double-counted, once for each endpoint, so that $|E| \leqslant \sum_{u \in V \backslash N(v)} \deg(u)$.

To make a certain parallelism appear, recalling that v was actually arbitrary, we can bound $\sum_{u \in V \backslash N(v)} \deg(u) \leqslant |V \backslash N(v)| \max_u(\deg(u))$, which specified for a maximum degree vertex v_M yields $|E| \leqslant |V \backslash N(v_M)| \deg(v_M) = (|V| - |N(v_M)|).|N(v_M)|$. Finally, applying inequality $(n-x)x \leqslant \frac{n^2}{4}$ (which is equivalent to $(n-2x)^2 \geqslant 0$), we get $|E| \leqslant \frac{n^2}{4}$, and since we're dealing with integers $|E| \leqslant \left\lfloor \frac{n^2}{4} \right\rfloor$.

Is this a tight bound, or are we overestimating here ?
Let's seen what we get from having equality in our previous inequalities.
First, $|E| = \sum_{u \in V \backslash N(v_M)} \deg(u)$ means that no double-counting may occur: all edges have exactly one endpoint in $V \backslash N(v_M)$. This means that we're actually dealing with a bipartite graph, with bipartition sets $V \backslash N(v_M)$ and $N(v_M)$. Next, $(n - |N(v_M)|).|N(v_M)| = \left\lfloor \frac{n^2}{4} \right\rfloor$ can actually only be achieved by a unique particular integer $|N(v_M)|$. It must be $|N(v_M)| = \left\lfloor \frac{n}{2} \right\rfloor$ or $|N(v_M)| = \left\lceil \frac{n}{2} \right\rceil$, since $x \mapsto (n-x)x$ is increasing

om $\left[0, \frac{n}{2}\right]$, and decreasing on $\left[\frac{n}{2}, n\right]$.

Now, for a complete bipartite graph $K_{\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil}$, which is triangle-free, the number of edges is $\left\lfloor \frac{n}{2} \right\rfloor \left\lceil \frac{n}{2} \right\rceil$. For even n, this equals $\left\lfloor \frac{n^2}{4} \right\rfloor$ as we can drop the floor and ceiling, and for odd $n = 2k+1$, we have $\left\lfloor \frac{n}{2} \right\rfloor \left\lceil \frac{n}{2} \right\rceil = k(k+1)$, and $\left\lfloor \frac{n^2}{4} \right\rfloor = \left\lfloor \frac{4k^2+4k+1}{4} \right\rfloor = k^2 + k = k(k+1)$.

1.1.2 Clique-free graphs and Turán's theorem

The previous problem can be generalized by asking how many edges a graph can have, if we request that it doesn't contain subgraphs of a certain family of graphs we fix.

Extremal problem:

For a family $\mathcal{H}$ of graphs, we consider the **extremal number** $ex(n, \mathcal{H})$ to be the maximum number of edges an n-vertex graph may have so that it contains no graphs of $\mathcal{H}$ as induced subgraphs.
Often, $\mathcal{H} = \{H\}$ is a single graph an we write $ex(n, H)$ instead of $ex(n, \mathcal{H})$.

For one more example, we consider the case of forbidding cliques K_{r+1}, for which we have:

Turán's theorem:

For $r \geqslant 1$ and for $n \geqslant r+1$, we have $(r-1)\left(\frac{n}{r} - 1\right)\frac{n}{2} \leqslant ex(n, K_{r+1}) \leqslant \left(1 - \frac{1}{r}\right)\frac{n^2}{2}$,
with a tight upper-bound when $r|n$.

We'll try to build an extremal graph by analogy to Mantel's theorem. In Matel's theorem, the bipartiteness of the extremal graph prevented the appearance of tringles. *Does an r-partite graph prevent the appearance of K_{r+1} ?*

Yes it does. If it contained a K_{r+1}, then by labeling its vertices with their partition sets, we'd have to distribute r labels on $r+1$ vertices, so that two vertices at least would end up with the same label. Since vertices of a same partition set can't be connected by an edge, this contradicts the fact that they would be connected in the hypothetical K_{r+1}, which therefore may not exist.

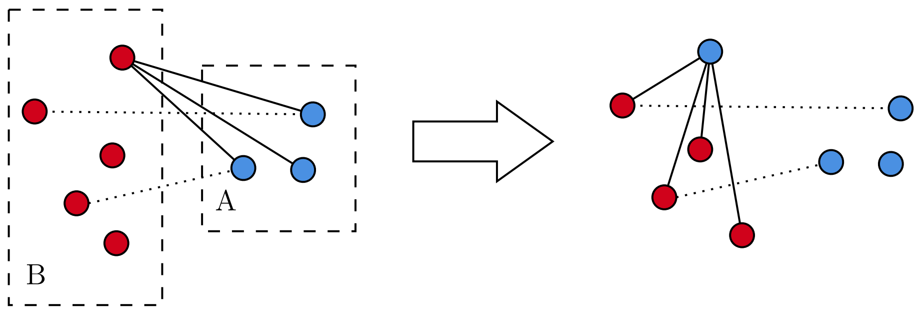
Now, *what's a r-partite graph that has the most edges as possible, for a fixed number n of vertices ?*
You can try a few examples for $r = 3$ and $n = 7$ and conjecture that there are the most egdes when the partition sets are "balanced", tin the sense that that have the same number of vertices. In fact, we can prove that:

Maximum edge r-partite graphs:

It's atteind be a complete r-partite graph in which partition sets have size $\left\lfloor \frac{n}{r} \right\rfloor$ or $\left\lceil \frac{n}{r} \right\rceil$.

Of course, the graph should be complete (in the sense that a vertex is connected to all other vertices of different partition sets), for otherwise, we could add an edges between vertices of different partition sets and contradict hypothetical maximality.

To see that the partition sets should be "balanced", we show a sort of augmentation lemma: if A and B are partition sets such that $|A| + 2 \leqslant |B|$, then we can get an r-partite graph with more edges. Indeed, we can pick a vertex in B, and assign it to A: to maintain r-partiteness, we delete its edges leading to A, but we may now add edges leading to the rest of B. This leads to a net change of $(|B| - 1) - |A| > 0$ number of edges, hence an increase.
So in a maximum edge r-partite graph, partition sets can only differ by 1 in size.
Here's a figure illustrating the proof:



Next, if we fix a first partition set A and we consider a partition set B with on less vertex, and another C with one more partition vertex, then B and C differ by more then one vertex. Since this is impossible, all partition sets must either be of sizes $|A|$ or $|A| - 1$, or all must either be of sizes $|A|$ or $|A| + 1$. If we denote by s the smaller size and by s' the number of such partition sets, and by t the greater size and by t' their number, then $s's + t't = n$ and $s' + t' = r$. In particular, $\frac{s'}{r}s + \frac{t'}{r}t = \frac{n}{r}$ is a convex combination of two successive integers, so that $s = \left\lfloor \frac{n}{r} \right\rfloor$ and $t = \left\lceil \frac{n}{r} \right\rceil$, for $\frac{n}{r}$ is only in the successive-integer-interval $\left[\left\lfloor \frac{n}{r} \right\rfloor, \left\lceil \frac{n}{r} \right\rceil\right]$.

This proves our little lemma.
As a remark, t' is the rest of n by r, since $n = (s' + t') \left\lfloor \frac{n}{r} \right\rfloor + t' = s' \left\lfloor \frac{n}{r} \right\rfloor + t' \left\lceil \frac{n}{r} \right\rceil$.

How many edges do such maximum r-partite graphs have ?
Using $|E| = \frac{1}{2} \sum_{v \in V} \deg(v)$ and the fact that in such a graph a vertex is connected to the vertices of $(r - 1)$ other partition sets, each such set having at most $\left\lceil \frac{n}{r} \right\rceil$ vertices, we can bound $|E| \leqslant \frac{n}{2}(r - 1) \left\lceil \frac{n}{r} \right\rceil$. However, we can do a better analysis.
For the ss' vertices that are in partition sets of size $\left\lfloor \frac{n}{r} \right\rfloor$, we have degrees $(s'-1)s+t't$, and for the tt' vertices in partition sets of size $\left\lceil \frac{n}{r} \right\rceil$, we have degrees $s's+(t'-1)t$, so that $\sum_{v \in V} \deg(v) = (s'-1)s+t't+s's+(t'-1)t$.
Next, we can make use of identity $\frac{s'}{r}s+\frac{t'}{r}t = \frac{n}{r}$, so that we can rewrite the terms $r.ss' \left(\frac{(s'-1)}{r}s + \frac{t'}{r}t\right) = rss'\frac{n-s}{r}$ and $r.tt' \left(\frac{s'}{r}s + \frac{(t'-1)}{r}t\right) = rtt'\frac{n-t}{r}$. Adding them up, we get $r\frac{n^2 - s's^2 - t't^2}{r}$.

With convexity of the square, we have $\frac{s'}{r}s^2 + \frac{t'}{r}t^2 \geqslant \left(\frac{n}{r}\right)^2$, so that we can bound $r\frac{n^2 - s's^2 - t't^2}{r} \leqslant n^2 - r\left(\frac{n}{r}\right)^2 = \left(1 - \frac{1}{r}\right)n^2$, which provides us the upper-bound from Turán's theorem (don't forget the $\frac{1}{2}$ preceding the sum).

We can also get the lower bound from Turán's theorem for the number of edges of these graphs. Indeed, we lower-bound the degrees by $(r-1)\left\lfloor \frac{n}{r} \right\rfloor > (r-1)\left(\frac{n}{r} - 1\right)$ and sum them up.

So the the number of edges of maximum edge r-partite graphs are candidates for $ex(n, K_{r+1})$.

We'll now actually show that the maximum edge r-partite graphs truely are the maximum egde garphs that contain no K_{r+1}, so that the bounds from the theorem statement follow from our discussion so far.

Following the proof of Mantel's theorem, which is the case $r = 2$ as triangles are K_3, we consider the neighbourhoods of vertices. For a vertex v, this time, $N(v)$ may have edges, but to prevent a K_{r+1} from appearing, we must not find a K_r as a subgraph of the graph induced by $N(v)$: otherwise, adding v to it would produce a K_{r+1} in the initial graph.
This calls for an attempt at an induction on r. The base case holds, as for $r = 1$, forbidding K_2 is the same as forbidding edges, so that $ex(n, K_{r+1}) = 0$ and the bounds from the theorem statement hold. The r-partite graphs are edgeless graphs in this context, so that they are indeed the maximisers.

In the induction step, our goal is to show that any given K_{r+1}-free graph has less edges then an r-partite graph, so that in particular it has less edges that the maximum edge r-partite graph.
We look at $E(N(v))$, which we can bound by $ex(|N(v)|, K_r)$, which we know to be attained by an $(r-1)$-partite graph on $|N(v)|$ vertices, if $|N(v)| \geqslant (r-1)$.

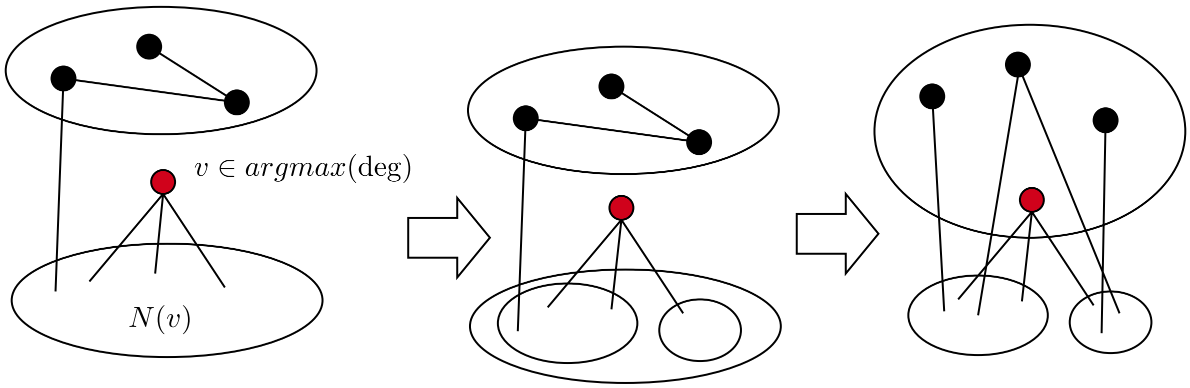
If the however maximum degree is less then $(r-1)$, so that $|N(v)| < (r-1)$ at all vertices, then we have bound $(r-1)\frac{n}{2}$ on edges, which is already less then the lower-bound $(r-1)\left(\frac{n}{r} - 1\right)\frac{n}{2}$, since $n \geqslant r+1$, so that we can discard these cases already.

So we know that when we find some vertex for which $|N(v)| \geqslant (r-1)$, we could get a graph with more edges by replacing the edges induced by $N(v)$ with those of the $(r-1)$-partite graph on $|N(v)|$ vertices. This graphs may however contain a K_{r+1}.
We can then make this graph an r-partite one (which has no K_{r+1}), by deleting all edges induced by $V \backslash N(v)$. Indeed, $V \backslash N(v)$ would then be the rth partition set appended to the $(r-1)$ ones we got from replacing the graph as $N(v)$ by an $(r-1)$-partite one. With this operation however, we lost edges. We can however get edges back by adding edges between vertices of $N(v)$ and $V \backslash N(v)$, which maintains r-partiteness. We need the number of edges added to be greater then those deleted, to get a net increase in edges.

The number of edges in the cut $\delta(N(v), V \backslash N(v))$ in the new r-partite graph will be $|N(v)||V \backslash N(v)|$: one for every pair of vertices in $N(v)$ and not in $V \backslash N(v)$. On the other hand, the edges other then those induced by $N(v)$ in the old graph could be upper-bounded by $\sum_{u \in V \backslash N(v)} \deg(u)$, which counts edges in the cut $\delta(N(v), V \backslash N(v))$ once and edges induced by $V \backslash N(v)$ twice. If we bound this further by $|V \backslash N(v)| \max_{u \in V \backslash N(v)} (\deg(u))$, we see that the new graph will have more edges then the old one if $|V \backslash N(v)| \max_{u \in V \backslash N(v)} (\deg(u)) \leqslant |N(v)||V \backslash N(v)|$ aka. $\max_{u \in V \backslash N(v)} (\deg(u)) \leqslant |N(v)|$.

This happens when we chose v to be the maximum degree vertex in the original graph. For this vertex, $|N(v)| \geqslant (r-1)$ unless we're in the case where all vertices satisfied $|N(v)| < (r-1)$, a case we already handled.

So in conclusion, for v the maximum degree vertex in the original graph, this construction will yield an r-partite grah with more edges then the original one. So our whole strategy works out.
We summarize some steps of it with this figure:



1.1.3 Edge-, graph- and Turán densities

The bounds from Turáns theorem are beggin us to take a limit $n \to \infty$ and look at the asymtotics of $ex(n, K_{r+1})$. This is where we have a first appearance of the notion of density:

Densities (part 1):

The **Turán edge density** for the forbidden family $\mathcal{H}$ at n is $\frac{ex(n,\mathcal{H})}{\binom{n}{2}}$, or in words:
the ratio of the maximum edges forming a $\mathcal{H}$-free n-vertex graph to the total possible number of edges of an n-vertex graph.

Indeed, we have:

Asymptotic Turán:

$$ex(n, K_{r+1}) \sim_{n\to\infty} \left(1 - \frac{1}{r}\right)\binom{n}{2}$$

We have $\dfrac{(r-1)\left(\frac{n}{r}-1\right)\frac{n}{2}}{\binom{n}{2}} \leqslant \dfrac{ex(n,K_{r+1})}{\binom{n}{2}} \leqslant \dfrac{\left(1-\frac{1}{r}\right)\frac{n^2}{2}}{\binom{n}{2}}$.

By writing $(r-1)\left(\frac{n}{r}-1\right)\frac{n}{2} = \left(1-\frac{1}{r}\right)\frac{(n-r)n}{2}$ (introduce $\frac{1}{r}r$ between the first factors), and using the facts that $\frac{(n-r)n}{n(n-1)} \xrightarrow[\infty]{n} 1$ and $\frac{n^2}{n(n-1)} \xrightarrow[\infty]{n} 1$, we have sandwich convergence of $\frac{ex(n,\mathcal{H})}{\binom{n}{2}}$ to $\left(1-\frac{1}{r}\right)$.

An interesting question to ask is: *does the Turán edge density always have a limit ?*
The answer starts with studying on whether the density increases or decreses with n. The key fact is that subgraphs of $\mathcal{H}$-free graphs are also $\mathcal{H}$-free (as subgraphs of subgraphs are subgraphs).
So if we pick a graph G attaining the maximum in $ex(n+1, \mathcal{H})$, then for any n-vertex subgraphs S of G, of which there are $\binom{n+1}{n} = n+1$, we have $|E_S| \leqslant ex(n, \mathcal{H})$. To use this fact to bound $|E_G| = ex(n+1, \mathcal{H})$, we will double count edges as follows. When ranging over the $n+1$ subgraphs S, and edge of G is counted in the E_S for all S except for the two in which the endpoint of the edge is missing (as n-vertex subgraphs S can be obtained by ignoring a single vertex of G). Therefore $\sum_S |E_S| = (n+1-2)|E_G|$, so that we can bound $(n+1).ex(n, \mathcal{H}) \geqslant (n-1).ex(n+1, \mathcal{H})$. Together with the relation $\binom{n+1}{2} = \frac{n+1}{n-1}\binom{n}{2}$, the inequality shows that Turán edge densities decrease with n.
Since they are positive, this means that the limit as $n \to \infty$ always exists. So we can name it:

Densities (part 2):

The **Turán density** for the forbidden family $\mathcal{H}$ is $\pi(\mathcal{H}) = \lim_{n\to\infty} \frac{ex(n, \mathcal{H})}{\binom{n}{2}}$.

For example, we've just shown that $\pi(K_{r+1}) = \left(1 - \frac{1}{r}\right)$.

We have so far only studied the case where $\mathcal{H}$ had a single member, which was the complete graph. In the next sections, we will study the cases in which $\mathcal{H}$ had a single member and relate these problems to graph coloring and use Turán densities to show existence of many forbidden structures.

For now, we'll discuss a few generalisations of the problems in different directions. We've asked how many edges an $\mathcal{H}$-free graph can have. We could also have asked how many triangles such a graph can have. Phrased generally:

Densities (part 3):

For a family $\mathcal{C}$ of graphs who's appearance we count, and a family $\mathcal{H}$ of graphs we forbid, we can ask for the maximum number of appearances of graphs of $\mathcal{C}$ as subgraphs of an n-vertex graph containing no graph of $\mathcal{H}$. We denote this number $ex(n, \mathcal{C}, \mathcal{H})$.

When $\mathcal{C}$ is a single graph A, we write $ex_A(n, \mathcal{H})$ for the maximum number of subgraphs A an $\mathcal{H}$-free n-vertex graph can have.

We can then define the **Turán A-density at** n as $\pi_A(n, \mathcal{H}) = \frac{ex_A(n, \mathcal{H})}{\binom{n}{|A|}}$, which is the ratio of maximum number of subgraphs A an $\mathcal{H}$-free n-vertex graph can have to the total possible number of subgraphs an n-vertex graph can have.

As we did before, we can consider the asymptotic behavior of Turán densities:

Densities (part 4):

We denote the **Turán A-density** by $\pi_A(\mathcal{H}) = \lim_{n\to\infty} \frac{ex_A(n,\mathcal{H})}{\binom{n}{|A|}}$.

The limits exists due to a similar argument as before.
Again, for G attaining $ex_A(n+1,\mathcal{H})$, any n-vertex subgraph S of G is $\mathcal{H}$-free, so that the number of times $m_A(S)$ we find A is S is $m_A(S) \leqslant ex_A(n,\mathcal{H})$. Summing these bounds over the $n+1$ possibilities of S, a copy of A as subgraph is counted $n+1-|A|$ times: it's counted by all S except for those that miss a vertex of the copy of A, this missing vertex completely determining S as it differs from G in a single vertex, so that there are exactly $|A|$ subgraphs S that won't count the copy of A, on for each vertex of the copy. Hence, $(n+1-|A|)ex_A(n+1,\mathcal{H}) = \sum_S m_A(S) \leqslant (n+1)ex_A(n,\mathcal{H})$, so that with identity $\binom{n+1}{|A|} = \frac{(n+1)}{(n-|A|+1)}\binom{n}{|A|}$, we see that the densities form a decreasing sequence that must converge.

To close of this section, we'll discuss a last notion of density, which highlights Turán densities nature as optimization problems. This will become important for the notion of flag algebras:

Densities (part 5):

The **graph denisty** of a graph S in G, denoted $p(S,G)$ is $\frac{\left|\left\{\binom{V_G}{|V_S|} : S \simeq G\left[\binom{V_G}{|S|}\right]\right\}\right|}{\binom{|V_G|}{|V_S|}}$ or in words, the ratio of the number of subgraphs of G that are (isomorphic to) S to the total number of subgraphs of G on the same number of vertices as S.

This is also called the **induced subgraph homeomorphism density**, is denoted $t_{ind}(S,G)$ and can be interpreted as the probabilitythat a uniformly random injection $V_S \to V_G$ preserves the adjacency of S, in G.

The relation to Turán densities is that $\pi_A(n,\mathcal{H}) = \max\left(p(A,G) : \begin{cases}|V_G| = n \\ G\ is\ \mathcal{H}\ free\end{cases}\right)$.

A specific graph density is the **edge density** of a graph, $\frac{|E|}{\binom{|V|}{2}}$, aka. the ratio of edges to total possible edges, which can be interpreted as $p(K_2,G)$ in this context.
We then also have reformulation $\pi(\mathcal{H}) = \pi_{K_2}(\mathcal{H})$.

1.1.4 The Erdős–Stone–Simonovits theorem

Turán densities have links to other fields of graph theory. Here's what we mean:

The Erdős–Stone–Simonovits theorem:

For a graph G with at least an edge, we have $\pi(G) = 1 - \dfrac{1}{\chi(G) - 1}$.

Here, $\chi(G)$ is the chromatic number of the graph, and the condition ensures that $\chi(G) \geqslant 2$.

Full proofs of this result are quite involved. We state it here to give you one more reason to be interested in determining Turán densities. In this section, we explain the connection between graph densities and graph coloring. A full proof would take to long and take us to far, so we'll develop the theory until we reach the original Erdös-Stone theorem, which we won't prove.

The connection can be visulaized as follows: given a graph H and coloring with r colors of it, picture grouping all vertices of the same color together. Since the vertices of same color can't have edges in common, this is a r-partite graphs with colors determining partition sets.

Another connection can be made through the following remark. For graphs G and H where H is a subgraph of G, we must have $\chi(H) \leqslant \chi(G)$, which rephrases under contraposition as saying that when $\chi(H) > \chi(G)$, G must be H-free.

So a graph H with chromatic number $\chi(H)$ forms a an $\chi(H)$-partite graph, and it can't be a subgraph of an $(\chi(H) - 1)$-partite graph, as we can color such graphs with $(\chi(H) - 1)$ or less colors.

The second latter remark leads us to $ex(n, K_{\chi(H)}) \leqslant ex(n, H)$, as the left part is attained by an $(\chi(H)-1)$-partite graph via Turán's theorem, which is H-free, so that the bound holds by definition of extremal numbers.

The first remark will lead us to $ex(n, H) \leqslant ex(n, K_{\chi(H)}(s))$ for some s, where $K_{\chi(H)}(s)$ is the s-blow-up of $K_{\chi(H)}$, which is obtained by replacing each vertex of $K_{\chi(H)}$ by s copies, that we connect to copies of other vertices if their respective vertices where connected in the initial graph. Hence, in general, $K_r(s)$ is a complete r-partite graph with each partition set of size s. To see that $ex(n, H) \leqslant ex(n, K_{\chi(H)}(s))$, note that H formed an $\chi(H)$-partite graph, which we can embed in complete $\chi(H)$-partite graphs, if the sizes of partition sets of the latter are large enough. To be precise, if $s \geqslant \max(|V_i|)$ where V_i denotes the partition sets (the color sets) of H, we can embed H in $K_{\chi(H)}(s)$ by assigning vertices of H of the same color to vertices of the same partition sets in $K_{\chi(H)}(s)$, of which there are enough to host them, finding the edges of H among in those of $K_{\chi(H)}(s)$, which contains all possible edges of this kind.
Now, in general, is A is a subgraph of B, $ex(n, A) \leqslant ex(n, B)$, since the A-free graph attaining $x(n, A)$ must also be B-free, for it would contain a copy of A in the copy of B, if it weren't B-free. So $ex(n, H) \leqslant ex(n, K_{\chi(H)}(s))$.

Summarising, we have $ex(n, K_{\chi(H)}) \leqslant ex(n, H) \leqslant ex(n, K_{\chi(H)}(s))$. Next, we have the following bound of $ex(n, K_{\chi(H)}(s))$ that will lead to a sandwich bound on $ex(n, H)$ involving $ex(n, K_{\chi(H)})$:

Original Erdös-Stone:

For $r \geqslant 2$, $s \geqslant 1$, and any $\varepsilon > 0$, we can find an integer N such taht for all $n \geqslant N$, $ex(n, K_r(s)) \leqslant ex(n, K_r) + \varepsilon n^2$.

Phrased differently, this is equivalent to saying that any n-vertex graph with more then $ex(n, K_r) + \varepsilon n^2$ edges does contain an $K_r(s)$ as subgraph. This is the result we don't show: a proof based on the regularity lemma can be found in Diestel's book *Graph Theory* and Zhao's notes *Graph Theory and Additive Combinatorics*, the latter of which also contains a second, different, proof.

We finally, have $ex(n, K_{\chi(H)}) \leqslant ex(n, H) \leqslant ex(n, K_r) + \varepsilon n^2$, so that we get $\pi(K_{\chi(H)}) \leqslant \pi(H) \leqslant \pi(K_{\chi(H)}) + \varepsilon$ in the limit, and since ε was arbitrary, this means $\pi(K_{\chi(H)}) = \pi(H)$.
By Turán's theorem, we know that $\pi(K_{\chi(H)}) = 1 - \dfrac{1}{\chi(G) - 1}$, so that the Erdös-Stone-Simonovits theorem follows.

A subtlety has to be pointed out here. So far, we considered subgraphs in the sense of induced subgraphs. In the argument of this section however, when embedding H in $K_r(s)$, we obtain it as a possibly un-induced subgraph. However, the upper-bound still holds due to the fact that the extremal number for general subgraphs is lower then that for induced subgraphs: indeed, induced subgraphs are general subgraphs, so that when forbidding general subgraphs, we also forbid the induced ones, hence each H-free graph in the general sense is also H-free in the induced sense, and in particular so is the H-free graph in the general sense that attains the extremal number for general subgraphs.

1.1.5 Supersaturation

We now discuss one more reason to determine Turán densities. In the process, we show an important property of graph densities that will relate to flag algebras: density averaging.

Turán densities not only allow us to estimate when a certain subgraph is guaranteed to appear under the presence of a certain number of another type of subgraphs, they can also provide a lower bound on the number of appearing subgraphs. This phenomenon appears sufficiently often in asymptotic extremal combinatorics that it has a fancy name:

Supersaturation:

For any $\varepsilon > 0$ and any graphs A and H, there is a integer N so that for any n-vertex graph G, where $n \geqslant N$, that has more then $(\pi_A(H) + \varepsilon)\dbinom{n}{|A|}$ copies of A as subgraphs, G has $\Omega\left(n^{|V_H|}\right)$ copies of H as subgraphs.

The proof of this result will use a property of density that will also be useful in the future development.

A first result related to limits is that for any $\varepsilon > 0$, we can find an N such that for $n \geqslant N$, we have $\pi_A(\mathcal{H}) \leqslant \dfrac{ex_A(n, \mathcal{H})}{\binom{n}{|A|}} \leqslant \pi_A(\mathcal{H}) + \varepsilon$, since this is a decreasing convergent sequence. So for an n-vertex graph with more then $(\pi_A(\mathcal{H})+\varepsilon)\binom{n}{|A|} \geqslant ex_A(n,\mathcal{H})$ copies of A, the definition of the extremal number guarantees us an appearance of a copy of H as subgraph. In terms of densities, the condition stated rephrases as $p(A,G) > \pi_A(H)+\varepsilon$.

Anticipating the future of the proof, we'll consider here the case of $\varepsilon := \frac{\varepsilon}{2}$ and $n := N$, saying that any N-vertex graph with more then $\left(\pi_A(\mathcal{H})+\frac{\varepsilon}{2}\right)\binom{N}{|A|}$ copies of A has a copy of H in it.

The density property we'll make use of is:

Density averaging:

For an n-vertex graph G and $|A| \leqslant k \leqslant n$, we have $p(A,G) = \frac{1}{\binom{n}{k}} \sum_{S\subseteq G, |S|=k} p(A,S)$.

By interpreting densities as probabilities, this turns out to be total probability. Indeed, if $p(A,G)$ is the probability of getting A as subgraph in G by choosing $|A|$ vertices at random, then by disjoining on A being a subgraph of k-vertex subgraphs of G, we have $p(A,G) = P(A \subseteq G) = \sum_{S\subseteq G,|S|=k} P(A\subseteq H \wedge H = S)$ and

$P(A \subseteq H \wedge H = S) = P(A\subseteq H | H = S)P(S) = p(A,S)\frac{1}{\binom{n}{k}}$.

Alternatively, we can use a counting argument and identities of binomial coefficients to show this.
In the sum, each copy of A in G is counted as a copy of A in S when S contains the vertices of A. There are $\binom{n-|A|}{k-|A|}$ ways to choose such S containing the vertices of A, which we can build by taking the vertices of A and adding $k-|A|$ vertices among $n-|A|$ available ones.
Hence, the number of times a copy of A in G is counted in the sum is $\binom{n-|A|}{k-|A|}$.
We can then use the identity $\binom{n}{k}\binom{k}{|A|} = \binom{n-|A|}{k-|A|}\binom{n}{|A|}$, which one can check by performing cancellations in $\frac{\binom{n}{k}\binom{k}{|A|}}{\binom{n}{|A|}} = \frac{\frac{n...(n-k+1)}{k!}\frac{k...(k-|A|+1)}{|A|!}}{\frac{n...(n-|A|+1)}{|A|!}} = \frac{(n-|A|)...(n-k+1)}{(k-|A|)!} = \binom{n-|A|}{k-|A|}$, to see that $p(A,G) = \frac{1}{\binom{n}{k}} \sum_{S\subseteq G,|S|=k} p(A,S)$ really does hold.

The plan is now the following vague formulation: from a lower bound $p(A,G) > \pi_A(H)+\varepsilon$, density averaging will tell us that not to few $p(A,S)$ can be below this bound. Since this bound guarantees the the existence of an H, we will get multiple H for each S for which $p(A,S)$ is above the bound. We will then account for the possible double-counting of the H obtained this way.

So, we pick an n-vertex graph G, where $n \geqslant N$ with at least $(\pi_A(H)+\varepsilon)\binom{n}{|A|}$ edges, or equivalently $p(A,G) > \pi_A(H)+\varepsilon$. We then use density averaging for $k = N$, so that the subgraphs S are N-vertex graphs. In the averaging sum, we disjoin on the S for which $p(A,S) > \pi_A(H)+\frac{\varepsilon}{2}$.

They must in fact make up probability mass $\geqslant \frac{\varepsilon}{2}$, or equivalently, $\geqslant \frac{\varepsilon}{2}\binom{n}{N}$ of the S must satisfy $p(A,S) > \pi_A(H) + \frac{\varepsilon}{2}$. Otherwise, if p is the probability that $p(A,S) > \pi_A(H) + \frac{\varepsilon}{2}$, so that we assume $p \leqslant \frac{\varepsilon}{2}$, then by bounding $p(A,S) - \pi_A(H) \leqslant 1$ for those S and $(1-p) \leqslant 1$ for the other S, we can bound $p(A,G) - \pi_A(H) = \frac{1}{\binom{n}{N}} \sum_{S \subseteq G, |S|=N} (p(A,S) - \pi_A(H)) \leqslant p.1 + (1-p)\frac{\varepsilon}{2} \leqslant \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$, which contradicts $p(A,G) > \pi_A(H) + \varepsilon$.

So we get at least $\frac{\varepsilon}{2}\binom{n}{N}$ N-vertex subgraphs of G which have $p(A,S) > \pi_A(H) + \frac{\varepsilon}{2}$, so that by our choice of N, each must contain a copy of H, which is also a copy of H in G.

Did we over-count ?
One such copy of H can be produced this way by at most $\binom{n-|V_H|}{N-|V_H|}$ subgraphs S. Indeed, we enumerate these subgraphs by considering the vertices of H, then choosing the remaining $N - |V_H|$ among the $n - |V_H|$ available ones.
This means that we have the guarantee of at least $\frac{\frac{\varepsilon}{2}\binom{n}{N}}{\binom{n-|V_H|}{N-|V_H|}}$ copies of H in G.
We'll now polish this bound to a $\frac{\varepsilon}{2}\frac{(n...(n-N+1))((N-|V_H|)!)}{(N!)((n-|V_H|)...(n-N+1))} = \frac{\varepsilon}{2}\frac{n...(n-|V_H|+1)}{N...(N-|V_H|+1)} = \Omega\left(n^{|V_H|}\right)$.

1.2 More topics in terms of densities

1.2.1 The Erdös pentagon problem

1.2.2 Monotone subsequences in permutations

1.2.3 Ramsey theory via densities

2 Flag Algebras in the context of graphs

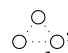
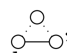
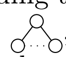
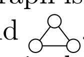
2.1 Mantel's theorem via densities and their limits

We will now show the asymptotic version of Mantel's theorem using densities. The proof technique we devellop is what inspired Razborov to devellop flag algebras. We'll devellop notions and tools useful in subsequent chapters.

Recall density averaging:

Density averaging:

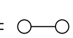
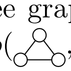
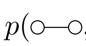
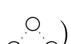
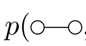
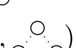
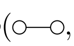
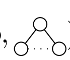
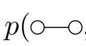
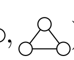
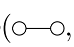
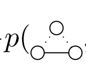
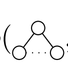
For an n-vertex graph G and $|A| \leqslant k \leqslant n$, we have $p(A,G) = \frac{1}{\binom{n}{k}} \sum_{S \subseteq G, |S|=k} p(A,S)$.

The subgraphs S on k vertices can be sorted into equivalence classes according to graph isomorphism. For example, for triangles, there are four classes, corresponding to: , , , and .
Let's call F_k the set of these classes and denote by $C \in F_k$ one such class. the point is that for subgraphs of G S and S' in the same class C, $p(A,S) = p(A,S')$, as any copy of A in one corresponds to a copy of A in the other, through the isomorphism that lets S and S' be in the same class C. We can therefore write $p(A,C)$ for these densities.

Now, in the sum $\sum_{S \subseteq G, |S|=k} p(A,S)$, we can sort subgraphs according to their isomorphims class, so as to get $\sum_{C \in F_k} \sum_{S \subseteq G, S \in C} p(A,S)$, which is $\sum_{C \in F_k} \sum_{S \subseteq G, S \in C} p(A,C)$ with our notation. We then see that we can factor out the term in the last sum, so that $\sum_{C \in F_k} \sum_{S \subseteq G, S \in C} p(A,C) = \sum_{C \in F_k} p(A,C) |\{S : S \subseteq G, S \in C\}|$.
The term $p(C,G) = \frac{1}{\binom{n}{k}} |\{S : S \subseteq G, S \in C\}|$ can be interpreted as the density of any graph representing C in G. Then, we get $p(A,G) = \frac{1}{\binom{n}{k}} \sum_{S \subseteq G, |S|=k} p(A,S) = \sum_{C \in F_k} p(A,C) p(C,G)$.

Chain rule:

For an n-vertex graph G and $|A| \leqslant k \leqslant n$, we have $p(A,G) = \sum_{C \in F_k} p(A,C) p(C,G)$.

In the context of Mantel's theorem, we can apply this chain rule to get interesting results.
For any triangle free graph G, we can express the edge density with the chain rule by using $A =$  and $k = 3$. Then, $p($$, G) = 0$, since G was assumed triangle-free, and we can compute edge densities $p($, $) = 0$, $p($, $) = \frac{1}{3}$, and $p($, $) = \frac{2}{3}$ and, even if it's not relevant $p($, $) = 1$.
This yields $p($$, G) = \frac{1}{3} p($$, G) + \frac{2}{3} p($$, G)$ for any triangle-free graph on more then 3 vertices.

We'll investigate the density $p(\overset{\circ}{\circ\cdots\circ}, G)$ by counting the occurrences of $\overset{\circ}{\circ\cdots\circ}$ at each vertex, using similar arguments to other proofs of Mantel's theorem.
For a given vertex v, we can produce a $\overset{\circ}{\circ\cdots\circ}$ by choosing two edges from its neighbourhood: these vertices won't be connected, as this third edge would close a triangle. Conversely, each $\overset{\circ}{\circ\cdots\circ}$ was counted this way, once for its unique vertex of degree 2. Therefore $p(\overset{\circ}{\circ\cdots\circ}, G) = \frac{1}{\binom{n}{3}} \sum_{v \in V_G} \binom{\deg_G(v)}{2}$.

We will however reformulate this passage to local quantities in terms of densities. For a fixed vertex v of G, we can ask for densities of subgraphs that contain v. In our context, we can ask for $p(\overset{\bullet}{\circ\cdots\circ}, G)$, which is the ratio of the number of occurence of $\overset{\bullet}{\circ\cdots\circ}$, where the full vertex represents v, to the number of all possible 3-vertex subgraphs that use v. Alternatively, $p(\overset{\bullet}{\circ\cdots\circ}, G)$ is the probability of choosing two vertices other then v uniformly, and getting induced subgraph $\overset{\bullet}{\circ\cdots\circ}$. The notation $p(\overset{\bullet}{\circ\cdots\circ}, G)$ can be misleading, as it depends on the vertex v, which will become clear in a moment. We won't use the following notation, but we could have written $p(\overset{\bullet}{\circ\cdots\circ}_v, G_v)$ to emphasise that we're counting subgraphs in which the black vertex is the vertex v of G.

We can compute this by noting that these subgraphs are built by choosing a first vertex of v's neighbourhood with probability $\frac{\deg_G(v)}{n-1}$ and a second one with probability $\frac{\deg_G(v)-1}{n-2}$,
so that $p(\overset{\bullet}{\circ\cdots\circ}, G) = \frac{(\deg_G(v))(\deg_G(v)-1)}{(n-1)(n-2)}$. Now, with our first definition, $p(\overset{\bullet}{\circ\cdots\circ}, G)\binom{n-1}{2}$ should be the number of occurrences of $\overset{\bullet}{\circ\cdots\circ}$, since $\binom{n-1}{2}$ is the number of all possible 3-vertex induced subgraphs that use v. And indeed, we have $\binom{\deg_G(v)}{2} = p(\overset{\bullet}{\circ\cdots\circ}, G)\binom{n-1}{2}$, as one can check to be coherent with $p(\overset{\bullet}{\circ\cdots\circ}, G) = \frac{(\deg_G(v))(\deg_G(v)-1)}{(n-1)(n-2)}$ by simplifying the expression.

So we finally get identity $p(\overset{\circ}{\circ\cdots\circ}, G) = \frac{1}{\binom{n}{3}} \sum_{v \in V_G} p(\overset{\bullet}{\circ\cdots\circ}, G)\binom{n-1}{2} = \frac{3}{n} \sum_{v \in V_G} p(\overset{\bullet}{\circ\cdots\circ}, G)$.

We'll take a brief moment to mention how this will be generalised in the context of flag algebras. We will soon define and discuss σ-flags, which are partially labeled graphs will labels corresponding to σ. We will then discuss the notion of densities of σ-flags, which are the probabilities of finding a certain graph labeled graph within another labeled graph, where labels are preserved.
Here, σ is just a single vertex v, and $p(\overset{\bullet}{\circ\cdots\circ}, G)$ can be thought of as a density of σ-flags, where the label is the vertex v of G that we fix.
This is where the name "flag" comes from: we picture σ as fixed, like a flag pole, and enumeration of subgraphs with fixed σ as a waving flag attached to the pole. More on that later.

Another identity we can reformulate in terms of densities is $\sum_{v \in V_G} \deg(v) = 2|E|$.
By the definition of edge densities, we have $|E| = p(\circ\!-\!\circ, G)\binom{n}{2}$.
We can also express $\deg(v)$ with a density: $p(\bullet\!-\!\circ, G)$, the ratio of edges incident to v to the number of all possible edges that could be incident to v, which is $(n-1)$ (one for each other vertex), or alternatively, the probability that a different vertex the v forms an edge as an induced subgraph with v.
Then, $\deg(v) = p(\bullet\!-\!\circ, G)(n-1)$.
So we get one more identity in terms of densities, $\sum_{v \in V_G} p(\bullet\!-\!\circ, G) = p(\circ\!-\!\circ, G)n$.

Now comes a reward for our reformulations in terms of densities.

The point of introducing densities is that for large graphs, things get blury and certain relations between densities start to hold. One such relation is that for n large enough, $p(\circ\!\overset{\bullet}{\wedge}\!\circ, G) \approx p(\bullet\!-\!\circ, G)^2$. Indeed, since G is triangle free, the first is just the probability of having two edges, which is almost that the probability of having two edges chosen independently with repeats, when G is large and there are many edges, so that picking the same is unlikely.
This would imply the identity $p(\circ\!\overset{\circ}{\wedge}\!\circ, G) \approx \dfrac{3}{n} \displaystyle\sum_{v \in V_G} p(\bullet\!-\!\circ, G)^2$. For now, let's see where this identity takes us: if it's in the right direction, then we'll try to make it rigorous. We will thus continue with this identity as an equality, for some sort of asymptotic behaviour as $n \to \infty$.

We can now combine this with the other identity we derived with the classic inequality trick which consists in recognizing that $\displaystyle\sum_{v \in V_G} p(\bullet\!-\!\circ, G)^2$ looks a like the square of a euclidean norm, and in that context, $\displaystyle\sum_{v \in V_G} p(\bullet\!-\!\circ, G)$ looks like a dot-product with the all-1 vector, so that Cauchy-Schwarz provides us with $n \displaystyle\sum_{v \in V_G} p(\bullet\!-\!\circ, G)^2 \geqslant \left(\sum_{v \in V_G} p(\bullet\!-\!\circ, G) \right)^2$ (since $\displaystyle\sum_{v \in V_G} 1 = n$). Finally, since $\left(\displaystyle\sum_{v \in V_G} p(\bullet\!-\!\circ, G) \right)^2 = p(\circ\!-\!\circ, G)^2 n^2$ by our second identity, we find that for large n, we have $p(\circ\!\overset{\circ}{\wedge}\!\circ, G) \geqslant 3p(\circ\!-\!\circ, G)^2$.

We can use the latter together with the very first relation we derived, stated as $p(\overset{\circ}{\circ\!-\!\circ}, G) + 2p(\circ\!\overset{\circ}{\wedge}\!\circ, G) = 3p(\circ\!-\!\circ, G)$, and the fact that $p(\overset{\circ}{\circ\!-\!\circ}, G) \geqslant 0$, to get $0 \leqslant 3p(\circ\!-\!\circ, G) - 6p(\circ\!-\!\circ, G)^2$, which implies that $0 \leqslant 1 - 2p(\circ\!-\!\circ, G)$, since $0 \leqslant p(\circ\!-\!\circ, G)$, which is equivalent to $p(\circ\!-\!\circ, G) \leqslant \dfrac{1}{2}$.
Now, $p(\circ\!-\!\circ, G) \leqslant \dfrac{1}{2}$ is precisely the asymptotic version of Mantel's theorem.

So the limit-identities payed of. Now, let's try to make out asymptotics work.
Let's start with some evidence for n large enough, $p(\circ\!\overset{\bullet}{\wedge}\!\circ, G) \approx p(\bullet\!-\!\circ, G)^2$.
Formally $p(\circ\!\overset{\bullet}{\wedge}\!\circ, G) = \dfrac{\deg(v)(\deg(v)-1)}{(n-1)(n-2)}$ and $p(\bullet\!-\!\circ, G) = \dfrac{\deg(v)}{(n-1)}$, so that asymptoticly, when $\deg(v) \xrightarrow[\infty]{n} \infty$, we get $\dfrac{\deg(v)(\deg(v)-1)}{(n-1)(n-2)} \sim_{n\to\infty} \left(\dfrac{\deg(v)}{(n-1)}\right)^2$, which translates to $p(\circ\!\overset{\bullet}{\wedge}\!\circ, G) \sim_{n\to\infty} p(\bullet\!-\!\circ, G)^2$.

In fact with a better analysis, we weaken our assumptions.
To see how big $|p(\circ\!\overset{\bullet}{\wedge}\!\circ, G) - p(\bullet\!-\!\circ, G)^2|$ can get, we study the function $f : x \mapsto \dfrac{x(x-1)}{(n-1)(n-2)} - \dfrac{x^2}{(n-1)^2}$ who's derivative is $f' : x \mapsto \dfrac{2x-1}{(n-1)(n-2)} - \dfrac{2x}{(n-1)^2}$ which is negative from 0 to $\dfrac{(n-1)}{2}$, and positive from $\dfrac{(n-1)}{2}$ to $n-1$. Since f has its roots at 0 and $(n-1)$ this means that f is negative, with minimum $\dfrac{1}{4}\left(\dfrac{(n-3)}{(n-2)} - 1\right)$ attained in $\dfrac{(n-1)}{2}$, we deduce that $|p(\circ\!\overset{\bullet}{\wedge}\!\circ, G) - p(\bullet\!-\!\circ, G)^2| \leqslant \dfrac{1}{4}\left(1 - \dfrac{(n-3)}{(n-2)}\right)$, which tends to 0 as $n \to \infty$. Therefore, our additional assumption on the degrees of the graph was unnecessary, letting n be large is enough.

Now, lets move on to $p(\overset{\circ}{\wedge}_{\circ\cdots\circ}, G) \approx \dfrac{3}{n} \displaystyle\sum_{v \in V_G} p(\bullet\!-\!\circ, G)^2$.

We can bound $\left| \dfrac{3}{n} \displaystyle\sum_{v \in V_G} p(\overset{\bullet}{\wedge}_{\circ\cdots\circ}, G) - \dfrac{3}{n} \sum_{v \in V_G} p(\bullet\!-\!\circ, G)^2 \right| \leqslant 3 \max_v \left| p(\overset{\bullet}{\wedge}_{\circ\cdots\circ}, G) - p(\bullet\!-\!\circ, G)^2 \right|$, where the latter is bounded by $\dfrac{1}{4}\left(1 - \dfrac{(n-3)}{(n-2)}\right)$, so that their distance really does go to 0.

However, the fact that for distances go to 0 as $n \to \infty$ is not enough to carry out our proof.
To get $p(\overset{\circ}{\wedge}_{\circ\cdots\circ}, G) \geqslant 3p(\circ\!-\!\circ, G)^2$, we assumed that the left of inequality $\dfrac{1}{n} \displaystyle\sum_{v \in V_G} p(\bullet\!-\!\circ, G)^2 \geqslant p(\circ\!-\!\circ, G)^2$ would also hold when replacing the left with its asymptotic equivalent.
This is not true in full generality: the sequences $\left(1 + \dfrac{1}{n}\right)_n$ and $\left(1 - \dfrac{1}{n}\right)_n$ may be equivalent, and $\left(1 + \dfrac{1}{n}\right) \geqslant 1$, but for any n, we do not have $\left(1 - \dfrac{1}{n}\right) \geqslant 1$.

However, inequalities are maintained by their limits if the sequences converge!
This leads us to investigate the convergence of our densities. So far, all we cared about was having a sequence of triangle-free n-vertex graphs. Such sequences exist: we can consider bipartitie graphs, for example.

Now, we seek a sequence of triangle-free n-vertex graphs for which the densities converge.
First, we can note that densities are in $[0, 1]$, so that Bolzano-Weierstrass guarantees us that we can extract a convergent subsequence. However, we want limits for all densities considered, those of $\circ\!-\!\circ$, $\triangle$, $\overset{\circ}{\wedge}_{\circ\cdots\circ}$, $\overset{\circ}{\circ\!-\!\circ}$ and $\overset{\circ}{\circ\cdots\circ}$ and this for the same sequence of graphs. Since this family of subgraphs we care about is finite, we can just carry out successive extraction from successive application of Bolzano-Weierstrass an finite number of times to get this done.

However, for a more general context, we note that the family of graphs $\mathcal{G}$ is countable. Their densities for a sequence of triangle-free n-vertex graphs G_n can be considered as a sequence in $[0, 1]^{\mathcal{G}}$. We may then use Tychonoff's theorem to extract a convergent sequence from the compact metric space $[0, 1]^{\mathcal{G}}$ with metric $d(x, y) = \sup\limits_{H \in \mathcal{G}} |x(H) - y(H)|$. For this metric, all coordinate sequence converge too.

So for any graph H, we can denote by $\phi(H)$ the limit of $p(H, G_{\varphi(n)})$, where $G_{\varphi(n)}$ is the limit of triangle-free $\varphi(n)$-vertex graphs obtained from the extraction, where $\varphi(n) \xrightarrow[\infty]{n} \infty$.

So, $p(\overset{\circ}{\wedge}_{\circ\cdots\circ}, G)$ converges, and therefore so does $\dfrac{3}{n} \displaystyle\sum_{v \in V_G} p(\bullet\!-\!\circ, G)^2$, to the same limit $\phi(\overset{\circ}{\wedge}_{\circ\cdots\circ})$. We may then pass to the limit in the inequalities that followed from Cauchy-Schwartz to get $\phi(\overset{\circ}{\wedge}_{\circ\cdots\circ}) \geqslant 3\phi(\circ\!-\!\circ)^2$. We can also pass to the limit in $p(\overset{\circ}{\circ\!-\!\circ}, G) + 2p(\overset{\circ}{\wedge}_{\circ\cdots\circ}, G) = 3p(\circ\!-\!\circ, G)$. Finally, note that for all graphs H, $\phi(H) \in [0, 1]$. So by the same arguments as before we can obtain the asmyptotic version of Mantel's theorem: $\phi(\circ\!-\!\circ) \leqslant \dfrac{1}{2}$.

When we say "asmyptotic version of Mantel's theorem" we actually mean a result that is much weaker then Mantel's theorem. $\phi(\circ\!-\!\circ) \leqslant \dfrac{1}{2}$ only states that there is a sequence of triangle-free graphs of increasing vertex number, who's edge density limit is less then $\dfrac{1}{2}$.

However, we can recover a better result from this. We start by noting that the initial sequence of triangle-free n-vertex graphs we considered was completely arbitrary. What if we let G_n be a sequence of graphs attaining the maximum number of edges over triangle-free n-vertex graphs ?

Their edge densities are $\dfrac{ex(n, \triangle)}{\binom{n}{2}}$. We know that this sequence converges to the Turán density $\pi(\triangle)$.

So when we extract a subsequence to get convergence for the densities of any subgraph, we get $\phi(\circ\!\!-\!\!\circ) = \pi(\triangle)$, as these limits must be the same, since the first is the limit of a subsequence of an already convergent sequence.

Thus for such an extremal initial sequence, we recover the result $\pi(\triangle) \leqslant \dfrac{1}{2}$, which is more deserving of the name of "asmyptotic version of Mantel's theorem".

This also highlights that $\phi(H)$ is quite ambiguous: it depends on the initial sequence of graphs chosen, as well as on the extraction we performed.

As a final note, we just mention that we used the axiom of choice when considering the sequence G_n of graphs attaining the maximum number of edges over triangle-free n-vertex graphs, as for all n, there exists such a graph, and we chose one for our sequence.

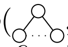
Now, let's draw some conclusions.

This proof of Mantel's theorem in its asymptotic version was very cumbersome. However, it highlights many key ideas we'll develop in flag algebras. A brief list of the take-aways of this proof are:

- We're interested in bounding densities
- We do so by deriving relation among densities
- We consider densities in which induced subgraphs can't range over the full graph anymore, but are restricted to having certain labeled vertices (subgraphs) in them.
- Some relations between densities arise in the limit, when we consider graphs of a given family with a large number of vertices. To this end, we have to select a sequence of the family of graphs of increasing order, for which all density limits exist.
- The limits depend on the sequence of graphs chosen: for the right choices, we can get results on extremal density limits.

In the context of flag algebras, we'll develop more systematic techniques of bounding densities, that contrast with the patch-work-proof we just gave. These systematic techniques make flag algebras a useful tool in combinatorics, that has allowed for new results to be discovered.

2.2 Flags and their densities

As we've seen in the previous section, it's of interest to consider both the densities of subgraphs with no constraints on vertices, such as $p($$, G)$, as well as the densities of subgraphs in which some vertices are fixed, such as $p($●—○$, G)$ for a fixed $v \in V_G$. We will now generalise these notions, and make them a bit more rigorous with clarifying notation.

Flags:

A **σ-flag** is a pair (F, θ) of a graph F and a embedding θ of a graph σ in F. Formally, θ is a injection from V_σ to V_F such that the graph induced by V_σ in F, $F[\theta(V_\sigma)]$, is isomorphic to σ via θ. We think of it as a partially vertex-labeled graph, where the labels originate from the image of V_σ under θ. We refer to $\theta(V_\sigma)$ as the flag-pole and to $V_F \backslash \theta(V_\sigma)$ as the flag-cloth.

We denote their set with $\mathcal{F}^\sigma$, and specify the slices $\mathcal{F}^\sigma_n$ made of σ-flags (A, θ) where A is an n-vertex graph.

Often, we consider flags in a forbidden subgraph context. If $\mathcal{H}$ is a family of graphs we forbid, σ is $\mathcal{H}$-free, and in a σ-flag (F, θ) we constrain F to be $\mathcal{H}$-free.

We consider two σ-flags (A, θ) and (B, μ) to be the same, i.e. to be ismorphic, if there exists a label preserving isomorphism between them, in the sense that there is an isomorphism of graphs $\gamma : A \simeq B$ such that $\gamma \circ \theta = \mu$.

Of particular interest are $\emptyset$-flags, where $\sigma = \emptyset$, which we use to consider unlabeled graphs.

If the last statement got you frowning, we'll recall some foundational stuff for a moment.
There exists at least one map $\theta : \emptyset \to V_F$. Indeed, in the standard set theoretic foundations, a map is a set of pairs of elements of the domain and the codomain, such that for all elements in the domain, there exists an element in the codomain such that these elements for a pair. Now, since $\forall x \in \emptyset, ...$ is true, the set $\emptyset = \emptyset \times V_F$ qualifies as a map $\theta : \emptyset \to V_F$.

All $\emptyset$-flags are isomporphic precisely when their graphs are isomorphic, as to contradict $\gamma \circ \theta = \mu$, there has to exist an $i \in \emptyset$ for which $\gamma \circ \theta(i) \neq \mu(i)$, but the first part of this statement is already false.

Before moving to densities, we dicuss some technicallities.
As in the previous section, we will consider densities for sequences of flags who's graphs having increasing order. This raises the question is such sequences exist, depending on what family of graphs $\mathcal{H}$ we forbid. We call σ non-degenerate wrt. $\mathcal{H}$, if for $n \geqslant |V_\sigma|$, the sets $\mathcal{F}^\sigma_n$ are non-empty.

In the case of triangle free graphs, we noted that bipartite graphs allowed us to consider sequence of triangle free graphs of increasing order. In the discussion that follows, we assume that σ non-degenerate wrt. $\mathcal{H}$. Remeber however that whenever we apply the theory of flag algebras to concrete cases, we have to explicitly justify this step !

It's possible to construct pathological non-degenerate cases: for example, if we first pick graph σ and then let $\mathcal{H}$ be the family of graphs with more then $|V_\sigma| + 1$ that contain σ as an induced subgraph, then σ is degenerate wrt. $\mathcal{H}$. Indeed, σ is $\mathcal{H}$-free, since it has less vertices then all of those of $\mathcal{H}$. But already starting from $n \geqslant |V_\sigma| + 1$, the sets $\mathcal{F}^\sigma_n$ will actually all be empty, as they can't contain a σ-flag (A, θ), since A would then contain a copy of σ induced by $\theta(V_\sigma)$, and since they have $n \geqslant |V_\sigma| + 1$ vertices, this is the precise definition of A belonging to $\mathcal{H}$.

We can define the density of flags through the following concept.
For two σ-flags (A,θ) and (B,μ), we consider (A,θ) to be a **subflag** of (B,μ) if A is an induced subgraphs of B for some embedding $\gamma : V_A \to V_B$, so that $\gamma \circ \theta = \mu$. So an isomorphism of flags corresponds to the case where γ is bijective.
Now, the density of (A,θ) in (B,μ) can be defined to be the probability of an injective map $\gamma : V_A \to V_B$, for which $\gamma \circ \theta = \mu$, to be an embedding of A as a subgraph of B, under uniform probability on this constraint set. This corresponds to choosing $|V_A \backslash \theta(V_\sigma)|$ vertices in $V_B \backslash \gamma \circ \theta(V_\sigma) = V_B \backslash \mu(V_\sigma)$ uniformly at random, so that the event that these vertices, together with $\mu(V_\sigma)$, induce A as a subgraph of B, occurs.

We'll generalize this notion in a moment: for now, we just use it to make the connection to the previous section. Indeed, $p(\bullet\!-\!\circ, G)$ can now be interpreted as a density of flags!
Here we consider $\sigma = ([1], \emptyset)$, the single vertex graph, and σ-flags $(\bullet\!-\!\circ, \theta)$ and (G,μ), where θ maps 1 to the black vertex, and μ maps 1 to the vertex v. Previously, we just said that the black vertex was v, and $p(\bullet\!-\!\circ, G)$ was the density of edges in G with v as one of their endpoints.
Now, let's check that the generalised definition of density can recover the previous one. We now seek the probability of choosing $|V_{\bullet\!-\!\circ} \backslash \theta(V_\sigma)| = 1$ vertices in $V_G \backslash \mu(V_\sigma) = |V_G| - 1$, so that together with $\mu(V_\sigma) = v$, this vertex induces $\bullet\!-\!\circ$ as a subgraph of G, which corresponds to their being an edge between the two vertices. So this probability is $\dfrac{\deg_G(v)}{|V_G|-1}$, as before.

We can generalise the notion of density to the following, by asking for simultaneous (but separate) appearance of substructures:

Density of σ-flags:

We consider t σ-flags F_i of order n_i and a σ-flag (G,θ) of order n.
Next, we focus on the flags of sizes $l_i = n_i - |V_\sigma|$ and $l = n - |V_\sigma|$, and assume that $l \geqslant \sum_{i\in[t]} l_i$.

We now draw t sets of vertices among the flag-cloth $V_G \backslash \theta(V_\sigma)$ of sizes l_i unifromily and independently. We denote with A_i the event for the ith set S_i chosen, $G[S_i \cup \theta(V_\sigma)] \simeq \sigma$, or in words, the vertices S_i, together with the flag-pole $\theta(V_\sigma)$ induce σ as a subgraph of G, i.e. F_i is a subflag of G.
We denote with B the event that all sets are pairwise disjoint.

Now, the density of flags F_i in G is defied to be $p(F_1, ..., F_t|G) = P(A_1 \cap ... \cap A_t|B)$.

So the density of flags can be thought of a fixing a flag-pole and considering he likelihood of getting t specified subgraphs in a graph all overlapping precisely in this pole.

Asking for disjointness of the flag-cloths allows us to decompose flags in a sense.
In our running example of Mantel's theorem, we can show that, using our new notation , $p(\circ\!-\!\bullet\!-\!\circ|G) = p(\bullet\!-\!\circ, \bullet\!-\!\circ|G)$. Indeed, the occurrence any two different $\bullet\!-\!\circ$, corresponds to the occurrence of a $\circ\!-\!\bullet\!-\!\circ$, due to there being no triangles in G. Now, technically each occurrence of a $\circ\!-\!\bullet\!-\!\circ$ should account for two occurrences of pairs of $\bullet\!-\!\circ$, as we distinguish order in our probability space.
However, we still get the same densities.
Indeed, when computing $p(\circ\!-\!\bullet\!-\!\circ|G)$ from our definition, B is a sure event as a single set is always disjoint from non-existent others. So it's the probability that a unordered pair of different vertices are both in the neighbourhood of the flag-pole, seeing as there are no triangles:

this is $\dfrac{\binom{\deg_G(\theta(1))}{2}}{\binom{n-1}{2}} = \dfrac{\deg_G(\theta(1))(\deg_G(\theta(1))-1)}{(n-1)(n-2)}$.
On the other hand, when computing $p(\bullet\!\!-\!\!\circ, \bullet\!\!-\!\!\circ|G)$, $P(B)$ is the probability that the two chosen singletons of vertices are different, which is $P(B) = 1 - \dfrac{(n-1)}{(n-1)^2} = \dfrac{(n-2)}{(n-1)}$. Next, $P(A_1 \cap A_2 \cap B)$ is the probability of getting two different vertices in the neighbourhood, where order now matters so that $P(A_1 \cap A_2 \cap B) = \dfrac{\deg_G(\theta(1))(\deg_G(\theta(1))-1)}{(n-1)^2}$. We the see that a $(n-1)$ is canceled in the fraction providing $p(\bullet\!\!-\!\!\circ, \bullet\!\!-\!\!\circ|G)$, so that we indeed find $p(\circ\!\!-\!\!\bullet\!\!-\!\!\circ|G) = p(\bullet\!\!-\!\!\circ, \bullet\!\!-\!\!\circ|G)$.

We will find another way of deriving this with the chain rule, in a moment
For now, we'd like to know how flag density, which was conditioned on having vertex sets disjoint, compares to the probability in which this isn't the case. Since we choose vertex sets independently, and each A_i depends only on set i, the A_i are independent, so that $P(A_1 \cap ... \cap A_t) = \prod_{i\in[t]} P(A_i) = \prod_{i\in[t]} P(A_i|B) = \prod_{i\in[t]} p(F_i|G)$. The first equality is due to independence, the second due to B being the sure event when a single set is considered, and the last one is due to the definition of densities.
Our goal is to see if the $p(\circ\!\!-\!\!\bullet\!\!-\!\!\circ, G) \approx p(\bullet\!\!-\!\!\circ, G)^2$ from Mantel's theorem generalises in this context.

Asymptotic independence:

With our previous notation:

$$\left| p(F_1, ..., F_t|G) - \prod_{i\in[t]} p(F_i|G) \right| \leqslant 2t^2 \left(1 - \left(\frac{l - 2\min_i(l_i) + 1}{l} \right)^{\min_i(l_i)} \right) \xrightarrow[\infty]{l} 0$$

First, let's get a big picture with $A = A_1 \cap ... \cap A_t$, so that $p(F_1, ..., F_t|G) = P(A|B)$ and $\prod_{i\in[t]} p(F_i|G) = P(A)$. The distance to be bound is $|P(A) - P(A|B)|$. By writing $P(A) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B})$, we note that $|P(A) - P(A|B)| = P(\overline{B})|P(A|B) + P(A|\overline{B})| \leqslant 2P(\overline{B})$ (where we used the fact that probabilities are in $[0,1]$ and the triangular inequality). Now, $\overline{B}$ is the event that there exist two indices $i \neq j$ such that the flag-cloths S_i and S_j to be chosen share at least one element. We may therefore use a union bound to get $P(\overline{B}) \leqslant \sum_{i\neq j\in[t]} P(C_{ij})$, where C_{ij} is the event that S_i and S_j intersect.

To bound $P(C_{ij})$, its easier to consider the opposite event, that S_i and S_j are disjoint. Now, to compute this probability, note that the number of ways we can choose S_i and S_j are $\binom{l}{l_i} \times \binom{l}{l_j}$, and that among these (ordered) pairs, we can count disjoint ones by enumerating the $\binom{l}{l_i + l_j}$ ways vertices make they up together, followed by the the $\binom{l_i + l_j}{l_i}$ ways we can form the first set (and therefore the second as complement) among these vertices, so that $P(C_{ij}) = 1 - \dfrac{\binom{l}{l_i+l_j} \times \binom{l_i+l_j}{l_i}}{\binom{l}{l_i} \times \binom{l}{l_j}}$. Simplifying the fraction leads to $\dfrac{l_j!((l-l_i) \times ... \times (l - l_i - l_j + 1))}{(l \times ... \times (l - l_j + 1))}$, which we may lower-bound by $\dfrac{(l - l_i - l_j + 1)^{l_j}}{l^{l_j}}$, and finally by $\left(\frac{l-2\min_i(l_i)+1}{l}\right)^{\min_i(l_i)}$. So each $P(C_{ij}) \leqslant 1-\left(\frac{l-2\min_i(l_i)+1}{l}\right)^{\min_i(l_i)}$, and since there are at most t^2 such pairs of indices (technically $t(t-1)/2$) we get the bound from the statement.

Now, when $l \to \infty$, $\frac{l-2\min_i(l_i)+1}{l} \to 1$, so that the bound tends to 0.

This is one of the key desired properties of densities verified. Next, we'll generalise the chain rule:

Chain rule:

We consider t σ-flags F_i of order n_i, and σ-flags G of order n.
Next, we focus on the flags of sizes $l_i = n_i - |V_\sigma|$ and $l = n - |V_\sigma|$, and assume that $l \geqslant \sum_{i\in[t]} l_i$ and for an $s \in [t]$ and an $r \leqslant l$, we have $r \geqslant \sum_{i\in[s]} l_i$. Then:

$$p(F_1, ..., F_t|G) = \sum_{F\in\mathcal{F}^\sigma_r} p(F_1, ..., F_s|F)p(F, F_{s+1}, ..., F_t|G)$$

3 Semidefinite Programming

4 Flag Algebras in the abstract context

5 More applications

5.1 The Rado multiplicity problem in vector spaces over finite fields

5.2 The maximum quartet distance between phylognentic trees