Conversational Combinatorics

Yves Jäckle
Version 1

Contents

1 Polyominos

Book in the 1001-folder

2 Distinct representatives and Hall's theorem

Consider a collection of n sets S_i that may intersect. Our goal is to find **distinct representatives** of this set, in the sense that we find distinct $s_i \in S_i$ for each set S_i. The trouble seems to arise when we pick these representatives without care: one set may run out of elements not already chosen as representatives for other sets, if it intersects with many other sets.

We can represent the problem as a bipartite graph, where the n sets S_i are represented by bipartition set U, and the elements of the "universe" $\cup_{i\in[n]}S_i$ are represented by bipartition set W. we include the edge $\{S_i, s\}$ if $s \in S_i$. Then finding a system of distinct representatives is equivalent to finding a matching of U into W, which is a set of edges with no common endpoints that covers all of U. *Under what assumptions is there such a matching ?*

Let's start by looking for some easy necessary conditions that such a matching must satisfy. Since we "inject" U into W, we must expect $|U| \leqslant |W|$. More precisely, we have $|U| \leqslant |N(U)|$ for the neighbourhood of U. In fact, this is true for any subset of U, so that for all $S \subseteq U$, we have $|S| \leqslant |N(S)|$.
The amazing thing is that this condition is already sufficient:

Hall's theorem:

If $|S| \leqslant |N(S)|$ for all $S \subseteq U$, then we can match U into W.

We'll try induction on the size of U by matching its elements one at a time.
In the initial case of $|U| = 1$, the condition says that the single element of u has a neighbour, that we can then match it with to get the desired result.
For the step, we'd like to match a vertex u of U to some $w \in W$ so that the condition $|S| \leqslant |N(S)|$ is maintained. The only way that $|S| \leqslant |N(S)|$ can be violated by such a move is if we had $|S| = |N(S)|$ for some S for which $w \in N(S)$ but $u \notin S$, as after deletion of u and w, we'd have $|S| > |N(S)|$, as $|N(S)|$ decreased by 1, but $|S|$ didn't.

In fact, the case in which $|S| < |N(S)|$ for all (non-empty proper) $S \subseteq U$ can be handled by taking any $u \in U$ and matching it to one of its neighbours, then deleting the pair and using induction to handle the remaining graph. To see that Hall's assumption is inherited by the remaining graph, take a set S for the initial graph and disjoin these cases. If S contained u, then $N(S)$ contained w, hence both sides of $|S| < |N(S)|$ decrease by 1 and the condition is maintained. If S didn't contain u, it may happen that $N(S)$ contains w, but even when $|N(S)|$ decreases by 1, we still have $|S| \leqslant |N(S)| - 1$ as they where integers. If $N(S)$ didn't contain w, then none of the sides of $|S| < |N(S)|$ decreased, and we keep the condition.

Now, how do we handle the problematic case of $|S| = |N(S)|$ for some $S \subseteq U$?
An idea is to use the induction hypothesis to match S into $N(S)$ and then match the rest $U\backslash S$ by induction again. The first problem in realising this approach is that to apply induction, S has to be proper and non-empty so as the sizes of S and $U\backslash S$ to be less then hat of U. This is where the assumption on S (in parentheses) in the previous paragraph came from. If the empty set and U are the only to satisfy $|S| = |N(S)|$, then we are in that case that we already handled.
The next step in successfully applying the induction assumption is to check Hall's condition. In the graph induced by the S and $N(S)$, for which $|S| = |N(S)|$, we also have $|S'| \leqslant |N(S')|$ for all $S' \subseteq S$, as they are also subsets of U. So we can match S into $N(S)$.
Now, we ask if the graph induced by $U\backslash S$ and $W\backslash N(S)$ has this property too. If some $S' \subseteq U\backslash S$ satisfied $|S'| > |N(S') \cap W\backslash N(S)|$, then by adding $|S| = |N(S)|$ on both sides, using disjointness of S' and S, and the fact that $N(S \cup S')$ can be partitioned into $N(S)$ and $N(S') \cap W\backslash N(S)$, we'd get contradiction

$|S \cup S'| > |N(S \cup S')|$ to the Hall condition on the initial graph.
Therefore, we may apply induction and all cases worked out.

COMPLETE: add the other proofs from Diestel ?

3 Chains, antichains and a Sperner's lemma

We're interested in the following two types of structures on a finite set X:

Chains and antichains:

A **chain** is a set $C \subseteq 2^X$ such that for all $S, U \in C$, we have $S \subseteq U$ or $U \subseteq S$.
An **antichain** is a set $A \subseteq 2^X$ such that for all $S \neq U \in A$, we have $S \subsetneq U$ and $U \subsetneq S$.

We'll first explain the name of chains with the following charcterisation:

Why they're called chains:

A chain C is a finite set $C \subseteq 2^X$ of form $C = \{C_1, .., C_k\}$ with $C_1 \subseteq C_2 \subseteq ... \subseteq C_k$.

Had we defined chains like this, the property we actually defined them with would have followed immediately, since $C_i, C_j \in C$, we have $C_i \subseteq C_j$ for $i < j$ and $C_j \subseteq C_i$ for $j < i$.
To see that chains are of the form we describe, picture a digraph who's vertices are the different sets of C, that we connect by an edge from S to U if $S \subseteq U$. The chain definition translates to this digraph being complete. A sequence of different sets $C_1 \subseteq C_2 \subseteq ... \subseteq C_k$ then translates to a dipath in the digraph. Indeed, a diwalk that would loop on itself would correspond to a sequence $C_1 \subseteq C_2 \subseteq ... \subseteq C_k \subseteq C_1$, so that all its sets must actually be equal, a case we excluded by considering one one per different set.
What can we about this graph ?
In our remark, we just shown that the digraph is acyclic. Such digraphs have a source (a vertex that is only the tail of edges) and a sink (a vertex that is only the head of edges). They are start and end of a longest dipath in the graph. Indeed, since we can't prolong such a maximum dipath, the vertices not on the dipath can't lead to its start, for otherwise, we could prolong our dipath along them. Also the vertices on the dipath can't lead back to the start, for we'd create a dicycle this way. So all edges can only lead into the start, which is herefore a source. A similar argument explains why the end is a sink.
If we delete the source and sink, aka. their corresponding sets from the chain, then we still have a complete directed acyclic graph on the remaining vertices, aka. we still have a chain. This leads us to conclude with induction. For chains of sizes 1 and 2, this is clear, and if a chain of size k or less can always be written as $\{C_1, .., C_k\}$ with $C_1 \subseteq C_2 \subseteq ... \subseteq C_k$, then when considering a chain of size $k+1$, we can delete its source (that we'll rename C_{k+1}), get a chain of size k in form $C_1 \subseteq C_2 \subseteq ... \subseteq C_k$, so that $C_1 \subseteq C_2 \subseteq ... \subseteq C_k \subseteq C_{k+1}$, since C_{k+1} is a source, and the induction step holds.

How long can chains be ?
The least we can "add" at each step in $C_1 \subset C_2 \subset ... \subset C_k$ is a single element. Intuitively, we can build a long chain by starting with the empty set $\emptyset$, and adding elements of X to the subset one at a time to make the sets of the chain. This produces chains of length $|X| + 1$.
This is actually the longest type of chains we can find. To see why this is the case, note that a longest chain must contain the empty set $C_1 = \emptyset$, as we could prolong it wih the empty set if it didn't, contradicting maximality, and that the cardinality of the $|C_i|$ must strictly increase, so that we can bound $|X| \geqslant |C_k| - 0 = \sum_{i=2}^{k} (|C_i| - |C_{i-1}|) \geqslant k - 1$, which rephrases to $k \leqslant |X| + 1$.

How long can antichains be ?
This is a more difficult question to answer. A first candidate for a large antichain could be the **r-sets** of X, which are the subsets of X of size r, for some $r \leqslant |X|$. They form an antichain of X, since they can't be included in one another, unless they're equal, due to having the same size.
There are $\binom{|X|}{r}$ r-sets of X, which is biggest for $r = \left\lfloor \frac{|X|}{2} \right\rfloor$ and $r = \left\lceil \frac{|X|}{2} \right\rceil$.
Again, we'd like to know if we can get bigger then in this construction.
The answer will be based on a study of the interaction between chains and antichains:

Chains-antichains interaction:

For any antichain A and any chain C, the chain C can contain at most one element of the antichain A. So, if we partition X into p chains PC_i, in the sense that the PC_i are chains, are pairwise disjoint, and so that any $S \subseteq X$ is in some PC_i, for some i, then we'll know that $|A| \leqslant p$.

This will be our strategy for showing $|A| \leqslant \binom{|X|}{\left\lfloor \frac{|X|}{2} \right\rfloor}$: we'll build a partition of size $p = \binom{|X|}{\left\lfloor \frac{|X|}{2} \right\rfloor}$ into chains. But first, we'll prove the claims we just made.

The first part is due to sets of a chain being included in one another, something that is precisely prohibited in antichains. The second part follows from the first, as all sets of A, which are subsets of X, are at least one chain, but two sets of an antichain can't be in the same chain, as this is prohibited from the first part, so that there can't be more chains then sets in the antichain, aka. $|A| \leqslant p$.

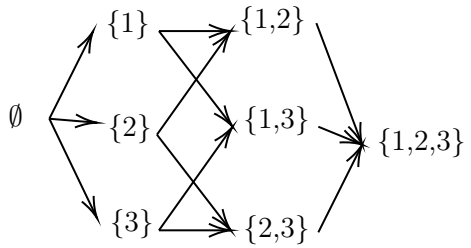
We're now ready for:

Sperner's lemma:

For any antichain A, we have $|A| \leqslant \binom{|X|}{\left\lfloor \frac{|X|}{2} \right\rfloor}$.

With the previous strategy, our goal is to find a partition using as few chains a possible to get an upper-bound that is as tight as possible on the size of an antichain. *How do we find such a partition ?*

To get few chains, we want our chains to be as long as possible. This can be done by choosing chains that add a single element at each subsets.
Let's seek inspiration in a figure of the situation. Below, we show the digraph representing the subsets of $X = \{1, 2, 3\}$, where edges represent inclusion, where the set at the head has exactly one element more that that at the tail. In this representation a subset is included in another if there is a path from one to the other. We've ordered the vertices by layers that correspond to set size of the subsets. In fact, we'll refer to layer r to describe the layer of r-sets.



A partition by chains corresponds to a partition into dipaths. To get long paths, aka. long chains, we'll try to build up them layer by layer, as follows. We start with a dipath $\{\emptyset\}$. Then, at each layer, we look at the previous one. If we're on layer r and there are less vertices on the previous layer (which corresponds to $r \leqslant \frac{|X|}{2}$), we take those that are the endpoints of the dipaths we had so far, and extend these paths, and we extend these dipaths by adding edges who's tails are in the previous layer and heads are in the current layer. If the edges we use for extension form a matching of the vertices of layer $r-1$ to layer r, then the dipaths will still be disjoint, as non of their new ends coincide.
We have to prove that such a matching exists. Let's try to get it from Hall's theorem, since the graph induced by layers $r-1$ and r is bipartite. If W is any set of nodes on layer $r-1$, and $\Gamma(W)$ is the set of its neighbours on layer r, then by double-counting edges as $\sum_{v\in W} \deg(v) = \sum_{v\in\Gamma(W)} \deg(v)$, where $\deg(v) = (|X| - (r-1))$ for $v \in W$, since we can add one of $|X| - (r-1)$ remaining elements of X to the $(r-1)$-set represented by v to get an r-set, and $\deg(v) = r$ for $v \in \Gamma(W)$, since we can delete one of r elements of the r-set represented by v to get an $(r-1)$ set, we get $\frac{|\Gamma(W)|}{|W|} = \frac{(|X| - (r-1))}{r} = \frac{|X|-1}{r} - 1$, so that $|W| \leqslant |\Gamma(W)| \Leftrightarrow r \leqslant \frac{|X|-1}{2}$, which is the case when $r \leqslant \frac{|X|}{2}$. So Hall's theorem applies and such a matching does indeed exist.
Now, there are still nodes on layer r hat aren't in a dipath it this stage. We'll simply let them be new one-vertex dipaths, and move to the next layer.
This construction works until we arrive at layer $\frac{|X|}{2}$ for even $|X|$, and layer $\left\lfloor \frac{|X|}{2} \right\rfloor$ for odd $|X|$. In the odd case, we have to match layers $\left\lfloor \frac{|X|}{2} \right\rfloor$ and $\left\lceil \frac{|X|}{2} \right\rceil$. Here, Hall still applies as $|W| = |\Gamma(W)|$, since $|X| - \left(\left\lceil \frac{|X|}{2} \right\rceil - 1 \right) = \left\lceil \frac{|X|}{2} \right\rceil$. Finally, we consider the layers $r > \left\lceil \frac{|X|}{2} \right\rceil$. There, we'll have $|W| \geqslant |\Gamma(W)|$, but we can still use Hall, this time matching the vertices of layer r into those of $r-1$ (so in the opposite orientation). This means that we can prolong some dipaths, while others will have to end at layer $r-1$.

The dipaths we built are disjoint and all vertices are contained in some dipath, which is in a sense a loop invariant of our implicit algorithm. We've therefore built a partition into chains this way. *How many of these dipaths/chains are there in our partition ?*
The crucial observation is that all paths pass through layer $\left\lfloor \frac{|X|}{2} \right\rfloor$. This allows us to conclude that there are $\binom{|X|}{\left\lfloor \frac{|X|}{2} \right\rfloor}$ paths, since all vertices $\binom{|X|}{\left\lfloor \frac{|X|}{2} \right\rfloor}$ of the layer are in some path, no path can contain more then one vertex per layer, and all paths have a vertex on that layer. To see that paths pass through that layer, note that paths where prolonged until that layer, so that they reach the layer in particular, or start there.

This allows us to conclude with Sperner's lemma, using the previous relation between partition size and antichain size.

4 Shadows, compression, and shifts

4.1 Workspace

We can measure proximity of sets by using the distance $d(A,B) = |A\Delta B|$. Indeed, this qantity is positive, symmetric and zero precisely if $A = B$. To see that the triannguular inequality holds, consider $A\Delta C$, which we want to relate to $A\Delta B$ and $B\Delta C$ and include $A\backslash C$ into $A\backslash B \cup B\backslash C$ by disjoining on whether elements are in B, and similarly for including $C\backslash A$ into $B\backslash A \cup C\backslash B$. So $A\Delta C \subseteq A\Delta B \cup B\Delta C$ by rearranging, and the size of the union being less then the sum of that of its parts. An intersting quantity related to a set family is its **diameter** $diam(\mathcal{F}) = \max\limits_{A,B\in\mathcal{F}} |A\Delta B|$.

The operation of **compressing/squashing** a family in a element x is that of forming $S_x(\mathcal{F})$, which is the image of $\mathcal{F}$ under $c_x : A \mapsto \begin{cases} A\backslash x, \ if \ x \in A \ and \ A\backslash x \notin \mathcal{F} \\ A, \ else \end{cases}$. It's the family obtained by by replacing A with $A\backslash x$ if it contained x and if $A\backslash x$ wasn't already in the family.

Compression maintains the size of the family, so that $|S_x(\mathcal{F})| = |\mathcal{F}|$. Indeed, the map is injective, so that there is a correspondence between sets: two different uncompressed sets won't have the same image, if a compressed set $c_x(A) = A\backslash x$ has the same image as an uncompressed $c_x(B) = B$, in the sense that $A\backslash x = B$, then we reach a contradiction in the fact that $A\backslash x \notin \mathcal{F}$ yet $B \in \mathcal{F}$, and finally, if a compressed set $c_x(A) = A\backslash x$ has the same image as a compressed set $c_x(B) = B\backslash x$, in the sense that $A\backslash x = B\backslash x$, then $A = B$ must both have been in $\mathcal{F}$, aka. they were the same set, so that injectivity holds.
So compression reduces the appearances of an element in the family, without affecting its size.

Another interesting property of compression is that it can only decrease the diameter of a family, so that $diam(S_x(\mathcal{F})) \leqslant diam(\mathcal{F})$. To see this, we disjoin cases on the images of A and B under compression. If none of them or both of them were compressed, then $c_x(A\Delta B) = c_x(A)\Delta c_x(B)$ stays the same. If only one of them, say A, was compressed, we disjoin cases on whether B contained x. If it didn't, $c_x(A)\Delta c_x(B) = (A\backslash x)\Delta B$ will decrease by x, as $x \in A\backslash B$, and we're good, but if it did, $c_x(A)\Delta c_x(B)$ will decrease by x, as $x \in B\backslash(A\backslash x)$. For this case, since B contained x but wasn't compressed, it must have been the case that $B' = B\backslash x \in \mathcal{F}$. Then the image can be bounded by $c_x(A)\Delta c_x(B) = (A\backslash x)\Delta B \subseteq (A\backslash x)\Delta B' = c_x(A)\Delta c_x(B')$. So in all cases $|c_x(A)\Delta c_x(B)|$ is bounded by $diam(\mathcal{F})$. Therefore the maximum of the left side satisfies the inequality too, which is what we wanted to prove.

Compression also decreases the trace of a family. The **trace** of family $\mathcal{F}$ on set T is the family $\mathcal{F}_{|T} = \{A \cap T | A \in \mathcal{F}\}$, so the intersctions of the sets of $\mathcal{F}$ with T. We have $|S_x(\mathcal{F})_{|T}| \leqslant |\mathcal{F}_{|T}|$. To see this, note that there is a surjection from $\mathcal{F}_{|T}$ to $S_x(\mathcal{F})_{|T}$ that to $A \cap T \in \mathcal{F}_{|T}$ associates $c_x(A) \cap T \in S_x(\mathcal{F})_{|T}$. By definition of $S_x(\mathcal{F})_{|T}$, all of its elements have an antecedent under this map.

A quick remark on repeated compression is that order matters! For example, $S_x(S_y(\{\{x,z\},\{y,z\})) = \{\{x,z\},\{z\}\}$ and $S_y(S_x(\{\{x,z\},\{y,z\})) = \{\{y,z\},\{z\}\}$, so they#re not the same.

We can use compression to form a **independence system** / a **hereditary family** from $\mathcal{F}$, denoted $\tilde{\mathcal{F}}$, which will have the propery that for $A \in \mathcal{F}$ and $B \subseteq A$, we must have $B \in \tilde{\mathcal{F}}$, aka. $\tilde{\mathcal{F}} = \cup_{A\in\mathcal{F}} 2^A$. The set system $\tilde{\mathcal{F}}$ is formed as the union of all $S_{x_1}(...S_{x_k}(\mathcal{F})..)$, for all $k \geqslant 0$ and all $x_i \in X$ where X is the ground set of the family (the case $k = 0$ can be understood as $\mathcal{F}$). It has the property that for $A \in \tilde{\mathcal{F}}$, and any $x \in A$, we have $A\backslash x \in \tilde{\mathcal{F}}$. This implies subset-stability, as we can get any subset by repeatedly

deleting elements (formally, we induct on $|A\backslash B|$ where $B \subseteq A$).
Now, to see the property, assume for contradiction that there is some $A \in \mathcal{F}$ and $B \subseteq A$ such that $B \notin \tilde{\mathcal{F}}$. We can the consider the inclusion maximal such subset of A that isn't in $\tilde{\mathcal{F}}$, which we'll call B too, ignoring the previous one. The case $B = A$ can be excluded as $A \in \mathcal{F} \subseteq \tilde{\mathcal{F}}$. All C in the sandwich $B \subsetneq C \subseteq A$ must be in $\tilde{\mathcal{F}}$, by maximality. So by taking any $x \in A\backslash B$, we know that we can write $B \cup x$ as $c_{x_1}(...c_{x_k}(F)..)$ for some $F \in \mathcal{F}$. Since $B \notin \tilde{\mathcal{F}}$ and $\mathcal{F} \subseteq \tilde{\mathcal{F}}$, we must have $B \notin \mathcal{F}$. So we have compression $c_x(B \cup x) = B$. Thus, B can be written as $c_x(c_{x_1}(...c_{x_k}(F)..))$, so that we actually end up with $B \in \tilde{\mathcal{F}}$.

Another way of forming a hereditary family $\mathcal{F}^{cpr}$ from $\mathcal{F}$ using compression is with the following algorithm. While there is an x in the groung set X such that $c_x(\mathcal{F}) \neq \mathcal{F}$, keep iterating on family $c_x(\mathcal{F})$. This loop will terminate, as $c_x(\mathcal{F}) \neq \mathcal{F}$ requires there to be an $A \in \mathcal{F}$ such that $c_x(A) = A\backslash x$, so that the size of one of the sets of the family decreases in each iteration, which may happen only finitely many times. At termination, we end up with $\mathcal{F}^{cpr} = c_{x_k}(...c_{x_1}(\mathcal{F}))$ where we had k iterations and compressed along the sequence $x_1, ..., x_k$. This family has the property that $c_x(\mathcal{F}^{cpr}) = \mathcal{F}^{cpr}$ for all $x \in X$, meaning that for any $x \in \bigcup_{A \in \mathcal{F}^{cpr}} A$, we must have $A\backslash x \in \mathcal{F}^{cpr}$. This is how we can clonclude that all subsets of set in $\mathcal{F}^{cpr}$ are also in $\mathcal{F}^{cpr}$: delete elements one at a time until reaching the subset, repeatedly using argument $A\backslash x \in \mathcal{F}^{cpr}$ to deduce that the sets where all in $\mathcal{F}^{cpr}$.

A different operation is **shifting**. It corresponds to replacing a particular element by another. We apply the map $s_{ij}(A) = \begin{cases} (A\backslash j) \cup i, \ if \ j \in A, i \notin A, \ (A\backslash j) \cup i \notin \mathcal{F} \\ A, \ else \end{cases}$ to get its image $S_{ij}(\mathcal{F})$, in which sets containing j but not i are replaced by those were i replaces of j, unless that replacement was already in the family.
Just as for compression, we have $|S_{ij}(\mathcal{F})| = |\mathcal{F}|$. This is because s_{ij} is injective: if $s_{ij}(A) = s_{ij}(B)$, then if none or bother were shifted, we end up with $A = B$ or $(A\backslash j) \cup i = (B\backslash j) \cup i \Rightarrow A = B$, and if only one, say A was shifted, we'd get a contradiction from $B = (A\backslash j) \cup i \notin \mathcal{F}$.

A family $\mathcal{F}$ is t-**intersecting** if for any two $A, B \in \mathcal{F}$, we have $|A \cap B| \geqslant t$.

Shifting preserves the property of being t-intersecting. Indeed, we disjoin cases on $|s_{ij}(A) \cap s_{ij}(B)|$. If one or both were shifted, then in both cases $|s_{ij}(A) \cap s_{ij}(B)| = |A \cap B|$ or $|s_{ij}(A) \cap s_{ij}(B)| = |(A \cap B)\backslash j \cup i| = |A \cap B|$ (since in that case, both contained j and not i). If only one of them, say A, was shifted, we disjoin cases on B. If it wasn't shifted because $j \notin B$, then if $i \notin B$, $|s_{ij}(A) \cap s_{ij}(B)| = |((A\backslash j) \cup i) \cap B| = |A \cap B|$ as i and j don't appear in the intersection anyway, and if $i \in B$, $|((A\backslash j) \cup i) \cap B| > |A \cap B|$ as now i appears in the intersection. If $j \in B$ and B wasn't shifted because $i \in B$, then $|((A\backslash j) \cup i) \cap B| = |A \cap B|$ as in the intersection, we replaced j with i. In the final case were $j \in B, i \notin B$, but $B' = (B\backslash j) \cup i \in \mathcal{F}$ we do have the unfortunate $|((A\backslash j) \cup i) \cap B| < |A \cap B|$ which would seem compromise the intersecting property. Indeed, if $S_{ij}(\mathcal{F})$ had a pair of sets who's intersection has size less then t, then it must be for sets such as in this case, as in the other cases $|s_{ij}(A) \cap s_{ij}(B)| \geqslant |A \cap B|$ would contradict the assumption that $\mathcal{F}$ is t-intersecting. However, this case actually can't occur as it yields a contradiction: we have $|A \cap B'| = |A \cap (B\backslash j) \cup i| = |((A\backslash j) \cup i) \cap B| = |s_{ij}(A) \cap s_{ij}(B)| < t$, which contradicts the t-intersecting property of $\mathcal{F}$ at the pair $A, B' \in \mathcal{F}$.

We can define the p-**shadow** of a set family as the set of p-sized subsets of set from $\mathcal{F}$, denoted $\sigma_p(\mathcal{F}) = \{P : \ |P| = p, \exists A \in \mathcal{F}, P \subseteq A\}$. In the case that $\mathcal{F}$ was made of k-sized sets, $\sigma_{k-1}(\mathcal{F})$ is often simply called the **shadow** of the family, and is denoted by $\partial\mathcal{F}$.

We have $\sigma_p(S_{ij}(\mathcal{F})) \subseteq S_{ij}(\sigma_p(\mathcal{F}))$. To show this inclusion, we take a $P \in \sigma_p(S_{ij}(\mathcal{F}))$, so that P has size p and there is a $B \in S_{ij}(\mathcal{F})$ such that $P \subseteq B$, and we prove that it's in $S_{ij}(\sigma_p(\mathcal{F}))$. We show this by

case disjunctions on B.

First, we consider the case in which B was shifted: $B = (A\backslash j) \cup i$ for some $A \in \mathcal{F}$ that contained j but not i, so that B contains i and not j, and $B \notin \mathcal{F}$.
If $i \notin P$, we have $P \subseteq A$, so that $P \in \sigma_p(\mathcal{F})$. Simultaneously, since $P \subseteq B$, P can't contain j either, so that it wont be shifted, so that $s_{ij}(P) = P$. Thus $P \in S_{ij}(\sigma_p(\mathcal{F}))$.
If $i \in P$, we consider the p-set $(P\backslash i) \cup j$ which is a subset of A as $j \in A$ and all elements other then i and j of P where in $B = (A\backslash j) \cup i$. So $(P\backslash i) \cup j \in \sigma_p(\mathcal{F})$. Now, $(P\backslash i) \cup j$ will be shifted, if the last condition, which here is $P \notin \mathcal{F}$ holds. If the latter isn't the case, then $P \in \mathcal{F}$ so that $P \in \sigma_p(\mathcal{F})$, and since P didn't contain j, $s_{ij}(P) = P$, so that $P \in S_{ij}(\sigma_p(\mathcal{F}))$ in this case too. Now, if $P \notin \mathcal{F}$ so that $s_{ij}((P\backslash i) \cup j) = P$, then $P \in S_{ij}(\sigma_p(\mathcal{F}))$ since we just saw that $(P\backslash i) \cup j \in \sigma_p(\mathcal{F})$.

This concludes the case of a shifted B. If B wasn't shifted, then $B \in \mathcal{F}$ and $s_{ij}(B) = B$, and we disjoin three cases: either it's because, $j \notin B$, or it's because $j \in B$ but $i \in B$, or it's because $j \in B$, $i \notin B$ but $(B\backslash j) \cup i \in \mathcal{F}$.

In the first case, we have $P \subseteq B$, so that $P \in \sigma_p(\mathcal{F})$. Simultaneously, since $P \subseteq B$, P can't contain j either, so that it wont be shifted, so that $s_{ij}(P) = P$. Thus $P \in S_{ij}(\sigma_p(\mathcal{F}))$.

In the second case, if $i \in P$, we still have $P \subseteq B$, so that $P \in \sigma_p(\mathcal{F})$, and $s_{ij}(P) = P$, as P has i. Thus $P \in S_{ij}(\sigma_p(\mathcal{F}))$. If $i \notin P$, we disjoin cases on whether j is in P. So if $j \notin P$, we have $s_{ij}(P) = P$ and conclude $P \in S_{ij}(\sigma_p(\mathcal{F}))$. If however $j \in P$, we can use the fact that $(P\backslash j) \cup i \in \sigma_p(\mathcal{F})$ (since $(P\backslash j) \cup i \subseteq B$ is a p-set too in this case), thus when applying shift s_{ij} to the family $\sigma_p(\mathcal{F})$, $(P\backslash j) \cup i \in \sigma_p(\mathcal{F})$ prevents P from getting shifted, so that $s_{ij}(P) = P$, so that $P \in S_{ij}(\sigma_p(\mathcal{F}))$ again.

In the last case, we disjoin cases on whether j is in P. If it isn't $s_{ij}(P) = P$, so that $P \in S_{ij}(\sigma_p(\mathcal{F}))$ again. If it is, then we use $(P\backslash j) \cup i \subseteq (B\backslash j) \cup i \in \mathcal{F}$, for which $(P\backslash j) \cup i \in \sigma_p(\mathcal{F})$, so that when applying shift s_{ij} to the family $\sigma_p(\mathcal{F})$, $(P\backslash j) \cup i \in \sigma_p(\mathcal{F})$ prevents P from getting shifted, so that $s_{ij}(P) = P$, so that $P \in S_{ij}(\sigma_p(\mathcal{F}))$ again.

4.2 Kruskal-Katona theorem

An interesting case of set families are when there is an order on their elements, for example set families of finite sets of $\mathbb{N}$. In that case, if we perform shifts s_{ij} where $i < j$, we can expect the values of the sums over all elements to decrease. Indeed, consider $\sum_{A\in\mathcal{F}}\sum_{a\in A} a$. If $S_{ij}(\mathcal{F}) \neq \mathcal{F}$, then one of the sets A of $\mathcal{F}$ must have been shifted, and since $\sum_{a\in A} a - \sum_{a\in(A\setminus j)\cup i} a = j - i > 0$, we have $\sum_{A\in\mathcal{F}}\sum_{a\in A} a > \sum_{B\in S_{ij}(\mathcal{F})}\sum_{b\in B} b$. This quantity will thus decrease for repeated applications of shifts s_{ij} where $i < j$, and since it's in $\mathbb{N}$, this can't happen forever, so that at by denoting $\mathcal{F}_0 = \mathcal{F}$ and for the kth shift $s_{i_kj_k}$ where $i_k < j_k$, $\mathcal{F}_k = s_{i_kj_k}(\mathcal{F}_{k-1})$, there must come an n such that $\mathcal{F}_n = s_{ij}(\mathcal{F}_n)$ for all $i < j$. Such a family is called **shifted**.

Integer Kruskal-Katona theorem:

We consider a finite set family $\mathcal{F}$ of finite sets of $\mathbb{N}$, so that the sets are k-sets, then for $p < k$ and $n \geqslant k$, we have $|\mathcal{F}| \geqslant \binom{n}{k} \Rightarrow |\sigma_p(\mathcal{F})| \geqslant \binom{n}{p}$.

We first note that if we've prove the theorem for family $S_{ij}(\mathcal{F})$, then it also holds for $\mathcal{F}$. This is bacause $|S_{ij}(\mathcal{F})| = |\mathcal{F}|$ and $|\sigma_p(\mathcal{F})| = |S_{ij}(\sigma_p(\mathcal{F}))| \geqslant |\sigma_p(S_{ij}(\mathcal{F}))|$ by our previous results, so that the bounds on $|\mathcal{F}|$ imply those on $S_{ij}(\mathcal{F})|$, which will imply the lower-bounds on $|\sigma_p(S_{ij}(\mathcal{F}))|$, hence on $|\sigma_p(\mathcal{F})|$. We can therefore assume wlog. that $\mathcal{F}$ is shifted, as we can apply shifts to it until it is, prove the theorem for the shifted family, then gain the bounds on $\mathcal{F}$ with the observation we just made. Note also that if $\mathcal{F}$ was made of k-sets, then so is $S_{ij}(\mathcal{F})$ (elements are only swapped, so the sizes stay the same), so that we can carry that assumption over to the shifted family as well.

We will show this by induction. The base case $n = k$ corresponds to $|\mathcal{F}| \geqslant 1 \Rightarrow |\sigma_p(\mathcal{F})| \geqslant \binom{k}{p}$ which is true since $\mathcal{F}$ has at least one k-set, who's p-shadow has $\binom{k}{p}$ elements.

For the step, we consider the case $p = k - 1$ first.
We use notations $\mathcal{F}_x = \{A\setminus x : A \in \mathcal{F}, x \in A\}$ and $\mathcal{F}_{\neg x} = \{A : A \in \mathcal{F}, x \notin A\}$.
As a first lemma, we have $\sigma_{k-1}(\mathcal{F}_{\neg 1}) \subseteq \mathcal{F}_1$. Indeed, for any $B \in \sigma_{k-1}(\mathcal{F}_{\neg 1})$, there is a $1 \neq j \notin B$ such that $B \cup j \in \mathcal{F}_{\neg 1} \subseteq \mathcal{F}$. Since $1 < j$, by shiftedness of the family, $S_{1j}(\mathcal{F}) = \mathcal{F}$, so that the image of $B \cup j$ is in $\mathcal{F}$ too, hence $B \cup 1 \in \mathcal{F}$. The latter means that $B \in \mathcal{F}_1$.

From this, we can deduce that $|\mathcal{F}_1| \geqslant \binom{n-1}{k-1}$. We start by noting that $|\mathcal{F}_x| + |\mathcal{F}_{\neg x}| = |\mathcal{F}|$, as the conditions forming their sets partition $\mathcal{F}$. So if, for a contradiction, we had $|\mathcal{F}_1| < \binom{n-1}{k-1}$, then $|\mathcal{F}_{\neg 1}| = |\mathcal{F}| - |\mathcal{F}_1| > \binom{n}{k} - \binom{n-1}{k-1} = \binom{n-1}{k}$. So $\mathcal{F}_{\neg 1}$ is a candidate for applying the induction hypothesis (note that it's also made of k-sets), so that $|\sigma_{k-1}(\mathcal{F}_{\neg 1})| \geqslant \binom{n-1}{k-1}$ (apply the case $p = k - 1$), hence $|\sigma_{k-1}(\mathcal{F}_{\neg 1})| > |\mathcal{F}_1|$, which contradicts the $\sigma_{k-1}(\mathcal{F}_{\neg 1}) \subseteq \mathcal{F}_1$ we previously established.

We can partition $\sigma_{k-1}(\mathcal{F})$ into the subsets in which we erase 1, aka. $\mathcal{F}_1$, and those in which we

didn't erase 1. The latter can be interpreted as the subsets of $\mathcal{F}_1$ obtained by erasing an element, so as $\{B \cup 1 : B \in \sigma_{k-2}(\mathcal{F}_1)\}$ (recall that the $\mathcal{F}_1$ are $(k-1)$-sets). So we have $|\sigma_{k-1}(\mathcal{F})| = |\mathcal{F}_1| + |\sigma_{k-2}(\mathcal{F}_1)|$ by counting the sets according to this disjunction of type.

Now, since $|\mathcal{F}_1| \geqslant \binom{n-1}{k-1}$, it's a candidate for applying the induction hypothesis (recall that the $\mathcal{F}_1$ are $(k-1)$-sets) and therefore$|\sigma_{k-2}(\mathcal{F}_1)| \geqslant \binom{n-1}{k-2}$ (use $p = (k-1)-1$).

Thus $|\sigma_{k-1}(\mathcal{F})| = |\mathcal{F}_1| + |\sigma_{k-2}(\mathcal{F}_1)| \geqslant \binom{n-1}{k-1} + \binom{n-1}{k-2} = \binom{n}{k-1}$, which is the statement of the theorem for $p = k-1$.

To get the cases of lower p, will will do a reverse finite induction on p. The base case is then $p = k-1$, and the step involves showing the case $p = k-(i+1)$ from $p = k-i$. The finite-induction assumption is that $|\sigma_{k-i}(\mathcal{F})| \geqslant \binom{n}{k-i}$.

To get this done, we split $\sigma_{k-(i+1)}(\mathcal{F}) = \sigma_{(k-i)-1}(\sigma_{k-i}(\mathcal{F}))$, since subsets of size $(k-i)-1$ are subsets of subsets of size $(k-i)$. By writing $k' = k-i$ and $p' = k'-1$, we have $\sigma_{k-(i+1)}(\mathcal{F}) = \sigma_{p'}(\sigma_{k'}(\mathcal{F}))$ and assumption $|\sigma_{k'}(\mathcal{F})| \geqslant \binom{n}{k'}$.

Since in the previous part of the proof in which $p = k-1$, k was arbitrary, we can use the exact same arguments to show that $|\sigma_{k'}(\mathcal{F})| \geqslant \binom{n}{k'} \Rightarrow |\sigma_{k'-1}(\sigma_{k'}(\mathcal{F}))| \geqslant \binom{n}{k'-1}$, so that $|\sigma_{k-(i+1)}(\mathcal{F})| \geqslant \binom{n}{k-(i+1)}$ as desired.

4.3 Perles-Shelah-Sauer-Vapnik-Chervonensky theorem

Perles-Shelah-Sauer-Vapnik-Chervonensky theorem:

Consider a family $\mathcal{F}$ and a $k < |X|$ such that $|\mathcal{F}| > \sum_{i=0}^{k} \binom{|X|}{i}$.
Then there is a $(k+1)$-set $T \subseteq X$ such that $|\mathcal{F}_{|T}| = 2^{|T|} = 2^{k+1}$, aka. we can get any subset of T ans an intersection of T and a set of $\mathcal{F}$, or equivalently $\mathcal{F}_{|T}$ is hereditary.

We will show this by contradiction, so that we assume that for all $(k+1)$-set $T \subseteq X$, we have $|\mathcal{F}_{|T}| \neq 2^{|T|}$. Note that this means $|\mathcal{F}_{|T}| < 2^{|T|}$, as $\mathcal{F}_{|T}$ is made of subsets of T.
We can smell the contradiction in using the notion of $\mathcal{F}^{cpr}$. On the one hand, we know that since $\mathcal{F}^{cpr} = c_{x_k}(...c_{x_1}(\mathcal{F}))$, and compression maintains family size, $|\mathcal{F}^{cpr}| = |\mathcal{F}|$, and since compression decreases trace size, $|\mathcal{F}^{cpr}_{|T}| \leqslant |\mathcal{F}_{|T}| < 2^{|T|} = 2^{k+1}$. What changed with $\mathcal{F}^{cpr}$ is that it's hereditary. This is why $\mathcal{F}^{cpr}$ can't contain sets H of size $k+1$ or greater: it would have to contain all that sets subsets (so wlog. H has size $k+1$), which would also be in $\mathcal{F}^{cpr}_{|H}$ so that $|\mathcal{F}^{cpr}_{|H}| \geqslant 2^{k+1}$. This allows us to bound $|\mathcal{F}^{cpr}| \leqslant \left|\binom{X}{[k]}\right| = \sum_{i=0}^{k} \binom{|X|}{i}$, which with $|\mathcal{F}^{cpr}| = |\mathcal{F}|$ contradicts our assumption.

4.4 Erdös-Ko-Rado theorem

Erdös-Ko-Rado theorem:

We consider an (1-) intersecting family $\mathcal{F}$ on a ground set of size n, made of k-sets, for some k satisfying $2k \leqslant n$. We then have bound $|\mathcal{F}| \leqslant \binom{n-1}{k-1}$.

4.5 Katona's union theorem and Kleitman's diameter theorem

We call a family $\mathcal{F}$ on a ground set of size n and **s-union** if for all $A, B \in \mathcal{F}$, we have $|A \cup B| \leqslant n - s$. For the case $s = 1$, we just call it a union.

Katona union theorem:

We consider an s-union $\mathcal{F}$ on a ground set of size n.
If $n - s = 2p$ is even, then $|\mathcal{F}| \leqslant \sum_{k=0}^{p} \binom{n}{k}$, and if $n - s = 2p + 1$ is odd, then $|\mathcal{F}| \leqslant 2 \sum_{k=0}^{p} \binom{n-1}{k}$.

Kleitman diameter theorem:

We consider a family $\mathcal{F}$ on a ground set of size n with diameter at most d.
If $d = 2p$ is even, then $|\mathcal{F}| \leqslant \sum_{k=0}^{p} \binom{n}{k}$, and if $d = 2p + 1$ is odd, then $|\mathcal{F}| \leqslant 2 \sum_{k=0}^{p} \binom{n-1}{k}$.

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

6 The regularity method

6.1 Introductory problems

COMPLETE: find introductory problem, move regularity definitions into that problem's solution

6.2 Szemerédi's graph regularity lemma

Edge density:

Denoting with $e(X,Y) = |\{\{u,v\} \in E : u \in X, v \in Y\}|$ the edges between vertex sets X, Y, we define the **edge density** $d(X,Y)$ to be $\dfrac{e(X,Y)}{|X||Y|}$.

This is also the probability of getting a $e(X,Y)$-edge graph in a bipartite graph with partition sets of size $|X|$ and $|Y|$, where edges are chosen uniformly. In our context, X and Y may share vertices, so that $|X||Y|$ may be larger then the number of possible edges.

(ε,δ)-regular pair:

For $\varepsilon, \delta \in]0,1]$, an **$(\varepsilon,\delta)$-regular pair** of vertex sets X and Y is a pair for which for all subsets $A \subseteq X$ and $B \subseteq Y$ not to relatively small, in the sense that $|A| \geqslant \varepsilon|X|$ and $|B| \geqslant \varepsilon|Y|$, we have close densities to the supset one: $|d(A,B) - d(X,Y)| \leqslant \delta$.

In a pair that is not (ε,δ)-regular, subsets $A \subseteq X$ and $B \subseteq Y$ with $|A| \geqslant \varepsilon|X|$ and $|B| \geqslant \varepsilon|Y|$, for which $|d(A,B) - d(X,Y)| > \delta$ are called **witnesses** of the irregularity.

We'll actually only need **ε-regular pairs**, which are $(\varepsilon,\varepsilon)$-regular pairs.
The reason we consider $\varepsilon, \delta \leqslant 1$ is that for ε, saying otherwise, $|A| \geqslant \varepsilon|X|$ would contradict $A \subseteq X$, and for δ, densities are always in $[0,1]$, so that the bound with $\delta > 1$ would always be true.
To get close densities over subsets, we can't expect to consider to small subsets in the definition. Consider a large bipartite graph with a single edge: the density of the bipartition-sets is small, but the density of the singleton subsets corresponding to the endpoints of the edge is 1. If we had ignore the constraints $|A| \geqslant \varepsilon|X|$ and $|B| \geqslant \varepsilon|Y|$ in the definition, the bipartition-sets wouldn't have qualifies as a ε-regular pair fro small epsilon.

ε-regular partition:

An **ε-regular partition** is a partition of vertices into k sets V_i, for some k, such that by denoting with R the pairs $(i,j) \in [k]^2$ for which V_i and V_j are ε-regular, and by $\overline{R}$ its complement (the irregular pairs), we have $\displaystyle\sum_{(i,j)\in\overline{R}} |V_i||V_j| \leqslant \varepsilon|V|^2$.

By taking to elements $x, y \in V$ uniformly and independently at random, the partition is ε-regular if the probability of x and y being in a pair of partition sets that aren't ε-regular, is less then ε. This translates to $\displaystyle\sum_{(i,j)\in\overline{R}} \frac{|V_i||V_j|}{|V||V|} \leqslant \varepsilon$, where $\dfrac{|V_i||V_j|}{|V||V|}$ is the probability of $x \in V_i$ and $y \in V_j$.

The road to finding an ε-regular partition is quite a curvy one.
It starts with the study of a particular random variable related to partitions. For partitions $(U_i)_{i\in[k]}$ of a subset of vertices U and $(W_i)_{i\in[l]}$ of a subset of vertices W, we pick two points $x \in U$ and $y \in W$ independently and uniformly at random, and we consider Z to be the edge-density of the partition sets they ended up in. This relates to ε-regular partitions in the sense that if $U = W = V$ and the partitions are the same and also ε-regular, then $P(Z > \varepsilon) \leqslant \varepsilon$. *What's Z on average ?*

We have $E(Z) = \sum_{i\leqslant k,j\leqslant l} \frac{|U_i||W_j|}{|U||W|} d(U_i, W_j) = \sum_{i\leqslant k,j\leqslant l} \frac{e(U_i,W_j)}{|U||W|} = \frac{1}{|U||W|} \sum_{i\leqslant k,j\leqslant l} e(U_i, W_j) = d(U, W)$, since $\sum_{i\leqslant k,j\leqslant l} e(U_i, W_j) = e(U, W)$, as we induce a partition on edges.
So the average density among partition sets is the density of the ground sets.

The next step is based on notions of energy from probability theory. We'll consider a quantity that will increase under refinement, the same way energy decreases in until equilibrium in physics. In an effort to compute the variance of Z, we consider $E\left(Z^2\right) = \sum_{i\leqslant k,j\leqslant l} \frac{|U_i||W_j|}{|U||W|} d(U_i, W_j)^2$, which we can combine with $E\left(Z^2\right) \geqslant E(Z)^2$ (convexity) to get $\sum_{i\leqslant k,j\leqslant l} \frac{|U_i||W_j|}{|U||W|} d(U_i, W_j)^2 \geqslant d(U, W)^2$.

Now, if U and W where already partition sets of a partition of V for example, say $U = V_k$ and $W = V_q$, we can make a certain pattern appear.

If p is the partition of the V_s, we can get a refinement p' of p by replacing V_k with the U_i, that we'll call $V_{k,i}$ and V_q by the W_j, which we'll call $V_{q,j}$, to get a new partition of V. Then, we have $\sum_{i\leqslant k,j\leqslant l} \frac{|V_{k,i}||V_{q,j}|}{|V_k||V_q|} d(V_{k,i}, V_{q,j})^2 \geqslant d(V_k, V_q)^2$.
To make similarities appear, we write it as $\sum_{i\leqslant k,j\leqslant l} |V_{k,i}||V_{q,j}| d(V_{k,i}, V_{q,j})^2 \geqslant |V_k||V_q| d(V_k, V_q)^2$.
We then define the "energy" of a pair A, B by $Q_{pr}(A, B) = \frac{|A||B|}{|V|^2} d(A, B)^2$. The normalization factor $\frac{1}{|V|^2}$ will be used to upper-bound energy, but for now just note that due to its positivity, the previous inequality becomes $\sum_{i,j} Q_{pr}(V_{k,i}, V_{q,j}) \geqslant Q_{pr}(V_k, V_q)$.
We can make this an identity about partitions by introducing the "energy" of a partition $Q((V_i)_{i\in[L]}) = \sum_{k,q\in[L]} Q_{pr}(V_k, V_q) = \sum_{r,s\in[L]} \frac{|V_r||V_s|}{|V|^2} d(V_r, V_s)^2$. If we refine a partition p into p', then we can add the inqualities $\sum_{i,j} Q_{pr}(V_{k,i}, V_{q,j}) \geqslant Q_{pr}(V_k, V_q)$ for all pairs V_k, V_q of p, as well as $Q_{pr}(A, B) \geqslant 0$ for the pairs of p' that havent shown up in the previous kind of inequalities (of which there aren't any, actually), to get $Q(p') \geqslant Q(p)$. So energy increases under refinement.

The reason we normalized the sum is so that energy is bounded: for any partition p, we have $Q(p) \leqslant 1$, because $\sum_{r,s\in[L]} \frac{|V_r||V_s|}{|V|^2} = 1$ (split the sum and recall we deal with a partition) and $d(V_r, V_s) \leqslant 1$ (density).

Since the normalization factor is positive, we still keep the increase under refinement property. We want our notion of energy to be bounded, since we'll create a method that refines partitions so as to increase energy by a fixed minimal amount. This will allow us to conclude that the process must terminate.
For now, let's just collect:

Energy:

To a partition $(V_i)_{i\in[L]}$ of V, we associate an **energy** $Q((V_i)_{i\in[L]}) = \sum_{r,s\in[L]} \frac{|V_r||V_s|}{|V|^2} d(V_r,V_s)^2$,

which is upper-bounded by 1.

For a refinement p' of a partition p, in the sense that we can group the sets of p' so that the groups form partitions of the sets of p, we have $Q(p') \geqslant Q(p)$.

To see that $Q(p') \geqslant Q(p)$ for any refinement p' of p, note that we can refine sets of p one at a time, increasing energy along the way by the argument we previously gave, where the refinement refine a single set of p, until we reach p'.

As we mentioned, we'll exploit the idea of energy having to increase under refinements, yet being upper-bounded. The key point that relates this study to ε-regular partitions, is that we'll find a way to exploit witnesses of irregularities so as to get refinements that increase energy by a minimum amount. Since energy is bounded, so increases can happen only finitely many times, and hence so must there cause, the irregularites. This will guarantee our algorithm to end, and with a partition with few irregular pairs as output.

So, how do witnesses of irregularities relate to energy ?
With our definition of energy, for the refinement p' of p as described previously (refines one set), we could write the pair-energy difference as $\sum_{i\leqslant k,j\leqslant l} Q_{pr}(V_{k,i},V_{q,j}) - Q_{pr}(V_k,V_q) =$

$\frac{|V_k||V_q|}{|V|^2}\left(\left(\sum_{i\leqslant k,j\leqslant l}\frac{|V_{k,i}||V_{q,j}|}{|V_k||V_q|}d(V_{k,i},V_{q,j})^2\right) - d(V_k,V_q)^2\right) = \frac{|V_k||V_q|}{|V|^2}\left(E\left(Z^2\right) - E(Z)^2\right)$,

so that $\frac{|V|^2}{|V_k||V_q|}(\sum_{i\leqslant k,j\leqslant l} Q_{pr}(V_{k,i},V_{q,j}) - Q_{pr}(V_k,V_q)) = Var(Z)$.

We know that this increase will be positive, but we'd like to lower bound it when this is possible. By expressing $Var(Z) = E\left((Z-E(Z))^2\right) = \sum_{i\leqslant k,j\leqslant l}\frac{|V_{k,i}||V_{q,j}|}{|V_k||V_q|}(d(V_{k,i},V_{q,j}) - d(V_k,V_q))^2$, we see a density difference appear, which relates to ε-regular pairs.

Indeed, if one of the partitions set pairs $V_{k,i},V_{q,j}$ is a witness to an irregular pair V_k,V_q, then we can bound $\frac{|V_{k,i}||V_{q,j}|}{|V_k||V_q|}(d(V_{k,i},V_{q,j}) - d(V_k,V_q))^2 > \varepsilon.\varepsilon.\varepsilon^2 = \varepsilon^4$. So if such a pair exists, we have also $Var(Z) > \varepsilon^4$ and thus energy boost $\sum_{i\leqslant k,j\leqslant l} Q_{pr}(V_{k,i},V_{q,j}) > Q_{pr}(V_k,V_q) + \frac{|V_k||V_q|}{|V|^2}\varepsilon^4$. This result has a nice name:

The red bull lemma:

If p' is a refinement of p such that p' contains a witness pair for an irregular pair V_k, V_q of sets of p, then $\sum_{i\leqslant k, j\leqslant l} Q_{pr}(V_{k,i}, V_{q,j}) > Q_{pr}(V_k, V_q) + \frac{|V_k||V_q|}{|V|^2}\varepsilon^4$.

We can already see the end of the proof of the regularity lemma: if $Q(p)$ is close to 1 and $\frac{|V_k||V_q|}{|V|^2}\varepsilon^4$ would make $Q(p')$ greater then 1, the fact that energy is at most 1 leads us to conclude that the pairs must be ε-regular.

The issue is that the increment term depends on the size of the partition sets, on which we have no bounds, but which we expect to decrease in size over successive refinements. However, the term $\frac{|V_k||V_q|}{|V|^2}$ is related to ε-regular partitions: if p isn't ε-regular, we can lower-bound the sum of the $\frac{|V_k||V_q|}{|V|^2}$ corresponding to irregular pairs by ε.

Now, assembling the inequalities, we get:
$Q(p) \leqslant \sum_{(r,s)\in R}\sum_{c,c'} Q_{pr}(C_{r,c}, C_{s,c's}) + \sum_{(r,s)\in\overline{R}}\left(\sum_{c,c'} Q_{pr}(C_{r,c}, C_{s,c's}) - \frac{|V_r||V_s|}{|V|^2}\varepsilon^4\right)$, so that, making $Q(p') = \sum_{r,s}\sum_{c,c'} Q_{pr}(C_{r,c}, C_{s,c's})$ appear and using bound we mentioned at the very begining of our argument, namely $\sum_{(r,s)\in\overline{R}} \frac{|V_r||V_s|}{|V|^2} > \varepsilon$ for partitions that aren't ε-regular, we get the following key result:

Energy boosting refinements for irregular partitions:

If p is a partition that is not ε-regular, then by refining it the way we described produces a partition p' such that $Q(p') \geqslant Q(p) + \varepsilon^5$.

We can now use the following procedure: start with the one-set partition of V into $\{V\}$ and repeatedly apply our refinement to boost energy, if for the current partition, it isn't ε-regular. The energy increase being at least ε^5, and energy being upper-bounded by 1, this can happen at most ε^{-5} times. In particular:

Szemerédi's graph regularity lemma:

ε-regular partitions exist for any $\varepsilon \in]0, 1]$

What can we say about their size of these partitions?
As we mentioned, at each step, at most $2^{|\overline{R}|}$ new partition sets appear per pair in $\overline{R}$, making it $|\overline{R}|2^{|\overline{R}|}$ new sets in total. This is largest when $|\overline{R}|$ is largest, where $|\overline{R}|$ is all of the pairs. Since we start with a single set, the next refinement will have $1.2^1 = 2$ sets. Then, the next one will have at most $2.2^2 = 8$ sets, and then at most 8.2^8 in the next round. A more transparent bound can be obtained by using $n2^n \leqslant 2^{(2^n)}$, so that repeated insertion yields $tower(2, 2i)$ for the ith iteration.

In the slowest case however, refinements will add a single partition set at each iteration. Since a partition can have size at most $|V|$, the algorithm must terminate after $|V|$ iterations. This doesn't necessarily make it a polynomial time algorithm, since we haven't discussed the efficiency of the step of finding a witness pair.

7 The Combinatorial Nullstellensatz and regular subgraphs

The big idea of the "polynomial method" of combinatorics is that of associating a polynomial to a certain combinatorial object, such that the roots or non-root points of the polynomial correspond to particular structures in the combinatorial object. Finding or counting these structures is then equivalent to finding or counting roots or non-root points.

An example is guaranteeing the existence of regular subgraphs in some graphs. For a graph (V, E), we can use indicator variables $x_e \in \{0, 1\}$ to indicate if the edge will be induced in the subgraph, and have the subgraph be induced by these edges. Using incidence coefficients $a_{v,e}$ for vertex v and edge e which is 1 if $v \in e$ and 0 otherwise, the sum $\sum_{e \in E} a_{v,e}x_e$ will represent the degree of v in the subgraph. A way to get regularity is by setting the goals $\sum_{e \in E} a_{v,e}x_e \equiv 0\ (p)$ for a prime p, and adding the assumption that the maximum degree of (V, E) is $2p - 1 < 2p$. Then, we know that the prime p divides $\sum_{e \in E} a_{v,e}x_e$, which isn't one of the multiples $2p, 3p, ...$ of p since $\sum_{e \in E} a_{v,e}x_e < 2p$ by the degree constraint, so that we must have $\sum_{e \in E} a_{v,e}x_e = 0$ (the vertex won't be part of the subgraph) or $\sum_{e \in E} a_{v,e}x_e = p$.

To relate the goals to a polynomial, we can reformulate $\sum_{e \in E} a_{v,e}x_e \equiv 0\ (p)$ to $1 - \left(\sum_{e \in E} a_{v,e}x_e\right)^{p-1} \neq 0\ (p)$ with Fermat's little theorem, as the only integers for which $(n)^{p-1} \neq 1\ (p)$ are $n \equiv 0\ (p)$. We can ask for this at any vertex, to get he existence of a spanning p-regular subgraph.

This can be expressed as $\prod_{v \in V} \left(1 - \left(\sum_{e \in E} a_{v,e}x_e\right)^{p-1}\right) \neq 0\ (p)$.

In fact, note that $\prod_{v \in V} \left(1 - \left(\sum_{e \in E} a_{v,e}x_e\right)^{p-1}\right)$ takes value 1 (mod p) at $x = 0$ and when x indicates a p-regular subgraph and value 0 (mod p) if it indicates a non-p-regular subgraph. We can actually also indicate $x = 0$ with $\prod_{e \in E}(1 - x_e)$.

We then consider $f(x) = \prod_{v \in V} \left(1 - \left(\sum_{e \in E} a_{v,e}x_e\right)^{p-1}\right) - \prod_{e \in E}(1 - x_e)$ as a polynomial over $\mathbb{F}_p$. If $x = 0$ (he empty subgraph), it takes value 0, and if $x \neq 0$, the only other roots correspond to non-empty non-p-regular subgraph.

So the existence of a point $x \in \{0, 1\}^E$ for which $f(x) \neq 0$ will imply the existence of a p-regular subgraph. Fortunately, we have:

The Combinatorial Nullstellensatz:

We consider an n-variate non-zero polynomial f over a field $\mathbb{F}$. Assume that for all multi-indices $(a_1, ..., a_n)$ of monomials of f, the sum $a_1 + ... + a_n$ is maximized by $(m_1, ..., m_n)$. Then for finite $S_i \subseteq \mathbb{F}$ with $|S_i| > m_i$, there is a point $x \in S_1 \times ... \times S_n$ such that $f(x) \neq 0$.

In our example, $S_i = \{0, 1\}$. To apply the theorem, we need to check which monomial maximises the

sum of its multi-index. We'll actually need to add an assumption on our graph to get this. Indeed, in $\prod_{e\in E}(1-x_e)$, the sum maximiser is $(1,1,...,1)$ with sum $|E|$. In $\prod_{v\in V}\left(1-\left(\sum_{e\in E}a_{v,e}x_e\right)^{p-1}\right)$, the sums are at most $(p-1)|V|$. We want to have $(p-1)|V|<|E|$, so that $(1,1,...,1)$ is the unique maximiser, so that $2=|S_i|>m_i=1$ holds and we may apply the theorem.

We can either set this condition directly, or embellish it: with $2|E|=\sum_{v\in V}\deg(v)$, we can see that $(p-1)|V|<|E|$ is equivalent to the graph having average degree greater then $2p-2$.
The condition is compatible with that on the maximum degree of the graph, as $\sum_{v\in V}\deg(v)\leqslant(2p-1)|V|$.
As a conclusion, we've shown that for a graph with average degree greater then $2p-2$ and maximum degree at most $2p-1$, we may find a non-empty p-regular subgraph in it.

Now, let's prove the combinatorial Nullstellensatz, shall we ?
First, we'll need some algebra. We want to get a better description of the polynomials f for which all the points of $S_1\times...\times S_n$ are roots. *Can we find a certain representation of them ? Are they generated by a particular family of polynomials ?*

We're used to factoring univariate polynomials in the form $\prod(x-r)$. We can then note that all polynomials of form $\sum_{i\in[n]}h_i\prod_{s\in S_i}(x_i-s)$, for any polynomials h_i, will vanish on $S_1\times...\times S_n$. In fact, we'll show that any polynomial that vanishes on $S_1\times...\times S_n$ can be written in this form.

We will derive the form recursively, so that we prove the result by induction.
The idea is to reduce the degree of f by subtracting terms of form $h_i\prod_{s\in S_i}(x_i-s)$, so that the property of vanishing on $S_1\times...\times S_n$ is maintained, but the multi-degrees are decreased. We'll the expect to get a polynomial with more roots then its degree allows, which must therefore be the zero polynomial, so that f equals the sum of terms we subtracted.

We go ahead by writing $\prod_{s\in S_i}(x_i-s)=x_i^{|S_i|}-\sum_{j=0}^{|S_i|-1}c_{ij}x_i^j$. If at the current iteration k, the polynomial $f-\sum_{i\in[n]}h_i^{(k)}\prod_{s\in S_i}(x_i-s)$ has a monomial $(m_1,...,m_n)$ in which x_i has power $\geqslant|S_i|$, then we subtract $h_i'\prod_{s\in S_i}(x_i-s)$ to get the next iteration, where h_i' is a monomial $(m_1,...,m_i-|S_i|,...m_n)$ with the same coefficient as the previous one. The result is that the monomial $(m_1,...,m_n)$ will get replaced by $\left(x_1^{m_1}...x_i^{m_i-|S_i|}...x_n^{m_n}\right)\sum_{j=0}^{|S_i|-1}c_{ij}x_i^j$. In the latter polynomials monomials, the powers of x_i will be $<m_i$.
So if we selected the monomial $(m_1,...,m_n)$ in which x_i has power $\geqslant|S_i|$ such that m_i was maximum for this property, we note that in each iteration, there is an i for which the highest power in a monomial will decrease in the iteration. The algorithm will therefore terminate, and with a $f-\sum_{i\in[n]}h_i^{(k)}\prod_{s\in S_i}(x_i-s)$ in which all monomials have $m_i<|S_i|$, for all i.

We will soon show as a quick lemma that such a polynomial must be the zero polynomial, so that from

the last iteration we get expression $f = \sum_{i\in[n]} h_i^{(k)} \prod_{s\in S_i}(x_i - s)$.

Indeed, consider a polynomial g that vanishes on $S_1 \times ... \times S_n$ and for who's monomials $m_i < |S_i|$, for all i. If we factor g as $g = \sum_{j=0}^{m_n} g_j x_n^j$, then by fixing a point in $S_1 \times ... \times S_{n-1}$, we get a univariate polynomial $\sum_{j=0}^{m_n} g_j x_n^j$ in x_n, which has $|S_n| > m_n$ roots, and must therefore be the zero polynomial. So for all j, the polynomials g_j vanish on all points of $S_1 \times ... \times S_{n-1}$. The property $m_i < |S_i|$ for all i is inherited by the g_j, as otherwise we'd contradict it for g. So we are in the same context of assumptions, but with less variables: this calls for a proof by induction. The base case is the univariate one, in which this is just the fact that a polynomial with more roots then the degree is the zero polynomial. So our lemma holds.

We are now ready to prove the combinatorial Nullstellenstatz. We'll prove it by contradiction, by showing that it can't vanish on $S_1 \times ... \times S_{n-1}$. Note the difference to the previous lemma: here, $(m_1, ..., m_n)$ maximises a sum, and there may be monomials $(a_1, ...a_n)$ in f with $a_i \geqslant |S_i|$ for some i, but which don't maximise their sum.
If it vanished, we could get $f = \sum_{i\in[n]} h_i \prod_{s\in S_i}(x_i - s)$ as we just described.

We have $f = \sum_{i\in[n]} h_i \prod_{s\in S_i}(x_i - s) = \sum_{i\in[n]} h_i \left(x_i^{|S_i|} - \sum_{j=0}^{|S_i|-1} c_{ij} x_i^j \right)$.

We can split the expression further into $f = \sum_{i\in[n]} \left(\sum_{j=m_i+1}^{|S_i|} -c_{ij} x_i^j h_i + \sum_{j=0}^{m_i} -c_{ij} x_i^j h_i \right)$ (by setting $c_{i|S_i|} = -1$).
In the expression on the right, there is the monomial $(m_1, ..., m_n)$. It must therefore also be on the left, meaning that there is at least one i such that the monomial is in $\left(\sum_{j=m_i+1}^{|S_i|} -c_{ij} x_i^j h_i + \sum_{j=0}^{m_i} -c_{ij} x_i^j h_i \right)$, which in turn means that there is at least a j such that $x_i^j h_i$ is that monomial. It can't be part of the left sum, where $j > m_i$, as it all monomials of $x_i^j h_i$ the power of m_i is $> m_i$. It could be on the left, but a closer look at the h_i will prevent this.

Recall how the h_i were formed. At iteration k, we considered $f - \sum_{i\in[n]} h_i^{(k)} \left(x_i^{|S_i|} - \sum_{j=0}^{|S_i|-1} c_{ij} x_i^j \right)$. We looked for a monomial $(a_1, ...a_n)$ of it where $a_i \geqslant |S_i|$, and subtracted $c \left(x_1^{a_1} ... x_i^{a_i - |S_i|} ... x_n^{a_n} \right) \left(x_i^{|S_i|} - \sum_{j=0}^{|S_i|-1} c_{ij} x_i^j \right)$ for the appropriate coefficient c, so that the $(a_1, ..., a_i - |S_i|, ...a_n)$ was added to $h_i^{(k+1)}$. Passing from $f - \sum_{i\in[n]} h_i^{(k)} \left(x_i^{|S_i|} - \sum_{j=0}^{|S_i|-1} c_{ij} x_i^j \right)$ to $f - \sum_{i\in[n]} h_i^{(k+1)} \left(x_i^{|S_i|} - \sum_{j=0}^{|S_i|-1} c_{ij} x_i^j \right)$ then corresponded to deleting $(a_1, ...a_n)$ and replacing it with monomials $(a_1, ..., a_i - q, ...a_n)$ where $1 \leqslant q \leqslant |S_i|$.
Let's now look at the multi-index sums. At each iteration, in the h_i, a monomial who's sum is less by $|S_i|$ then that of a monomial from $f - \sum_{i\in[n]} h_i^{(k)} \left(x_i^{|S_i|} - \sum_{j=0}^{|S_i|-1} c_{ij} x_i^j \right)$ is added to it. Also, the multi-index

sums of $f - \sum_{i\in[n]} h_i^{(k)} \left(x_i^{|S_i|} - \sum_{j=0}^{|S_i|-1} c_{ij} x_i^j \right)$ will decrease of the iterations, as we replace its $(a_1, ...a_n)$ with $(a_1, ..., a_i - q, ...a_n)$, where $1 \leqslant q \leqslant |S_i|$. We can therefore conclude that the highest an multi-index sum in h_i can be is $m_1 + ... + m_n - |S_i|$, as it's highest for the monomial of the first iteration.

We can now show that $(m_1, ..., m_n)$ can't be part of $\sum_{j=0}^{m_i} -c_{ij} x_i^j h_i$. Indeed, the multi-index sums of the monomials in the latter are $a_1 + ... + a_i + j + ...a_n + \leqslant m_1 + ... + m_n - |S_i| + m_i < m_1 + ... + m_n$ since $|S_i| > m_i$.
We therefore get the contradiction we desired.

8 Van der Waerden's theorem

For colorings with more then 1 color, can one find infinite arithmetic progrressions ?
The answer is no and the idea behind it is that we can construct colorings in which longer and longer segments have the same color, so that no step of an arithmetic progression can cross them.

Such a coloring could be coloring the $j + \sum_{i=0}^{n} i$ with $0 \leqslant j \leqslant n$ for even n in red, and of odd n in blue.

This will assign numbers a unique color, as each number can be represented uniquely as $j + \sum_{i=0}^{n} i$ with $0 \leqslant j \leqslant n$, since $\sum_{i=0}^{n} i \xrightarrow[\infty]{n} \infty$ and is incresing, so that for any number m, there is a unique n such that $\sum_{i=0}^{n} i \leqslant m < \sum_{i=0}^{n+1} i$, so that $j = m - \sum_{i=0}^{n} i < n + 1$.

Now, an infinite arithmetic progression has form $a, a + s, a + 2s, ...$ We write $a = j + \sum_{i=0}^{n} i$. We consider $m \geqslant 2\max(s, n)+1$. There must be a point $a+ks$ of the progression such that $a+ks < \sum_{i=0}^{m} i \leqslant a+(k+1)s$, since $a \leqslant \sum_{i=0}^{n+1} i \leqslant \sum_{i=0}^{m} i$ (as $2\max(1, n) \geqslant n + 1$) and the progression is increasing and diverges. But since $s \leqslant m$, we have $a + (k + 1)s = s + (a + ks) \leqslant m + \sum_{i=0}^{m} i$, so that $a + (k + 1)s$ is in the monochromatic interval between $\sum_{i=0}^{m} i$ and $m + \sum_{i=0}^{m} i$.

We will now show that $a + ks \geqslant \sum_{i=0}^{m-1} i$, so that $a + ks$ is in the monochromatic interval between $\sum_{i=0}^{m-1} i$ and $m - 1 + \sum_{i=0}^{m-1} i$, so that $a + ks$ and $a + (k + 1)s$ can't have the same color, by definition of our coloring and the fact that m and $m - 1$ have different parity.

The reason this is the case is that otherwise, $a + ks < \sum_{i=0}^{m-1} i < \sum_{i=0}^{m} i \leqslant a + (k + 1)s$, so that by taking differences, $s \geqslant m$ which is impossible.

9 Hales-Jewettes theorem

The Hales-Jewettes theorem is about colorings of gridpoints of a hypercube and the existence of monochromatic lines in it. We explain precisely what we mean by "line", state the theorem and show how it can be used to prove Van der Waerden's theorem, which will motivate the interest in the particular type of line we're about to define.

We consider the cube $C_t^n = [t]^n$ and an r-coloring of its points. We'll be intereseted in **combinatorial lines** L of the cube, which are made of t points $L(i) \in C_t^n$, such that for all dimensions j, the sequences $(L(i)_j)_{i\in[t]}$ are either constant or strictly increasing, for which case there is only the possibility $L(i)_j = i$, and such that at least on dimension is of the increasing type (otherwise, we'd get a single point).

Note that this does not cover all lines in the intuitive sense: for dimension $n = 2$, the diagonal $(1, t), (2, t-1)..., (t, 1)$ isn't a combinatorial line.

There is a nice notation for combinatorial lines, which also allows us to count them: we can represent lines by a vector of $([t] \cup \{*\})^n$, where the entries in $[t]$ represent values of the constant coordinate sequence, and $*$ represents a increasing coordinate sequence. For example, line $(1, 1), (1, 2), (1, 3)$ of C_3^2 will be represented as $(1, *)$. The constraint from the definition is that there must be at least one $*$ in that vector. So among the $(t+1)^n$ words of $([t] \cup \{*\})^n$, the t^n that don't contain a $*$ aren't lines, so that there are $(t+1)^n - t^n$ lines.

Hales-Jewettes theorem:

For any t and r, there is a number $HJ(t, r)$ such that for all $n \geqslant HJ(t, r)$, we can find a monochromatic combinatorial line in any r-coloring of C_t^n.

The Hales-Jewettes theorem actually implies the Van der Waerden theorem. The link between the two can be made by establishing a correspondence between numbers and gridpoints of a hypercube: the representation of numbers in bases. Indeed, we can associate a point $x \in C_t^n$ with the number $\sum_{i=0}^{n}(x_i - 1)t^i$, where $(x_i - 1) \in \{0, .., t-1\}$, and vice versa, for numbers in $0, ..., t^n - 1$. A coloring of one translates to a coloring of the other.
A combinatorial line with $*$-set S then corresponds to numbers $\tau\left(\sum_{i\in S} t^i\right) + \sum_{i\notin S}(x_i - 1)t^i$ with range $\tau \in \{0, .., t-1\}$, which form an arithmetic progression of length t for $a = \sum_{i\notin S}(x_i - 1)t^i$ and $s = \sum_{i\in S} t^i$. The converse isn't true, as not all steps s can be represented as a$\sum_{i\in S} t^i$ for some $*$-set S. Now, finding a monochromatic line in C_t^n will provide us with a monochromatic progression.

The proof we'll give is known as Shelah's proof.
COMPLETE!

10 Infinite graphs

An infinite graph is a set of vertices V and a set of edges E between them, where these sets many now be infinite. Since there are multiple kinds of infinity, we distinguish between the countably infinite graphs, for which V is countable, and the uncountable ones. Note that for the countable ones at least, E can't be uncountable, as the neighbourhood of each is countable.

10.1 Coloring

The notion of a (proper) k-coloring extends without problems to infinite graphs.

Since we have a lot of coloring results available for finite graphs, it's natural to ask if we can deduce k-colorability of the infinite graphs from that of it's finite subgraphs. Clearly, if the infinite graph infinite graph is k-colorable, then so are it's finite subgraphs, as we can use the same colors for them. The converse holds for countably infinite graphs:

k-colorability of infinite graphs:

If all finite subgraphs of a countably infinite graph are k-colorable, then the infinite graph is too.

Note that this type of result isn't really intuitive or systematic. For example, consider the property of having a leaf, aka. a vertex of degree one. The infinite graph with vertices $\mathbb{Z}$ and edges between i, j if $|i - j| = 1$ (a line) has all its finite subgraphs satisfying this property, yet it doesn't satisfy it itself.
Now, our goal is to build such a coloring by listing vertices as v_n and coloring them one after the other. At each stage, we know that the subrgarph induced by the first $n + 1$ vertices is k-colorable, but we have no control on this coloring being the same for the first n vertices as in the previous stage.

To make things more clear, we introduce partial k-colorings C_n, which are k-colorings on the graph induced by vertices $v_1, ..., v_n$. We can extend these definitions by calling partial colorings the colorings on induced subgraphs and saying that one extends the other if it colors all the others vertices at least, and with the same colors.
We say that C_{n+1} extends C_n if they have the same colors on $v_1, ..., v_n$. If we manage to find a sequence of colorings $(C_n)_{n\in\mathbb{N}}$ such that C_{n+1} extends C_n for all indices n, then we can consider the coloring $C(v_n) = C_n(v_n)$. It's a k-coloring as $C(v_n) = C_n(v_n) \in [k]$ and or all $\{v_n, v_m\} \in E$, $C(v_n) = C_n(v_n) = C_{\max(n,m)}(v_n) \neq C_{\max(n,m)}(v_m) = C_m(v_m) = C(v_m)$.
So, *how do we show that such a sequence of partial colorings extending their predecessors exists ?*

Start with vertex v_1: which color should we give it ? We know that for all n, there is a partial coloring Q_n of $v_1, ..., v_n$, since our assumption is that all finite subgraphs are k-colorable, but they have the flaw that the color of v_1 may differ from one coloring to the next, so that they don't extend each other. However, if we consider the sequence $Q_n(v_1)$, on of the k possible values this can take must appear inifinitly often. If we extract the subsequence in which v_1 has the same color with φ, we get an infinite sequence $Q_{\varphi(n)}$ of partial colorings in which v_1 has the same color.
As you might be able to see, we could then perform the same argument on v_2 and the sequence of $Q_{\varphi(n)}$ and similarly on and on. The point is that when we set C_1 to be the partial coloring which colors only v_1 with its color in the $Q_{\varphi(n)}$, we will be able to extend this partial coloring with this type of argument.

The key property of the C_n we'll build is that there are an infinite number of partial colorings extending them. This is true for C_1, as the infinite sequence $Q_{\varphi(n)}$ of partial colorings extends it, since v_1 has the

same color in all of them. We can show this for the other C_n by induction, where only the induction steps remains to be proved. So if C_n has an infinite number of partial colorings extending it, that we collect in sequence $(K_m)_m$, we look at $(K_m(v_{n+1}))_m$. One of the k colors must appear infinitely often in that sequence, and we can extract those colorings in which v_{n+1} has that color with φ. Then by letting $C_{n+1}(v_{n+1})$ be that color, $(K_{\varphi(m)})_m$ is a infinite sequence of partial colorings extends it, as the colors on the first $n+1$ vertices are the same, by construction for v_{n+1} and by the inductive assumption for $v_1, ..., v_n$. In particular, C_{n+1} will extend C_n, and our strategy works out.

10.2 More

From the chapter in Diestels book ?

11 Positional games

11.1 Tic-tac-toe

With high probability, you've played tic-tac-toe before, so we won't explain the rules of the traditional game to you. It should also come as no surprise to you that every player has a strategy that garagtees them a draw as outcome. This can be checked by case disjunction. If we considered the game on a 2×2 grid with similar rules, then case disjunction would lead us to show that the first player always wins: after the first two turns, in all 12 cases the grid can be filled out it, the first player can win in the third turn.

What about a game on an $n \times n$ grid ?
The rules are the same, except that the lines one has to fill out are now of length n. An astonishing fact is that one can prove, without massive case disjunctions, that each player has a strategy to make the game end in a draw no matter how their opponent plays.

Consider the viewpoint of the second player: how should they prevent the first player from getting any line filled out ? The least they must do to prevent a line from being filled out, is to get one tile-point on that line. If we had a set of distinct such tile-points for each line, then the second player could try to fill them all to guarantee that the first player can't fill out a line.
However, *what if the first player would fill out one fo these points ? Can we fix our idea ?*

If we got two points per line, then whenever the first player fill one of them, we can fill the other in the next turn, thereby preventing the line from being filled out by the first player later on. This is a strategy from positional game theory known as a **pairing startegy**.
Does $n \times n$ tic-tac-toe have such a set of pairs of points ?

We may not find one for $n \leqslant 3$, or even 4, but for 5 and 6 we can find some:
Pairing up points by marking them with the same numbers, where we use $*$ for points not needed in the strategy, for 5, we have $\begin{matrix} 11 & 1 & 8 & 1 & 12 \\ 6 & 2 & 2 & 9 & 10 \\ 3 & 7 & * & 9 & 3 \\ 6 & 7 & 4 & 4 & 10 \\ 12 & 5 & 8 & 5 & 12 \end{matrix}$, and for 6 we have $\begin{matrix} 13 & 1 & 9 & 10 & 1 & 14 \\ 7 & * & 2 & 2 & * & 12 \\ 3 & 8 & * & * & 11 & 3 \\ 4 & 8 & * & * & 8 & 4 \\ 7 & * & 5 & 5 & * & 12 \\ 13 & 6 & 9 & 10 & 6 & 13 \end{matrix}$.

To find these in practice, we could have modeled this as an integer program. There are $2n+2$ lines (n horizontal, n vertical, 2 diagonals), so that we need $n+1$ indices to label the pairs. We use binary variables $x_{i,j,m}$ which indicate if tile-point $(i,j) \in [n]^2$ has label m where m takes at most $n+2$ (don't forget $*$). Next, to indicate whether line L has two identically labeled m points on it, we use binary variable $d_{L,m}$ .We get n^2 constraints $\sum_m x_{i,j,m} = 1$ (exactly one label per point), and $(2n+2)(2n+5)$ constraints $\sum_{(i,j)\in L} x_{i,j,m} \leqslant 2$, $\frac{1}{2}\left(\sum_{(i,j)\in L} x_{i,j,m}\right) \geqslant d_{L,m} \geqslant \left(\sum_{(i,j)\in L} x_{i,j,m}\right) - 1$ (for $m \neq *$) and $\sum_{m\neq *} d_{L,m} \geqslant 1$, (and the binary and integral constraints) to ensure that any feasible point of this polytope is a solution to our problem.

It turns out that these will be the initial cases for a construction by induction/recursion of a set of pairs of points for grids of size n. Indeed, given a set of labeled pairs of points for a grid of size n, we can

find one for a grid of size $n+2$, by filling out the inner grid by induction, and the outer layer as follows:

$$\begin{array}{ccccc} * & 1 & 1 & * & * \\ * & ind & ... & ind & 2 \\ 4 & \vdots & & \vdots & 2 \\ 4 & ind & ... & ind & * \\ * & * & 3 & 3 & * \end{array}$$

Here, the *ind* are determined by induction. They already cover all lines of the initial grid, except for the 4 that make up the outer layer, which we get as with the labels we presented.

Since the step from n to $n+2$ preserves parity, and we have an initial odd pairing for $n=5$, and an initial even one for $n=6$, we conclude that such a pairing exists for all $n \geqslant 5$.

This way the second player can guarantee a draw, whatever the first does. If the first player fills on of the points from the set, the second fills its paired up point, if it hasn't been filled by the second player already, so that the line can't be filled by the first player. If the first player fills a point not from the set, the the second player can fill out an arbitrary point, from the set or not. This is how it may potentially happen that the point has already been filled out we second is in the case where it's supposed to fill out that point as the pair has just been filled out. In all cases, either the first player doesn't fill out any point on a given line, or the second player has or will fill another point of the line, if red fills it at a point at some turn, so that the first player can't completely fill out the line. Thus, in the worst case, the game ends in a draw for the second player.

What about the perspective of the first player ? The first player can adopt the same strategy as the second in the third turn, since this strategy dealt with the case of a point of a pair already being filled, as might be the case for the first player after the first turn.
Another approach is to note that there is always a $*$ for $n \geqslant 5$, which the first player can fill out, so that a pairing exists from they're perspective.

Since both players can guarantee themselves a draw in the worst case, with this straegy, none of them can win, when they both play with this strategy, so that the game must end in a draw.

We extended the grid-size of tic-tac-toe. *What if we extend the dimension ?*
We can play tic-tac-toe on a d-dimensional grid of size n. Here, generalizing the winning lines requires more precision. In 3-dimensional tic-tac-toe, we want the diagonal line obtained by fixing one coordinates, and letting the other two vary from 1 to n, to be winning sets, for example, though they're not in a slice of the grid, nor diagonals of the cubic grid.

The notion of lines we'll use is that of a sequence of gridpoints $a^{(1)}, ..., a^{(n)}$ with the following property: for each coordinate j, the sequence $a_j^{(1)}, ..., a_j^{(n)}$ is either a constant one, for constants in $[n]$, or it's $1, ..., n$ ("increasing") or $n, ..., 1$ ("decreasing"). Also, not all coordinate sequences can be constant, as this would produce the same girdpoint in all the sequence. More specifically, those are oriented lines (the sequence induces a direction). Some of them will correspond to the same line, such as the lines in which all but one of the coordinate sequences are constant: the directed lines where the non-constant coordinate sequence is increasing and decreasing actually correspond to the same set of gridpoints (in a different order). In general, two directed lines correspond to the same line if (and only if) their non-constant coordinate sequences are, when considered side by side for a fixed coordinate, increasing in one and decreasing in the

other.

First, let's ask: *how many lines are there in a d-dimensional grid of size n ?*
To build the directed lines we must chose the c coordinates we want to be constant, where c ranges over $[d-1]$, and for each such choice, there are n possible values for the constant. For these choices, there are the 2 choices (increasing/decreasing) for the other $d-c$ coordinates. So the total number of directed lines is $\sum_{c=0}^{d-1} \binom{d}{c} n^c 2^{d-c}$, which we can express, using the binomial expansion, as $(n+2)^d - n^d$. The number of lines is then half of the latter, as each line corresponds to exactly two directed lines.

As a "fun fact" let's convince ourselves that these lines are either equal, or meet in at most one point.
Assume two lines A and B where to meet at at least two points $a^{(i)} = b^{(i)}$ and $a^{(j)} = b^{(j)}$. Let's zoom in on the coordinates q. If a_q was increasing, then $b_q^{(i)} < b_q^{(j)}$, so b_q must have been increasing, as it can't be constant or decreasing, and similarly if a_q was decreasing. If a_q was constant with value m, then $b_q^{(i)} = b_q^{(j)} = m$, so the only possible type of b_q is constant at value m. Therefore, A and B are of same type on each coordinate, and they must be equal.

Let's get back to the game now. We'll try to find a pairing as in the 2-dimensional case, though with more sophisticated combinatorics. We'll try to find representatives among the points for the lines, which from the chapter on Hall's theorem, should lead us to the following.
We'll represent the situation as a bipartite graph with bipartition set U representing te lines, where we include 2 vertices per line, and bipartition set W representing the points, and edges between them if the point is on the line. The reason we included two vertices per line is that we'll try to match U into W, so that a line will be matched to two points in the end, once for each representing vertex.
We'll try to apply Hall's theorem, for which we need to verify the Hall condition. We'll try to get that condition from double-counting edges and using bounds on them. The bound we'll need is:

Bound:

The number of lines a point is in upper-bounded by $\frac{3^d-1}{2}$.

For each directed line passing through a given point, the coordinate sequences are of three types: increasing, deceasing, or constant with value that point's coordinate. There are at most 3^d such combinations of coordinate sequences. We discard the one in which all sequences are constant, so that we can upper-bound the number of directed lines by $3^d - 1$, and therefore the number of lines, which is half of directed ones, by $\frac{3^d-1}{2}$. This migth be an over-estimation for a given point. For example, for a point with its two first coordinate 2 and 1, it can't be on any lines in which the two first coordinate-sequences are increasing, as the first coordinate-sequence requires it to be the second point on the line, while the second coordinate-sequence requires it to be the first point on the line. However, in the case of an odd grid, the center of the grid truely is in $\frac{3^d-1}{2}$ lines, so that this bound is tight.

Back to Hall's application in our context. To compare the size of line-vertex set S to its neighbourhood $N(S)$ of points in those lines, we'll estimate the edges between them in two ways. First, since each line contains n points, we know that $\sum_{L \in S} |\delta(L)| = |S|n$.

On the other bipartition set, these edges are counted in $\sum_{p \in N(S)} |\delta(p)|$, which may be larger then $\sum_{L \in S} |\delta(L)|$, as it may be that $S \subsetneq N(N(S))$. $\sum_{p \in N(S)} |\delta(p)|$ can be bounded by recalling that any point is in at most $\frac{3^d - 1}{2}$ lines, so that $\sum_{p \in N(S)} |\delta(p)| \leqslant |N(S)| \frac{3^d - 1}{2}$. Combining the bounds, we get $|S| \frac{2n}{3^d - 1} \leqslant |N(S)|$. So in the case that $1 \leqslant \frac{2n}{3^d - 1}$, we get Hall's condition, and a pairing strategy exists.

High dimesnional tic-tac-toe:

For a d-dimensional grid of size n, with $n \geqslant \frac{3^d - 1}{2}$, both the first and second player have a strategy that guarantees them a draw in the worst case, since a pairing startegy exists.

11.2 Maker-breaker games

12 Social choice with Boolean analysis

12.1 Social choice

Social choice is the study of how to make take a decision based on opinions of a group of people. In the simplest case, n voters have to chose between two options 1 and -1, and given their collected opinions in the form of a vector x of $\{-1,1\}^n$, we use a social choice function $f : \{-1,1\}^n \to \{-1,1\}$ to determine which option $f(x)$ will be taken as a community.

We start by listing some examples:

Some social choice functions:

- The two **majority** functions $Maj_{\pm}(x) = sign_{\pm}\left(\sum_{i\in[n]} x_i\right)$, one per choice of sign, where we handle ties with arbitration, in the sense that $sign_{\pm}(y) = \begin{cases} 1, y > 0 \\ -1, y < 0 \\ \pm 1, y = 0 \end{cases}$.
- The **weighted majority** functions $Maj_{\pm,c}(x) = sign_{\pm}\left(c_0 + \sum_{i\in[n]} c_i x_i\right)$ which depend on a weight vector c. Note that c_0 introduces the possibility of a bias.
- The k**-junta** functions $J_I(x) = g(x_I)$ for some $I \subseteq [n]$ and some $g : \{-1,1\}^{|I|} \to \{-1,1\}$.
- In particular, the i**th dictator** function $d_i(x) = x_i$.
- We can define $and_n(x) = \begin{cases} -1, x = (-1,...,-1) \\ 1, \ else \end{cases}$ and $or_n(x) = \begin{cases} 1, x = (1,...,1) \\ -1, \ else \end{cases} = and_n(-x)$ to obtain the family of **tribe** functions $\tau_{(T_i)}(x) = or_t(and_{|T_1|}(x_{T_1}),...,and_{|T_t|}(x_{T_t}))$, where the (T_i) partition $[n]$ into t "tribes". Here, option -1 is chosen when there is a tribe in which all members desire -1, otherwise option 1 is the default.

We can describe properties for these functions:

Some properties for social choice functions:

- A social choice function f is **symmetric** if for all permuation matrices M_π (π is a permutation of $[n]$), we have $f(M_\pi x) = f(x)$. So the ordering of voters doesn't affect the choice.
- A social choice function f is **monotone** if for any $x \leqslant y$ (coordinate-wise), $f(x) \leqslant f(y)$.
  So if some voters change their vote from ± 1 to ∓ 1, where all voters that changed their minds voted and vote the same, then the result of the social choice is in their favour, in the sense that it can't go to ± 1, having previously been at ∓ 1.

- A social choice function f is **unanimous** if $f(\pm 1, ..., \pm 1) = \pm 1$.
  So if all voters agree, the social choice is the unanimous one.
- A social choice function f is **transitive-symmetric** if for all voters i, j, there is a permutation π such that $\pi(i) = j$ and $f(M_\pi x) = f(x)$ for all $x \in \{-1, 1\}^n$. This means that if voters i took j's rank, there would be a way of reordering th rest of voters so as to get the same social decision, in any voting scenario.

For example, the majority functions have all these properties, since $\sum_{i \in [n]} x_i$ is invariant under permutation, $sign_\pm$ is increasing and so is $\sum_{i \in [n]} x_i$ under the partial coordinate order, and $Maj_\pm(\pm 1, ..., \pm 1) = \pm 1$.

Transitive-symmetry and symmetry are not the same. A function that's symmetric will also be transitive symmetric, as any permutation does the job. The converse is however false. The tribe functions are a good example. They're not symmetric, as permuting two voters of different tribes leads to different results if one of these tribes (where voters where permuted) voted -1 uninimously, while all other tribes voted 1 unanimously. However, they're transitive-symmetric, as we can simply permute the rest of the two tribes, so as to essentially end up with the same tribes, but reordered, so that results are the same, by symmetry of or_t.

We can also measure numerical quantities about social choice functions. For example, for such an f, we can measure the impact that voter i has by considering the ratio of how often a change in vote of i results in a change in social decision over all choices of other voters, to the number of all possible scenarios for other voters. This is:

Influence:

The **influence/impact** of voter i is

$$Imp(f, i) = P_{X \sim U(\{-1,1\}^n)}(f(X_1, ..., 1_i, ..., X_n) \neq f(X_1, ..., -1_i, ..., X_n))$$

.

The **total influence** of f is $I(f) = \sum_{i \in [n]} Imp(f, i)$.

We can then ask to find a social choice function maximising total influence, under constraints on this functions, such as it being monotonous and unanimous, for example.

Finally, we'll discuss an approach of dealing with social choice between more then two options.
In the **Condorcet method** to determine an option among k of them, we compare options pair-wise, by using social choice function f to determine which option is preferred in a pair, for each pair.

We can then picture the situation with a digraph in which options are represented by node, and we draw an arc from on to the other if in the pair-comparison-vote, voters preferred the tail of the arc to its head. The graph obtained this way is a tournament. It may actually have cycles, corresponding to people

preferring an option a to b, b to c, but then also c to a. This is the case in the following example, where 3 voters v_i choose between 3 options, and we use the Condorcet method with a majority function:

	v_1	v_2	v_3	*result*
a (1) *vs.* b (-1) :	1	1	-1	$a > b$
b (1) *vs.* c (-1) :	1	-1	1	$b > c$
c (1) *vs.* a (-1) :	-1	1	1	$c > a$

This is actually a coherent scenario. If v_1 has preferences $a > b > c$, v_2 has preferences $c > a > b$, and v_3 has preferences $b > c > a$, then these are the votes we can expect to see.

A **Condorcet winner** is an option that is a source of the digraph, aka. an option that won all decisions between it and the other pairs. If it exists, it's unique (tow would have to beat each other), but as we've seen from the previous example, a Condorcet winner may not exist. This outcome did a priori depend on the social choice function used in the "matches".
We may then ask, *what are the social choice functions for which a Condorcet winner always exists, for all* $x \in \{-1, 1\}^n$ *?* This is what Arrow(-Kalai)'s theorem will be about.

We will soon discuss Boolean analysis, which investigates mimicking real analysis on Boolean functions, which is essentially what social choice functions are. The link to social choice theory is for example one like the following.
We can consider a discrete partial derivative of a Boolean function with $\partial_i f : \{-1, 1\}^n \to \{-1, 1\}$ where $\partial_i f(x) = \dfrac{f(x_1, ..., 1_i, ..., x_n) - f(x_1, ..., -1_i, ..., x_n)}{2}$. Its square $(\partial_i f(x))^2$ takes value 0 when a change in i doesn't affect the outcome, and value 1 if it does.
This allows us to rewrite $Imp(f, i) = E_{X \sim U(\{-1,1\}^n)} \left((\partial_i f(X))^2 \right)$. So the study of $\partial_i f$ will lead to results about influence.
Many more such connections will arise. For now, we start by studying Boolean analysis, in which we try mimicking real analysis on Boolean functions.

12.2 Boolean analysis

We start our study of Boolean functions by introducing the **multilinear expansion** or **Fourier expansion** of a Boolean function. Its second name gives away the fact that we'll use it to do stuff resembling Fourier analysis.
This expansion expresses the a Boolean function f as a (real) linear combination of monomials $x^S = \prod_{i \in S} x_i$ for $S \subseteq [n]$. *Does this expression even exist ? Is it unique ?*

Existence can be shown from a slightly different consideration: can we express the indicator functions of singletons $\{a\}$ for $a \in \{-1,1\}^n$ as such polynomials ?
If we can, then since $f(x) = \sum_{a \in \{-1,1\}^n} f(a)\chi_{\{a\}}(x)$, and the polynomials with monomials of multidegree at most $(1, ..., 1)$ form a vectors space, so that linear combinations of them result in the same type of polynomials, we can express any f as such a polynomial.
We can express $\chi_{\{a\}}(x)$ as such a polynomial: $\chi_{\{a\}}(x) = \prod_{i \in [n]} \left(\frac{1 + a_i x_i}{2} \right)$, since $\left(\frac{1 + a_i x_i}{2} \right) = 1$ if $x_i = a_i$ and $\left(\frac{1 + a_i x_i}{2} \right) = 0$ if $x_i \neq a_i$ since we deal with values ± 1.

Uniqueness will bring us closer to analysis. The journey starts in probabilistic considerations, which we've seen to show up in social choice considerations. If we pick a hypercube point X uniformly at random, so that $X \sim U(\{-1,1\}^n)$, we can study the random value of the function $f(X)$. For two such Boolean functions f and g, we now that $E(f(X)g(X))$, which shows up in the covariance of $f(X)$ and $g(X)$, can be interpreted as a dot-product of $f(X)$ and $g(X)$.
By considering Boolean functions $\{-1,1\}^n \to \mathbb{R}$ as a vector space, we can use this as inspiration to define the dot-product $\langle f, g \rangle = \frac{1}{2^n} \sum_{x \in \{-1,1\}^n} f(x)g(x)$ on this space. This can also be derived from interpreting Boolean functions as vectors from $\mathbb{R}^{(\{-1,1\}^n)}$.

The slightly surprising thing, which justifies the Fourier name, is that the family of monomials x^S is orthonormal for this dot-product. Indeed, $x^S x^F = \prod_{i \in S} x_i \prod_{i \in F} x_i = \prod_{i \in S \backslash F} x_i \prod_{i \in S \cap F} x_i^2 \prod_{i \in F \backslash S} x_i$ and since Boolean functions take values in $\{-1,1\}$, we get $x^S x^F = \prod_{i \in S \backslash F} x_i \prod_{i \in F \backslash S} x_i$. Now, unless $S \backslash F$ and $F \backslash S$ are empty, aka. unless $S = F$, we can fix an element j in one of them, and partition $\{-1,1\}^n$ in half according to the value of the jth coordinate: for each $a \in \{-1,1\}^n$, there is an $a' \in \{-1,1\}^n$ that differs only in he jth coordinate, so that this pair of terms cancels each other out in the sum $\frac{1}{2^n} \sum_{x \in \{-1,1\}^n} x^S x^F$, since the monomials will differ in sign, as they contain x_j. So unless $S = F$, $\langle x^S, x^F \rangle = 0$. Now, computing $\langle x^S, x^S \rangle$ resulting in $\frac{1}{2^n} \sum_{x \in \{-1,1\}^n} 1 = 1$ for all $S \neq \emptyset$, since $x^S x^S = \prod_{i \in S} x_i^2 = 1$ in that case, and by adopting the convention that $\prod_{i \in \emptyset} = 1$, this holds for any $S \subseteq [n]$.

The probabilisitic view can speed up the proof of orthonormality.
We compute $E_{X \sim U(\{-1,1\}^n)} \left(X^S X^F \right) = E \left(\prod_{i \in S \backslash F} X_i \prod_{i \in F \backslash S} X_i \right)$ by using independence of coordinates for

$X \sim U(\{-1,1\}^n)$, so that $E_{X\sim U(\{-1,1\}^n)}\left(X^S X^F\right) = \prod_{i\in S\setminus F} E(X_i) \prod_{i\in F\setminus S} E(X_i)$. Since $E(X_i) = \frac{1}{2} - \frac{1}{2} = 0$, this expectatio is 0 unless the product is empty, in which case its $E(1) = 1$.

Now if f could be written in two ways as a expansion, say $\sum_{S\subseteq[n]} c_S x^S$ and $\sum_{S\subseteq[n]} c'_S x^S$, we could take the dot-product with some x^F on both expressions to conclude that $c_F = c'_F$. This allows us to conclude with uniqueness.

Fourier expansion:

The Fourier expansion of a Boolean function exists and is unique.

The coefficients of the expansion are indexed by $S \subseteq [n]$ and could therefore be considered as Boolean functions. Indeed, we can represent Boolean function with the notation $f(S) = f\left(\left(\begin{cases} 1, i \in S \\ -1, i \notin S \end{cases}\right)_{i\in[n]}\right)$.
The problem is that they're not necessarily ± 1 valued, as the following example shows: as you can check by building it as we explained, in dimension 2, $Maj_+(x) = \frac{1}{2}(1 + x_1 + x_2 - x_1x_2)$.
We will still use function like notation for the Fourier coefficients of the expansion, calling them $\hat{f}(S) = c_S$. Also the notation suggest, they will have similar properties to the Fourier transform of a function. Just note that $\hat{f}$ isn't a Boolean function, as it may not be ± 1 valued.

Let's collect some definitions we've seen to be useful:

Definitions:

Boolean functions have unique **Fourier expansions** $f(x) = \sum_{S\subseteq[n]} \hat{f}(S)x^S$, where $x^S = \prod_{i\in S} x_i$.
We call $\hat{f}$ the **Fourier coefficient function** of f.
We have dot-product $\langle f, g\rangle = E_{X\sim U(\{-1,1\}^n)}(f(X)g(X))$ and norms $\|f\|_p = \sqrt[p]{E_{X\sim U(\{-1,1\}^n)}(|f(X)|^p)}$.

Let's see how far the analogy to Fourier analysis goes.

Some consequences of orthonormality are $\langle f, x^S\rangle = \hat{f}(S)$, **Plancharel's identity** $\langle f, g\rangle = \sum_{S\subseteq[n]} \hat{f}(S)\hat{g}(S)$, and **Parseval's identity** $\langle f, f\rangle = \sum_{S\subseteq[n]} \hat{f}(S)^2$, which for Boolean f implies $\sum_{S\subseteq[n]} \hat{f}(S)^2 = E\left(f(X)^2\right) = E(1) = 1$.

We take a brief moment to discuss the probabilistic view.
We have $E(f) = \langle f, 1\rangle = \hat{f}(\emptyset)$ and $Var(f(X)) = \langle f, f\rangle - E(f)^2 = \sum_{\emptyset\neq S\subseteq[n]} \hat{f}(S)^2$.
For Boolean functions, variance has an interesting alternative expression: with $E(f) = P(f(X) = 1) - P(f(X) = -1)$, $\langle f, f\rangle = 1 = P(f(X) = 1) + P(f(X) = -1)$ and identity $a^2 - b^2 = (a + b)(a - b)$, we

get $Var(f(X)) = 4P(f(X)=1)P(f(X)=-1)$.
We can also consider $Cov(f(X),g(X)) = \langle f,g\rangle - E(f)E(g) = \sum_{\emptyset \neq S \subseteq [n]} \hat{f}(S)\hat{g}(S)$.

What about expectations over other probability distributions ?
We'll base our approach on relative densities, so as to keep uniform density close. To get a probability distribution over $\{-1,1\}^n$, we can take any $\phi : \{-1,1\}^n \to \mathbb{R}_+$ that satisfies $E(\phi(X)) = 1$, and have use $Y \sim \phi$ to denote that Y has probability distribution $P(Y=y) = \frac{\phi(y)}{2^n}$. Any probability distribution on $\{-1,1\}^n$ can be expressed this way, by defining $\phi(y) = 2^n P(Y=y)$.
If we then seek to compute $E_{Y\sim\phi}(g(Y))$, we have identity $E_{Y\sim\phi}(g(Y)) = \frac{1}{2^n}\sum_{y\in\{-1,1\}^n} \phi(y)g(y) = \langle \phi, g\rangle$.

The Fourier analysis analogy would be incomplete without convolutions, so let's define them here. Had we considered Boolean functions with domain the vectors of $\mathbb{F}_2$, we would have been tempted to define $f\star g(x) = E_{y\sim U(\mathbb{F}_2^n)}(f(y)g(x-y))$. We can still do something similar over the Hamming cube $\{-1,1\}^n$ by noting that by associating $0_{\mathbb{F}_2} \approx 1$ and $1_{\mathbb{F}_2} \approx -1$, we have $x_{\mathbb{F}_2} - y_{\mathbb{F}_2} \approx x \circ y$, where $\circ$ is sometimes called the Hadamar product, which is coordinate-wise multiplication. So our convolutions will be $f\star g(x) = E(f(Y)g(x\circ Y))$.

We can check some of their properties. For example $(f\star g)(x) = (g\star f)(x)$, since $y \mapsto x\circ y$ is a bijection ($x\circ(y-z) = 0 \Rightarrow y = z$ and $y = x\circ(x\circ y)$), we may change variables in the sum $\sum_{y\in\{-1,1\}^n} f(y)g(x\circ y) = \sum_{z\in\{-1,1\}^n} f(x\circ z)g\left(\underbrace{x\circ(x\circ z)}_{z}\right)$. There's also associativity to show, but we'll just take it for granted.
Probability densities have a nice property in this context. If $Y\sim\phi$ and $Z\sim\psi$ are independent, then $Y\circ Z \sim \phi\star\psi$, which is the essence of convolutions. We have to show that $P(Y\circ Z = x) = \frac{(\phi\star\psi)(x)}{2^n}$ to get this result, aka. $P(Y\circ Z = x) = \frac{1}{2^n}\left(\frac{1}{2^n}\sum_{t\in\{-1,1\}^n}\phi(t)\psi(x\circ t)\right)$. First, we split $P(Y\circ Z = x) = \sum_{t\in\{-1,1\}^n} P(Y=t)P(Y\circ Z = x|Y=t)$, where $P(Y\circ Z = x|Y=t) = P(t\circ Z = x|Y=t)$ and $P(t\circ Z = x|Y=t) = P(Z = t\circ x|Y=t) = P(Z = t\circ x)$ with the bijection (idempotence) of $\circ$ in this context, and independence. So in the end $P(Y\circ Z = x) = \sum_{t\in\{-1,1\}^n} P(Y=t)P(Z=t\circ x) = \frac{1}{2^n}\frac{1}{2^n}\sum_{t\in\{-1,1\}^n}\phi(t)\psi(x\circ t)$, as desired.

Finally, we show the most important property that makes the Fourier analysis analogy complete: $\widehat{(f\star g)}(S) = \hat{f}(S)\hat{g}(S)$. To see it, we use Fourier expansions and binlinearity of convolution to get $f\star g(x) = \sum_{S,F\subseteq[n]} \hat{f}(S)\hat{g}(F)E\left(Y^S(x\circ Y)^F\right)$. Next, we have $(x\circ Y)^F = x^F Y^F$ (unpack the definitions), so that, recognizing the dot-product along the way $f\star g(x) = \sum_{S,F\subseteq[n]} \hat{f}(S)\hat{g}(F)x^F\langle x^S, x^F\rangle$, so that orthonormality implies $f\star g(x) = \sum_{F\subseteq[n]} \hat{f}(F)\hat{g}(F)x^F$. By uniqueness of the Fourier expansion, we conclude that $\widehat{(f\star g)}(S) = \hat{f}(S)\hat{g}(S)$.

Again, let's collect the properties we just proved:

Fourier-ish porperties:

COMPLETE

We can also introduce other notions of distance between Boolean functions:

Relative Hamming distance:

The **relative Hamming distance** of f and g is $dist(f,g) = P(f(X) \neq g(X))$.

We already mentioned the derivative $\partial_i f(x) = \frac{1}{2}(f(x_1,...,1_i,...,x_n) - f(x_1,...,-1_i,...,x_n))$.
It has similar properties to the real derivative. For example, f is monotone precisely when $\partial_i f(x) \geqslant 0$ for all i and x. Indeed, if f is monotone, since $(x_1,...,1_i,...,x_n) > (x_1,...,-1_i,...,x_n)$, we have $f(x_1,...,1_i,...,x_n) \geqslant f(x_1,...,-1_i,...,x_n)$ and thereofore $\partial_i f(x) \geqslant 0$. For the converse, take $x > y$ and choose an order on the indices where they differ, say x and y differ on $i_1,...,i_k$. We can use intermediates $z^{(q)}$ where $z^{(0)} = y$ and $z^{(q+1)}$ differs from $z^{(q)}$ on i_{q+1}, so that $y = z^{(0)} < z^{(1)} < ... < z^{(k)} = x$. We then use the fact that $\partial_{i_q} f\left(z^{(q)}\right) \geqslant 0$ and compute the telescopic sum $\sum_{q=1}^{k} \partial_{i_q} f\left(z^{(q)}\right) \geqslant 0$ to find that $f(x) \geqslant f(y)$.

Another similarity to Fourier analysis is the interaction between derivative and Frourier expansions.
First, let's express the Fourier expansion of $\partial_i f(x)$ as follows. Since ∂_i is linear, we can expand f and compute $\partial_i x^S$. Now, $\partial_i x^S = \frac{1}{2}\left(\prod_{j\in S, x_i=1} x_j - \prod_{j\in S, x_i=-1} x_j\right) = \begin{cases} 0, i \notin S \\ x^{S\setminus i}, i \in S \end{cases}$, so that $\partial_i f(x) = \sum_{S\subseteq[n]:i\in S} \hat{f}(S)x^{S\setminus i} = \sum_{F\subseteq[n]\setminus i} \hat{f}(F\cup i)x^F$. Therefore, $\widehat{\partial_i f}(S) = \begin{cases} 0, i \in S \\ \hat{f}(S\cup i), i \notin S \end{cases}$.

12.3 The theorems

Questions in social choice theory are typically formulated as finding all social choice functions satisfying a certain criterion. In a first example, we'll ask which functions maximise the expected number of voters which voted for the decision taken.

The first task is to express the quantity we want to maximise.
How do we indicate if i voted to the final decision ?
This happens when x_i and $f(x)$ have the same value, which is equivalent to $f(x)x_i = 1$. This is not yet an indicator, as it takes value -1 when the choices disagree. We'll still look at $f(x)\sum_{i=1}^{n} x_i$, which is the the number of voters that voted for the decision, minus that of voters that voted against it. If w denotes the first number, then the quantity we which to maximise is $E(w)$. We can the use then express $f(x)\sum_{i=1}^{n} x_i = w - (n-w)$ to get $E(w) = \frac{1}{2}\left(n + E\left(f(x)\sum_{i=1}^{n} x_i\right)\right)$.
So the problem is equivalent to maximising $E\left(f(x)\sum_{i=1}^{n} x_i\right)$. *Can you bound the latter ?*
With the absolute value, we have $E\left(f(x)\sum_{i=1}^{n} x_i\right) \leqslant E\left(\left|\sum_{i=1}^{n} x_i\right|\right)$, where $\left|\sum_{i=1}^{n} x_i\right|$ is also the number of voters that voted for the decision, minus that of voters that voted against it. We have $\left|\sum_{i=1}^{n} x_i\right| = (+1)\sum_{i=1}^{n} x_i$ if $\sum_{i=1}^{n} x_i \geqslant 0$, so if the number of voters that voted 1 is $\geqslant$ then that who voted -1 and similarly if $\left|\sum_{i=1}^{n} x_i\right| = (-1)\sum_{i=1}^{n} x_i$. This is precisely $f(x)\sum_{i=1}^{n} x_i$ when f is a majority function.
So the majority functions maximise $E(w)$.

A perhaps more surprising result is obtained by seeking to maximise the stability of a election procedure. First, we'll explain what one could understand under the notion of stability.
Assume the votes are mis-recorded as follows: for all i, independently with probability $p \in]0,1[$, the correct vote $y_i = x_i$ is recorded, but with probability $1-p$, the a random vote is recorded, which may be $y_i = 1$ or $y_i = -1$, both with probability $\frac{1}{2}$.This isn't a framework that models real-life mis-recording of votes, as there is usually a bias, so that y_i would depend on x_i's value. We'll stick to this model for its simplicity. We'd like to know how such a perturbation affects the outcome of the vote on average, in the sense that we want to compute $P(f(x) = f(y))$. *How do we proceed ?*

In a first reformulation, we'll make expectations appear, as they have dot-product properties. we can work with. Note that the product $f(x)f(y)$ is 1 if $f(x) = f(y)$ and -1 otherwise. So $E(f(x)f(y)) = P(f(x) = f(y)) - P(f(x) \neq f(y))$, and using that the sum of them is 1, we get $E(f(x)f(y)) = 2P(f(x) = f(y)) - 1$. So since $E(f(x)f(y))$ is proportional to $P(f(x) = f(y))$ and we expect it to be easier to work with, we define this to be the **stability** $Stab(f,p) = E(f(x)f(y))$.

In the sum $E(f(x)f(y))$, grouping the terms for a fixed x, we can factor $f(x)$ out in these terms, leaving us with $E(f(y)|x)$, so that $E(f(x)f(y)) = E(f(x)E(f(y)|x))$. Using the Fourier expansion of f and linearity of expectation, we can get a better picture of $E(f(y)|x)$ by computing $E\left(y^S|x\right)$ for $S \subseteq [n]$. We can then use Independence of the y_i among each other to get $E\left(y^S|x\right) = \prod_{i\in S} E(y_i|x)$. Next, we have $E(y_i|x) =$

$px_i + (1-p)\left(\frac{1}{2} - \frac{1}{2}\right) = px_i$, so that $E\left(y^S|x\right) = p^{|S|}x^S$. In conclusion $E(f(y)|x) = \sum_{S\subseteq[n]} p^{|S|}\hat{f}(S)x^S$, so that by on more Fourier expansion and orthogonality, $E(f(x)f(y)) = \sum_{S\subseteq[n]} p^{|S|}\hat{f}(S)^2$.

When is stability maximised by a Boolean function ?

We have a surprising result:

Stability maximisation:

The **stability** $Stab(f,p) = E(f(x)f(y))$ is uniquely maximised over all **unbiased** Boolean function, unbiased in the sense that $E(f(x)) = 0$, by the linear functions, and in particular the dictator functions.

Unbiasedness means that for uniformly random votes, the probabilities of the decisions are equal, $P(f(x) = 1) = P(f(x) = -1)$.
Proof: Recall that $E(f(x)) = \hat{f}(\emptyset)$, since $E(f(x)) = E(f(x).1) = E\left(f(x)x^{\emptyset}\right)$.
A bound on $\sum_{S\subseteq[n]} p^{|S|}\hat{f}(S)^2$ is obtained by using unbiasedness to get $\sum_{S\subseteq[n]} p^{|S|}\hat{f}(S)^2 = \sum_{\emptyset\neq S\subseteq[n]} p^{|S|}\hat{f}(S)^2$, followed by the use of $p > p^{|S|}$ on $]0,1[$ with $|S| > 1$ and $1 = \sum_{S\subseteq[n]} \hat{f}(S)^2 = \sum_{\emptyset\neq S\subseteq[n]} \hat{f}(S)^2$ to get $\sum_{S\subseteq[n]} p^{|S|}\hat{f}(S)^2 \leqslant p$, where equality holds precisely if $\hat{f}(S) = 0$ for all $|S| > 1$. The latter case are the linear functions.

Technically, only the dictator functions are candidates, as they are the only linear functions that are ± 1 valued on the cube $\{-1,1\}^n$. Visual intuition for this is that we're asking for hyperplanes passing the origin that have distance 1 to all vertices of the cube.

We computed the Fourier expansion of the majority function on 3 voters, and see now that it isn't as stable as the dictator function. This may seem counter-intuitive: it's unlikely that a lot of votes were changed in the perturbation, and the majority will hardly be affected, while the perturbation of the dictator's vote will affect the decision. A way of explaining this intuitively is that we took an average over voting profiles, and not all percentages are equally likely. Indeed, to get 50%, we have $\binom{n}{n/2}$ possibilities, which is a lot more then the about $\binom{n}{n/3}$ to get 33%. Yet, the 50% case is impacted a lot more by perturbations, as perturbations of a single voter changes the majoritary decision.

COMPLETE: Arrow's theorem with the Kalai proof.
Orient cycle on a, b, c so that 1 corresponds to correct direction. NAE then means coherence of a voter, as it prohibits cycles. Dependence of vars: if x is fixed, there are 2 cases out of 6 in which x and y are equal (it fixes z) and 4/6 in which they aren't (z is free). Now, $2/6 = 1/3 = \frac{1}{2} + \frac{1}{2}\left(-\frac{1}{3}\right)$ and $4/6 = \frac{1}{2} - \frac{1}{2}\left(-\frac{1}{3}\right)$, so that we're in the stability context with $p = -\frac{1}{3}$, or at least the reformulation with $y_i = \pm x_i$ with proba $\frac{1}{2} \pm \frac{1}{2}\left(-\frac{1}{3}\right)$.