Combinatorics Tips: Part I

Matthew Brennan
brennanm@mit.edu

1 General Tips

- What is your method trying to prove and are there relevant counterexamples?
- Use good examples to figure out where you will need to use certain conditions.
- If you're stuck, try to prove the facts that you know in multiple ways. It's possible that an alternative argument is the one that generalizes.
- When greedy or direct methods fail, introduce a little structure to your approach.

2 Solve a Different Problem First

The main message of this section is that in many hard #2/5/3/6 IMO combinatorics problems there exists a much easier problem P whose statement alone would be a huge hint for the original problem. Sometimes, when you are out of fruitful ideas it can be way easier to guess what P is than to guess whatever clever ideas are in the solution to the original problem! Plus, if you introduce and solve P you will likely get partial credit (proving nontrivial things always yields points) as opposed to vague handwaving which is a quick route to zero points.

> If you are stuck, invent and solve some related problem P first. This could be the original problem with conditions removed, added, simplified, generalized or a totally different statement that captures something about the problem. Ideally P turns out to be an easy Olympiad-level problem to get nontrivial ideas rolling in a tractable way.

Some common ideas to guess a good intermediate problem P:

- Terry Tao's advice for research applies: take your problem and tamper with some conditions!
- Add in some additional assumptions to the problem
- Restrict to a certain family of instances that allows you to prove something
- The problem probably has several parts interacting in a complex way – make one of them way simpler (e.g. $n \times n$ grids to $1 \times n$ or $2 \times n$, etc.)
- Make the problem way smaller (e.g. $n = 2, 3, 4, 5$, etc.)

- Abstract away details and generalize a part of the problem

In the first example, it is useful to try to first solve a strictly weaker version of the problem. Once we have proven something at all, it will be easier to analyze our approach and identify what conceptual improvements a full solution would need.

Example 1. *(RMM 2019) Given any positive real number ε, prove that, for all but finitely many positive integers v, any graph on v vertices with at least $(1+\varepsilon)v$ edges has two distinct simple cycles of equal lengths.*

Let's start off by being less ambitious and try to get any bound of the form cv to work instead of $(1+\epsilon)v$. We begin with some preliminary simple observations:

- $(1+\epsilon)v$ could not be replaced by anything $\leq v-1$ since a tree would not have any cycles
- If the graph has $\geq n-1$ cycles, then we are done by pigeonhole since each cycle has length between 3 and n

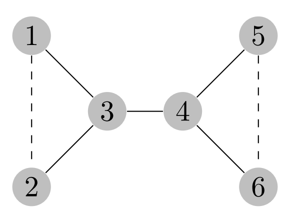
Now that we're thinking about trees we can construct a solution to the weaker problem where $(1+\epsilon)v$ is replaced by $2v-2$.

Toy Example 1. *If G is a connected graph on v vertices with at least $2v-2$ edges, then G contains two distinct simple cycles of the same length.*

Proof. Take a spanning tree T of G. For each edge e in $G\backslash T$, consider the unique cycle formed by adding e into T. This yields at least $v-1$ distinct cycles since each has a different edge not in T. By pigeonhole, we are done. □

So we need to improve 2 to $1+\epsilon$. Where was this argument loose? After writing the argument out cleanly as above, it becomes clear that we are only considering cycles using **one edge in $G\backslash T$** and ignoring a potentially huge number of cycles using more than one of these edges. Trying to handle cycles with more than one of these edges quickly becomes complicated. Let's try the simplest possible improvement: cycles with **two edges in $G\backslash T$**.

We would like to say something like many pairs of edges in $G\backslash T$ produce cycles. But there is an issue! Consider the following graph:



where the solid lines are edges of T and the dashed lines are edges of $G\backslash T$. It's possible that a pair of edges in $G\backslash T$ do not yield a cycle with T. This occurs exactly when the two cycles formed by adding the edges individually are disjoint. Given an edge $e \in G\backslash T$, let $C(e)$ denote the unique cycle in $T \cup \{e\}$. So we would like to show that there are a lot of pairs $e_1 \neq e_2 \in G\backslash T$ with $C(e_1)$ and $C(e_2)$ overlapping.

But this isn't true! With only ϵv edges in $G\backslash T$, if all of the $C(e)$ are small then they can be disjoint. But, returning to the original problem, we know that all of the $C(e)$ can't be too small otherwise some two of them would have the same size. So what does all of this **concretely** tell us?

1. We would like an argument that produces at least $(1-\epsilon)v$ many pairs $e_1 \neq e_2 \in G \backslash T$ with $C(e_1)$ and $C(e_2)$ overlapping

2. Any successful argument must use the fact that $|C(e)|$ are distinct for all $e \in G \backslash T$

What is the **simplest way** to accomplish (1)? If we pick some edge f of T and considering all $C(e)$ going through f, this gives many pairs of overlapping $C(e)$. What does (2) tell us? It tells us that the cycles can't be too small and hence some edge f is in many $C(e)$. Let $e_1, e_2, \dots, e_{\epsilon v}$ be the edges outside of T. Putting this all together, we can compute that the average number of cycles through each edge of T is

$$\frac{1}{v-1}\sum_{i=1}^{\epsilon v} |C(e_i)| \geq \frac{1}{v-1}\left(3+4+\cdots+(\epsilon v+2)\right) \geq \frac{\epsilon^2 v}{2}$$

Thus there is some edge f in at least $\epsilon^2 v/2$ cycles $C(e)$. All of the cycles formed by adding pairs of the corresponding edges e to T are distinct. This yields

$$\geq \binom{\epsilon^2 v/2}{2} > v-1$$

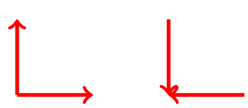
distinct cycles and we are done by pigeonhole.

What are the lessons in this problem? What we did was formally solve the easier task of solving the problem for $2v-2$ instead of $(1+\epsilon)v$. We then asked where this argument could be improved and followed a line of thought guided by (1) **simple choices** and (2) **counterexamples** to rule out approaches and make sure we were using conditions where we needed to. In general, start with an easier problem, start off with simple choices and let examples be your guide! Overall, remember that the solution will probably be nice enough to be cleanly written up in a page of the shortlist.

Here is another example where solving a different problem makes things way easier.

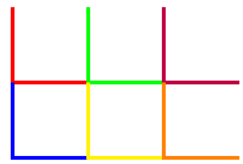
Example 2. *(RMM 2018) Ann and Bob play a game on the edges of an infinite square grid, playing in turns. Ann plays the first move. A move consists of orienting any edge that has not yet been given an orientation. Bob wins if at any point a cycle has been created. Does Bob have a winning strategy?*

The simplest and first thing to usually try in games is a pairing moves strategy: where for every move A, one player always responds with a predetermined move $f(A)$. What is the simplest pairing strategy that could be good for one of the players? Observe that the following local configurations are generally good for Ann, since they preclude a corner from being used by Bob in a cycle.



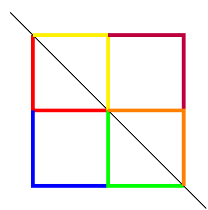
However, directly producing pairing strategy for Ana based on this does not seem to immediately work. Let's tamper with the problem statement. Right now, Ann and Bob have asymmetric goals and Ana is going first. Let's instead see what happens if Bob goes first.

Toy Example 2. *Same problem but Bob goes first.*



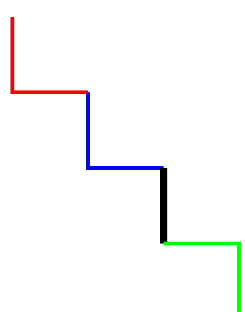
Solution. Ann wins by the following pairing strategy. For each unit square in the grid, pair its bottom and left edges. Observe that each edge of the grid is in exactly one pair. Whenever Bob plays in an edge of a pair, Ann plays in the other edge of the pair such that the edges are either both pointing towards or both away from their common vertex. This ensures no L-shaped corner can ever be in a cycle. Ann now wins since any cycle must contain an L-shaped corner. □

However, it is not hard to verify that if Ann goes first, Bob can completely break this pairing strategy. How stable is this strategy? Are there ways to modify it and still solve the Toy Example? The answer is yes. Below is another solution to the Toy Example.



Alternative Solutions to Toy Example 2. Clearly we could have also paired up the upper and right edges of each unit square. But we can also combine these two solutions! Divide the grid into two halves with a diagonal of points. Below the diagonal, pair up bottom and left edges and, above the diagonal, pair up upper and right edges. Any cycle must have a bottom-left L-shaped corner and an upper-right L-shaped corner. It is impossible for the bottom-left L-shaped corner to be above the diagonal and the upper-right L-shaped corner to be above the diagonal. □

We now need to construct a strategy where Ann can throw away the first move without consequence. The last solution to the toy example can be tweaked to make this happen. Consider a diagonal of edges and have Ann play at some arbitrary vertical edge e on the diagonal. Now have her pair bottom-left L-shaped corners on this diagonal to the left of e and upper-right L-shaped corners to the right of e. This works for the same reason as the solution above.



What are the lessons in this problem? It was helpful to (1) **tamper with the problem statement** and (2) **find alternative proofs for facts we've shown** in case we haven't find the right proofs that generalize.

Example 3. *(Russia 2005) 100 people from 25 countries, 4 from each county, are standing around a circle. Prove that one may partition them into 4 groups so that no group contains two people who were either from the same country or standing next to each other on the circle.*

What is the real problem here? It seems pretty likely that there is nothing special about 25 or 100 – these can probably be replaced by n and $4n$ without consequence. However, there might be something special going on with 4. A priori, we have no idea why there is a 4 here rather than 2, 3 or 2019. To understand more, let's make this part of the problem as simple as possible by replacing 4 with 2 and seeing what we can say.

Toy Example 3. *$2n$ people from n countries, 2 from each county, are standing around a circle. For what $(x, y) \in \{0, 1\} \times \{0, 1, 2\}$ can we always partition them into 2 groups so that every person has at most x other people from their country and at most y neighbors on the circle in their group?*

Let G be the graph on $2n$ vertices with a cycle of red edges corresponding to being neighbors on the circle and blue edges between the n pairs of people from the same country. First note that adding the right blue edge can produce an odd cycle so G is not bipartite and hence $(x, y) = (0, 0)$ is impossible. The red edges form an even cycle and hence the subgraph of G on the red edges is bipartite. This implies that $(1, 0)$ is always possible. Note that $(0, 2)$ is trivially possible by taking one from each country arbitrarily. So to fully answer the question in our toy example, it remains to determine whether $(0, 1)$ is possible.

Some small cases should have you convinced that $(0, 1)$ is always possible. Playing around with greedy strategies that try to directly solve the problem by moving people around from group to group should also have you pretty convinced that this general approach is not going to work. So what do we do?

Let's first revisit our goal: we have a 3-regular graph and want to split it into two groups so that each vertex is part of at most one edge in its group from a specific subset of the edges. This looks a lot like $(1, 0)$. For $(1, 0)$, we threw away a subset of the edges and the graph was bipartite. Is $(0, 1)$ actually different from $(0, 1)$? Not if you look at it the right way. The blue edges, the odd red edges and the even red edges are basically symmetric for the purposes of this problem. Throw away any one of those three sets and the rest of the graph is bipartite. This leads us to the following solution.

Solution to Toy Example for $(0, 1)$. Throw away every second red edge. Now each vertex is incident to one red and one blue edge. In particular, every vertex has degree 2 and thus the graph is a union of cycles. Each cycle alternates red-blue and thus is even in length. This implies that the graph is bipartite and we're done. □

Now we're happy. Not only do we understand the 2 case completely, but we have proven something nontrivial. Now let's capitalize on the fact that we've definitely started to make progress: let's try to generalize our argument to the original problem. The easiest thing to do is take exactly what we have proven and apply it directly. Split the 4 people from each country into two new countries of 2 people each. The $(0, 1)$ case above tells us that we can split the 100 people into 2 groups of 50 people so that each person is in the same group as at most 1 neighbor around the circle and no person from the same new country. Thus each of these two groups of 50 people contains exactly two people from each original country. Thus the subgraph on each of these two sets satisfies that each vertex is incident to at most one blue edge and at most one red edge. This can be verified to imply the subgraphs are each bipartite. Splitting each into its two colours yields the 4 desired groups and we're done.

For this problem, we were a little lucky and our toy problem turned out to totally solve the original problem. Often this won't be the case and we will have to look for more subtle ways to generalize the ideas in our toy problem. Notice we spent most of our time trying to solve something totally different just to get a better understanding of the setup! And it paid off. Another lesson here is that when direct greedy approaches failed, we introduced a little structure (throwing away every second red edge) and our toy example went from hard to easy. This kind of situation is common in Olympiad problems: add some structure to your goal and some natural approach emerges. For example, consider IMO 2017 #5 where once you figure out to group the people into blocks of size $2n$, the problem becomes significantly easier.

I will not go through the last example of this section (which is left as an exercise), but it too benefits immensely from considering the right easier problem. In general, gathering intuition about 3D objects can be unnecessarily difficult. Here it is helpful to instead examine 2D analogue first and see what it tells you.

Example 4. *(Russia 1997) An $n \times n \times n$ cube is divided into unit cubes. We are given a closed non-self-intersecting polygon (in space), each of whose sides joins the centers of two unit cubes sharing a common face. The faces of unit cubes which intersect the polygon are said to be distinguished. Prove that the edges of the unit cubes may be colored in two colors so that each distinguished face has an odd number of edges of each color, while each nondistinguished face has an even number of edges of each color.*

3 Induction

Induction is a trick you should always seriously try on a combinatorics problem you are stuck on. It's easy to forget or be tempted to try to solve problems in "more conceptual" or direct ways. But many problems are significantly easier with induction or intractable without it. Every problem is a good candidate for induction but some are *especially* good candidates:

1. A number of longstanding open problems in math were resolved with very long and complicated induction proofs. These proofs are somewhat mystical in nature – often elucidating very little about the reason why the result is true. So when can't figure out a good "reason" or make clean observations in a problem, it's quite possibly just induction!

2. Induction is often well-suited to complicated objects or conditions with complicated dependences that can be built up in a simpler way.

Let's begin with an example of a problem with a complicated condition that is way easier to work with inductively. This is an example of a problem which just becomes super complicated and unintuitive without induction.

Example 5. *(RMM 2017) Let n be an integer greater than 1 and let X be an n-element set. A non-empty collection of subsets $A_1, \dots, A_k$ of X is tight if the union $A_1 \cup \cdots \cup A_k$ is a proper subset of X and no element of X lies in exactly one of the A_is. Find the largest cardinality of a collection of proper non-empty subsets of X, no non-empty subcollection of which is tight.*

Let $X = \{1, 2, \dots, n\}$. The condition is such that any subfamily of the collection F must either contain all of X in its union or satisfy that there is some element in exactly one set. Call such a collection F *good*. Playing with this condition yields a few clean simple consequences, including:

$$\text{for each } i \in X, \text{ there is some element } j \text{ in exactly one set} \in F \text{ not containing } i$$

Experimenting with constructions should convince you that $2n - 2$ is the answer. An example of a construction achieving this is the set of all subsets of X of size $n - 1$ and any $n - 2$ singletons.

The main part of the problem is to now show that there is no collection of size $2n-1$. The given condition is fairly complicated and hard to extract meaning from. It might be simpler to work with inductively. What are some possible **ways to reduce the problem size** from n elements to $n-1$ elements? The simplest possible approach is:

- Remove some element i from X and all sets containing i from F

It can be verified that the remaining sets of F form a good collection. Furthermore, if i is in at most 2 sets of F, we're already done! However, this clearly fails to work for our tight example above – every element is in many sets. So we need some way to reduce the size of X even when every element is in many sets. This leads to our next inductive idea:

- Merge i and j and discard all sets of F containing exactly one of i or j

It can be verified that the remaining sets of F form a good collection. This seems way more promising. We want to show that there must be two elements i and j that appear together in all but at most two sets. But we already have something so close to this! Our observation above shows that, for any i, there is some j in exactly on set not containing i. We just want this j to also be in all but one of the sets containing i.

The last step of the problem is to resolve this issue by choosing i to be the element appearing the least in F. If i appears t times, then j must appear in at least $t - 1$ of the sets i appears in as otherwise it appears $\le t - 1$ times in total. Merging i and j and then removing the ≤ 2 sets containing exactly one of i or j completes the induction.

Altering the problem statement as in the previous section can be extremely useful to make inductive steps small enough to actually work. We next examine a classic example of this.

Example 6. *(USAMO 2007) Let S be a set containing $n^2 + n - 1$ elements, for some positive integer n. Suppose that the n-element subsets of S are partitioned into two classes. Prove that there are at least n pairwise disjoint sets in the same class.*

The weirdest things about this problem statement are definitely the quantity $n^2 + n - 1$ and the fact that n is both the size of the subsets and the number of disjoint sets we want to find. There should be no reason for this: it does not seem likely that a solution for (n, n) would not also work for the more general problem with (m, n). This yields a clear generalized problem statement. Playing around with some small examples can yield a good guess of what $n^2 + n - 1$ should be replaced with when m and n are not equal. This yields the following toy problem which can be solved by a classic induction argument.

Toy Example 4. *Let S be a set containing $m(n + 1) - 1$ elements, for some positive integers m and n. If the n-element subsets of S are partitioned into two classes, then there are m pairwise disjoint sets in the same class.*

Proof. We prove this by induction on m. If $m = 1$, then $|S| = n$ and S itself suffices. Assume the result is true for m and suppose that S satisfies $|S| = (m+1)(n+1) - 1$ and all of its n-element subsets have been labelled red or blue. Let T be any subset of S of size $m(n+1) - 1$. By the induction hypothesis, there are m pairwise disjoint subsets of T of the same color, which we assume is blue. The $n+1$ subsets of size n of the remaining $n+1$ elements must all be red, since otherwise the induction would be complete. Applying this for all such T yields that any two subsets that differ in one element must have the same color. Therefore all n-element subsets are of the same color, completing the induction. □

A cynical student might even think that the toy example was the original problem and was made artificially harder by setting $m = n$. This is a good way to think about many hard problems! There is some easier problem P underlying many hard IMO problems. If you're stuck, guessing P can be way easier than guessing the solution to the original problem out of thin air.

4 Problems

Only some of these problems are related to the ideas discussed. The main point is just to get some combinatorics practice in before the IMO! Many of these are very hard so feel free to ask for hints.

A1. An arrangement of chips in the squares of $n \times n$ table is called sparse if every 2×2 square contains at most 3 chips. Serge put chips in some squares of the table (one in a square) and obtained a sparse arrangement. He noted however that if a chip were placed in any free square then the arrangement is no longer sparse. For what n is this possible?

A2. Let $A_1, A_2, \ldots, A_{100}$ be subsets of a line, each a union of 100 pairwise disjoint closed segments. Prove that the intersection of all hundred sets is a union of at most 9901 disjoint closed segments.

A3. The total mass of 100 given weights with positive masses equals $2S$. A natural number k is called *middle* if some k of the given weights have the total mass S. Find the maximum possible number of middle numbers.

A4. There are N cities in a country. Any two of them are connected either by a road or by an airway. A tourist wants to visit every city exactly once and return to the city at which he started the trip. Prove that he can choose a starting city and make a path, changing means of transportation at most once.

A5. In a country there are 1993 towns, some pairs of which are joined by two-way roads so that it is possible to travel between any pair of towns by a sequence of roads. Prove that if at least 93 roads connect to each town, then it is always possible to travel between any pair of towns using at most 62 roads.

A6. There are 1001 towns in a country, and any two of them are joined by a one-way road. At every town exactly 500 roads start. A republic containing 668 of the towns leaves the country. Prove that it is possible to travel between any two cities in this republic without leaving the republic.

A7. Consider a convex 2000-gon, no three of whose diagonals have a common point. Each of its diagonals is colored in one of 999 colors. Prove that there exists a triangle all of whose sides lie on diagonals of the same color.

B1. 110 teams participate in a volleyball tournament. Every team has played every other team exactly once and there are no ties in volleyball. It turned out that in any set of 55 teams, there is one that lost to no more than 4 of the remaining 54 teams. Prove that in the entire tournament, there is a team that has lost to no more than 4 of the remaining 109 teams.

B2. There are some counters in some cells of 100×100 board. Call a cell nice if there are an even number of counters in adjacent cells. Can exactly one cell be nice?

B3. On a circle there are $2n+1$ points where $n \geq 2$, dividing it into equal arcs. Two players take turns erasing one point at a time. If after one player's turn, it holds that all the triangles formed by the remaining points on the circle are obtuse, then the player wins and the game ends. Who has a winning strategy: the starting player or the player who goes second?

B4. There are 999 scientists. Every 2 scientists are both interested in exactly 1 topic and for each topic there are exactly 3 scientists that are interested in that topic. Prove that it is possible to choose 250 topics such that every scientist is interested in at most 1 theme.

B5. Let G be a tournoment such that its edges are colored either red or blue. Prove that there exists a vertex of G like v with the property that, for every other vertex u there is a monochromatic directed path from v to u.

B6. You are given $\binom{n}{2}$ stones, divided into piles of various sizes. Each minute, you take one stone from each existing pile, and group them together into a new pile. Prove that eventually, you will have one pile of size i for each $1 \leq i \leq n$.

C1. Initially there are $n+1$ monomials on the blackboard: $1, x, x^2, \ldots, x^n$. Every minute, each of k students each simultaneously write the sum of some two polynomials that were written before on the blackboard. After m minutes, the polynomials $S_1 = 1+x, S_2 = 1+x+x^2, S_3 = 1+x+x^2+x^3, \ldots, S_n = 1+x+x^2+\cdots+x^n$ and some others are on the blackboard. Prove that $m \geq \frac{2n}{k+1}$.

C2. There are some markets in a city. Some of them are joined by one-way streets such that there are exactly two streets to leaving any market. Prove that the markets in the city may be partitioned into 1014 districts such that streets only join markets from different distrincts and all of the streets between two districts point the same way.

C3. There are some cards in a deck. For each pair of cards, there exists a winner. However, this winning relation is not transitive. The deck is divided into two piles, the contents and orders of the cards are known to you. Each turn, one of the two cards at the top of the piles wins; you must move the two top cards to the bottom of the winning deck, but you may decide the order. Prove that you can collect all cards in a single pile.

C4. Initially, there are red and blue horses on the lower left and right corners of a 2018×2018 board and owned by A and B, respectively. The players A and B alternate, with A starting. In a turn, a player of moves his or her horse 20 cells with respect to one coordinate and 17 cells with respect to the other while ensuring that no horse occupies a square previously occupied in the game. The player who can't make the move loss, who has the winning strategy?

C5. Let G be a directed graph with n vertices such that for any two vertices, there is some vertex that they both point to. Suppose that k and m are such that $2^{2^k+1} - 1 > n$ and $1 < m < 2k + 1$. Prove that G contains a directed cycle of length exactly m.