MATH 108 NOTES: COMBINATORICS

ARUN DEBRAY

JUNE 11, 2013

These notes were taken in Stanford's Math 108 class in Spring 2013, taught by Professor Kannan Soundararajan. I live-TEXed them using `vim`, and as such there may be typos; please send questions, comments, complaints, and corrections to `a.debray@math.utexas.edu`.

Contents

1. Introduction to Graph Theory: 4/1/13

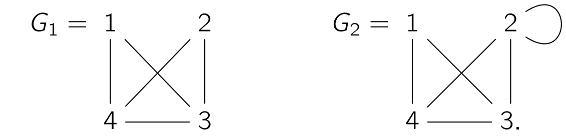
This course will cover graph theory and enumerative combinatorics. Graph theory will be covered in the first 9 or 10 chapters of the book (and perhaps some supplementary material) and enumerative combinatorics will be in chapters 14 and 15, as well as the first chapter of Stanley's book. If time permits, other topics will be discussed, such as coding theory or Hadamard matrices.

Definition. A graph is an object consisting of vertices V and a set of edges E such that each edge e is a pair of vertices $\{x, y\}$, such that edge e is thought to connect edges x and y.

Alternatively, one could think of sets V and E and a function $\phi : E \to V \times V$.

One says that x and y are on the edge e, or that they are incident on it. Note that $x = y$ is allowed.

Example 1.1. If $G_1 = (V, E_1)$ with $V = \{1, 2, 3, 4\}$ and $E_1 = \{\{1, 3\}, \{3, 4\}, \{2, 4\}, \{1, 4\}\}$ and $G_2 = (V, E_2)$ with $E_2 = E_1 \cup \{2, 2\}$, the graphs are



Definition. A simple graph is a graph that contains no loops and no multiple edges, such as G_1 in Example 1.1 above, but not G_2.

Definition. A finite graph $G = (E, V)$ is a graph in which both E and V are finite sets.

Definition. If $v_i, v_j \in V$ are joined by an edge, they are called adjacent. Then, the adjacency matrix of a graph G is a matrix A of size $|V|$ such that a_{ij} is equal to the number of edges joining v_i and v_j.

If the graph is undirected, then the matrix is symmetric.

In a simple graph, the adjacency matrix consists of only zeroes and ones. A symmetric matrix with real entries has nice properties, so the subject of spectral graph theory attempts to understand a graph by performing linear algebra on its adjacency matrix.

Definition. The degree of a vertex is the number of edges coming out of it. Note that for loops on an undirected graph, each loops counts twice. The degree of a vertex is also referred to as its valency.

Thus, $\deg(v_i) = \sum_{j=1}^{n} a_{ij}$ (summing a column of the adjacency matrix), which does require the convention given above.

Definition. A graph is called regular if all vertices have the same degree.

If A is the adjacency matrix of a k-regular graph, then k is an eigenvalue, because $A\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix} = \begin{pmatrix}k\\ \vdots\\ k\end{pmatrix}$. In particular, since the matrix is symmetric for an undirected graph, the eigenvalues are all real.

Understanding the adjacency matrix is extremely helpful for applications such as random walks on graphs.

Lemma 1.1 (Handshaking[1]). *In any graph, the number of vertices of odd degree is even.*

Proof. It can be assumed that there are no loops, since removing the loops of a graph doesn't change the parity of the degree of any vertex.

Consider every ordered pair (x, y) which forms an edge $\{x, y\}$, so that (x, y) is counted if (y, x) is. Thus, there are $2|E|$ such pairs.[2] However, this is also equal to the sum

$$\sum_{x \in V} \left(\sum_{\{x,y\} \in E} 1 \right) = \sum_{x \in V} \deg(x),$$

which is also equal to the sum of the elements in the adjacency matrix.

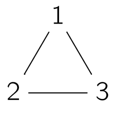
Thus, $\sum_{x \in V} \deg(x) = 2|E|$. Notice that throwing out the loops isn't necessary to get this result. Thus, the number of vertices with odd degree must be even; otherwise, the sum would be odd. ⊠

The lemma owes its name to a scenario where a bunch of people at a party exchange some handshakes, and then count the sum of the number of handshakes that each person made.

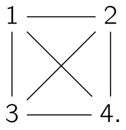
Some more examples of graphs:

(1) The complete graph on n vertices, denoted K_n, is a simple graph on vertices $V = \{1, \dots, n\}$ with $E = \{\{i, j\}, i, j \in V\}$ (i.e. all vertices are connected). For example,

$K_3 =$



$K_4 =$



K_n has $\binom{n}{2} = n(n-1)/2$ edges. Thus, if G is any finite simple graph with $|V| = n$, then $|E| \leq \binom{n}{2}$, since there are no more edges to add to K_n.

(2) A bipartite graph is a graph such that $V = A \cup B$, with A and B disjoint, such that the edges join vertices in A to vertices in B, and there are no edges between two elements of A or two elements of B.

The complete bipartite graph $K_{m,n}$, where $|A| = m$ and $|B| = n$, has $|V| = m + n$ and $|E| = mn$.

(3) Using probability, one can construct some interesting classes of graphs, such as the Erdös-Renye random graph: if $|V| = n$ and there is some probability $p \in [0, 1]$, then for each $v_i, v_j \in V$ (considering v_j, v_i and v_i, v_j in the same trial), add an edge between them with probability p. This leads to questions such as how this sort of graph might look. The expected number of edges is $|E| = p\binom{n}{2}$, since there are $\binom{n}{2}$ trials.

One could examine triangles in a graph: when does a graph G contain vertices $1, 2, 3$ and edges $\{1, 2\}, \{2, 3\}, \{1, 3\}$? Graphs without triangles certainly exist, such as bipartite graphs.

[1] This is of course not to be confused with the Hand*waving* Lemma.

[2] This is a common trick in combinatorics — double-counting something, or counting in two different ways, can show that two things must be equal.

Example 1.2. The complete bipartite graph $K_{m,n}$ has n^2 edges and $2n$ vertices, so it has half of the "maximum" number of edges. Nonetheless, it doesn't looks like a random graph (with $p = .5$) might look at all, since triangles would be expected for large n. More precisely, the number of possible triangles is $\binom{|V|}{3}$, so the probability is p^3 over $\binom{|V|}{3}$ trials. Thus, it is expected to happen fairly often. This can be generalized to an important topic of research: what kinds of structures tend to exist in large, seemingly random graphs?

The idea of investigating graphs probabilistically is called the probabilistic method, due to Erdös. Szemeredi also did a lot in this area (and won the Abel prize in 2012 for his work).

Exercise 1.1. If G is a k-regular bipartite graph (which already implies that there aren't any loops), then show that $-k$ is an eigenvalue of the adjacency matrix. (Hint: try to guess an eigenvector, such as indexing 1 if $x_i \in A$ and -1 if $x_i \in B$.)

Definition. A walk is an alternating sequence of vertices and edges $v_1, e_1, v_2, e_2, \ldots, e_{n-1}, v_n$, such that edge e_j connects v_k and v_{k+1}. This is exactly what you would think it is.

Notice that a walk can visit an edge or vertex many times.

Definition. A walk is called a trail if no edge is repeated in it.[3]

Example 1.3. $v_1, e_1, v_2, e_2, v_3, e_3, v_4, e_4, v_5, e_5, v_6, e_5, v_3, e_2, v_2, e_1, v_1$ is a walk that isn't a trail.

Definition. A closed walk is (again) exactly what you might think it to be: a walk in which the starting and ending vertices are the same. One can similarly define a closed trail.

Definition. A path is a walk in which all vertices are distinct.[4]

Definition. An Euler trail is a trail through a graph that traverses every edge exactly once.

Not every graph has an Euler trail: look at K_4.

Theorem 1.2 (Euler). *A connected graph has an Euler trail iff there are at most two vertices with odd degree.*

Connectedness will be defined in the next lecture, and its definition will not come as a surprise. The proof will also be given in the next lecture. Note also that by Lemma 1.1, the condition of at most 2 is equivalent to specifying either 0 or 2 such edges.

2. More Graph Theory: 4/3/13

Definition. A graph is connected if for any two vertices in the graph, there is a walk (equiv. path) from one to the other.

Definition. Similarly to how one defines a closed walk and a closed trail, one has a closed path. This requires the first and last vertices to be identical (which strictly implies that it isn't a path). This is also a cycle.

Last lecture, we also saw Euler trails. A closed Euler trail is also called an Euler tour. Euler was interested in a question about the bridges of Königsberg: specifically, if the graph created from the bridges and islands admitted an Euler tour. By Theorem 1.2, this isn't possible. So let's prove the theorem.

Proof of Theorem 1.2. Let G be a connected graph. If it has an Euler tour, then every edge going into a vertex must come out of it, so the degree of every vertex is even. If it only has an Euler trail, then the starting and ending point don't necessarily meet this, so they might have odd degree, but that makes for at most 2 such edges.

If G has only edges with even degree, start at an arbitrary point and build a trail by choosing distinct edges. Then, the trail won't get stuck anywhere, but it might not get all of the edges. Thus, remove the edges contained in this trail, leaving a smaller graph. This graph isn't necessarily connected, but it has several connected components, each of which has vertices of only even degree. By induction, we can assume that each of these components has an Euler tour, since they all have fewer edges than G.[5] Then, an Euler tour can be made by joining the edges in the original trail with each of these tours at some vertex.

If G has instead two vertices of odd degree, take the graph G' to be G with a new edge e between them. Then, G' has an Eulerian tour by above, so taking that tour and removing e gives an Eulerian trail on G. ⊠

Definition. A tree is a connected graph that has no cycles.

[3] Note that the textbook calls this sort of walk a *path*.

[4] The textbook calls this a simple path.

[5] This would need to be formalized for the graph with one edge and then making an inductive assumption, but it's just as valid.

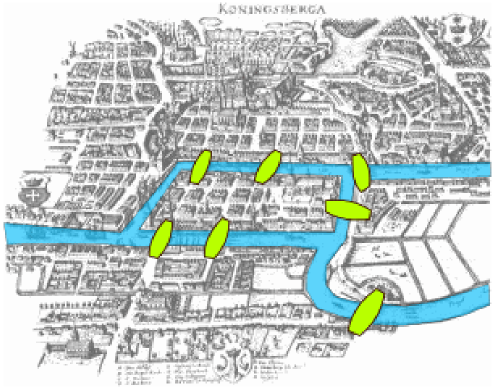


Figure 1. Schematic of the Bridges of Königsberg. Source

These are called trees because many of them can be made to look like trees with trunks and branches.

A tree with n vertices must have $n-1$ edges, which can be proven straightforwardly by induction. Additionally, any two vertices are connected by a unique path: a tree is connected, so such a path exists, but if there were two distinct paths p_1 and p_2, one could obtain a cycle from them, by taking $p_1 \cup p_2 \setminus (p_1 \cap p_2)$.

If one adds an edge to a tree without adding any vertices, then it connects two vertices v_1 and v_2. However, there was already a path between these vertices and this edge gives another path, so the result is not a tree. If one removes a path from a tree, the result is disconnected. Thus, if one removes an edge, it splits into two trees, so the formula for the number of edges (i.e. $|V|-1$) follows.

For any connected graph, define the distance between two edges x and y to be $d(x,y)$, the smallest number of edges in any path between x and y. This can be used to show that trees are bipartite: take some vertex v and let A be the set of vertices of even distance from v, and B be the set of vertices of odd distance from v. Then, having edges from A to be is possible, but there will never be an edge from A to A, or B to B, since this would cause a nontrivial cycle (since it's of odd length).

This is a special case of a more general theorem:

Theorem 2.1. *A graph that has no cycles of odd length is bipartite.*

This can be proven by a modification of the above argument, as well as reasoning that a bipartite graph cannot have cycles of odd length by trying to sort the vertices in the cycle into sets.

Definition. An isomorphism of graphs $G_21 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ called $f : G_1 \to G_2$ is a bijection $V_1 \to V_2$ that preserves edges.

For example, one can use an isomorphism to make K_4 planar, by moving one of the edges around the graph. There's a website called `planarity.net`, which allows one to try and make a graph planar.

Definition. An automorphism of a graph is an isomorphism with itself.

One can try to enumerate all trees of a given size. For example, the trees with 3 edges aren't very interesting; they're just lines of two edges. Thus, one might consider labelled trees, where the vertices are chosen among $\{1, \dots, n\}$ and the order of the labelling matters. There aren't 6 labelled trees with 3 vertices, though, because $1-2-3$ is isomorphic to $3-2-1$. Thus, there are 3 labelled trees with 3 vertices.

Theorem 2.2 (Cayley)**.** *There are n^{n-2} trees with n labelled vertices.*

The proof of this theorem will require the following lemma:

Lemma 2.3. *Every tree has at least two vertices of degree 1.*

Proof. Let d_i be the degree of vertex i; then, $\sum_{i=1}^n d_i = 2|E| = 2n-2$. Since $d_i \geq 1$, then at most $n-2$ have degree at least 2, or else one vertex would have to have degree 0. ⊠

3. Cayley's Theorem: 4/8/13

Returning to the problem of labelled trees, which is a different condition than isomorphism, we will prove Theorem 2.2.

Definition. A subgraph H of a graph G is a subset of the vertices and edges of the graph: $H = (V(H), E(H))$, such that the edges in $E(H)$ only connect the vertices in $V(H)$, with $E(H) \subseteq E(G)$ and $V(H) \subseteq V(G)$.

This is exactly what you would think it would be.

Definition. A subgraph of G is called spanning if it contains all of the vertices of G.

The most interesting spanning subgraphs are those whose edge sets are a proper subset of the edge set of G. Then, Cayley's Theorem can be reformulated in terms of the number of spanning trees contained in K_n.

First, it is useful to observe that every connected graph has a spanning tree. There are many ways to do this; for example, one could order the edges somehow and add the edges in order, adding only the edges that don't produce a cycle. Alternatively, you could add vertices in one per step; since the graph is connected, there is always another vertex to add (until all of them have been used).

Claim. If a connected graph has n vertices and $n - 1$ edges, then it is a tree.

Proof. Using Lemma 1.1, there is a vertex with degree 1. Then, remove it, and apply induction. ⊠

There are plenty of other ways to prove this. Now, it will be possible to prove Cayley's Theorem. Two proofs will be given:

Proof of Theorem 2.2.

Definition. A Prüfer code is a sequence of $n - 2$ numbers $[y_1, \dots, y_{n-2}]$ with $1 \leq y_1, \dots, y_{n-2} \leq n$.

Thus, there are n^{n-2} distinct Prüfer codes, and the goal is to show that there is a bijective correspondence between labeled trees and Prüfer codes. Generate a Prüfer code from a tree in the following way: of all of the vertices with degree 1, let x_1 be the vertex with the smallest value. Then, let y_1 be the unique vertex to which x_1 is connected. Then, remove the edge connecting them, and repeat; the next entry in the Prüfer code is the next y_i found by this algorithm. Then, continue until $n - 2$ numbers are found.

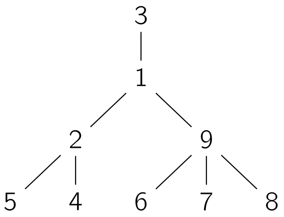


Figure 2. This tree's Prüfer code is $[1, 2, 2, 1, 9, 9, 9]$.

Claim. If $[y_1, \dots, y_{n-2}]$ is the Prüfer code for a tree, then $y_{n-1} = n$.

Proof. A vertex will be removed from the game if it is the least vertex in the tree, which will never be true for vertex n, and at the last step, there are only two vertices, n and a for some $a < n$, which are connected, so y_{n-1} is whatever is connected to a, which is n. ⊠

This is why the Prüfer code is written as an $(n - 2)$-tuple, rather than an $(n - 1)$-tuple; if it were of size $(n - 1)$, the last entry wouldn't have any meaning.

Now, given a Prüfer code, it is possible to obtain a tree. First observe that each vertex v of the tree will appear in its Prüfer code exactly $\deg(v) - 1$ times, since it is removed once (in which case it isn't added to the Prüfer code), but the remaining $\deg(v) - 1$ times the other vertex on the considered edge is removed, and v is added to the code. Thus, given a Prüfer code, such as $[1, 2, 2, 1, 9, 9, 9]$ as in Figure 2, one can reconstruct the tree according to the following algorithm:

- First, place vertex y_1 on the graph and add the smallest vertex with degree 0, connecting them.
- Remove y_1 from the Prüfer code and repeat, updating the degrees left to add based on the new Prüfer code. Then, repeat these steps.

Though this seems a little fuzzy, this is a bijective correspondence between a tree and its Prüfer code, since each one can be used to obtain the other, and a more rigorous proof of this can be given by induction.

Then, there are clearly n^{n-2} Prüfer codes, so there are n^{n-2} labelled trees. ⊠

The binomial coefficient is the number of ways to choose k elements out of an n-element set, $\binom{n}{k} = n!/(k!(n-k)!)$. This obeys the recurrence $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$, which can be seen in Pascal's triangle: pick a set S of k elements and a favorite element t; either $t \in S$, in which case you get the first term, or $t \notin S$, which gives the second term. Additionally, there is the following theorem:

Theorem 3.1 (Binomial[6]).

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

There are several ways to prove it: induction on n is straightforward, but one can also directly observe that when multiplying out $(x+y)$ and looking for a particular coefficient, one chooses k xs and $n-k$ ys.

The binomial coefficient can be generalized to the multinomial coefficient:

$$\left(\sum_{i=1}^{\ell} x_i\right)^n = \sum_{k_1+\cdots+k_\ell=n} \binom{n}{k_1,\dots,k_\ell} \prod_{i=1}^{\ell} x_i^{k_i}. \tag{1}$$

Here, $\binom{n}{k_1,\dots,k_\ell} = n!/(k_1!\cdots k_\ell!)$ is called the multinomial coefficient. This can be thought of as assigning n people to ℓ teams of sizes $k_1,\dots,k_\ell$, so the recursive formulation is

$$\binom{n}{k_1,\dots,k_\ell} = \sum_{j=1}^{\ell} \binom{n-1}{k_1,\dots,k_{j-1},k_j-1,k_{j+1},\dots,k_\ell}. \tag{2}$$

A complete definition does also require that $\binom{n}{k_1,\dots,k_\ell} = 0$ if some $k_j < 0$.

The multinomial coefficient reduces the the binomial coefficient of n with k and $n-k$.

Then, (1) will be used in a second proof of Cayley's Theorem to show that

$$(1+\cdots+1)^{n-2} = \sum_{\substack{k_1,\dots,k_n \geq 0 \\ k_1+\cdots+k_n=(n-2)}} \binom{n-2}{k_1,\dots,k_n}. \tag{3}$$

4. Another Proof of Cayley's Theorem: 4/10/13

Another proof of Cayley's Theorem can be given by enumerating all trees with given properties. This will be an example of enumerative combinatorics.

Second proof of Theorem 2.2. Let $t(n; d_1,\dots,d_n)$ be the number of trees on n vertices such that vertex i has degree d_i. Then, all of the $d_i \geq 1$ and $\sum_{i=1}^{n} d_i = 2n-2$. Thus, $\sum_{i=1}^{n}(d_i - 1) = n-2$, since n terms have been taken away.

Claim.

$$t(n; d_1,\dots,d_n) = \binom{n-2}{d_1-1,\dots,d_n-1}.$$

Supposing this claim, the total number of trees is

$$\begin{aligned}\sum_{d_1,\dots,d_n} t(n; d_1,\dots,d_n) &= \sum_{d_1,\dots,d_n} \binom{n-2}{d_1-1,\dots,d_n-1} \\ &= \sum_{\substack{k_1,\dots,k_n \geq 0 \\ \sum k_i = n-2}} \binom{n-2}{k_1,\dots,k_n} \\ &= (\underbrace{1+\cdots+1}_{n \text{ times}})^{n-2} = n^{n-2} \qquad \text{by (3).}\end{aligned}$$

Proof of the claim. It's probably worth checking this claim for $n=2$ or $n=3$ to build intuition, since these cases are pretty simple. The general case will be given by a recursive formula that is similar to the one given for multinomial coefficients in (2).

Of the n vertices, some vertex has degree 1: $d_i = 1$. Then, vertex i can be connected to any of the vertices $1,\dots,i-1,i+1,\dots,n$. If it's connected to vertex 1, the number of options is

$$t(n-1; d_1-1, d_2, d_3, \dots, d_{i-1}, d_{i+1}, \dots, d_n) = \binom{n-3}{d_1-2, d_2-1, \dots, d_{i-1}-1, d_{i+1}-1, \dots, d_n-1}$$

[6] "About the Binomial Theorem I am teeming with a lot o' news. . ."

by the inductive assumption,[7] so one can take the sum over all such vertices

$$\begin{aligned} t(n; d_1, \dots, d_n) &= \sum_{i=1}^{n} \binom{n-3}{d_1 - 2, d_2 - 1, \dots, d_{i-1} - 1, d_{i+1} - 1, \dots, d_n - 1} \\ &= \binom{n-2}{d_1 - 1, \dots, d_n - 1} \end{aligned}$$

by the multinomial theorem. ☒

Then, the theorem follows as above. ☒

Here are two applications of this:

(1) Take a connected graph and make it a weighted graph, or a graph in which every edge e has a weight $w(e) \geq 0$. These can be thought of as costs of traversal, or distances between the points, or travel times, etc. Then, if one starts at some vertex A, what is the shortest path to some other vertex? Here, "shortest" means the most economical: if a path uses edges $e_1, \dots, e_k$, then its weight is $w(e_1) + \cdots + w(e_k)$, and the goal is to find the path that minimizes the sum of the weights of the edges. There is an algorithm for this, which produces a spanning tree for G such that the path on this tree connecting A to any other point on the tree is the shortest.

   Algorithm (Dijkstra).
   - Start at A, or letting $S = \{A\}$.
   - At every stage, there is a set S of vertices such that the shortest path from A to these vertices is known. Here, find a $v \in V \setminus S$ such that the distance from A to some vertex $u \in S$ plus the distance from u to v is minimized, over all $u \in S$ adjacent to v, over all $v \in V \setminus S$.

   Since G is connected, there will always be such v and u, so this algorithm is correct. Of course, there's a bit more to think about, but this is not so bad.

   If the weights aren't assumed to be nonnegative, there is no unique shortest path, so the problem wouldn't be well-formed, and the problem wouldn't have as much physical meaning.

(2) Given some weighted, connected graph G, one problem is to find a minimal spanning tree for G: to find a spanning tree T such that $\sum_{e \in T} w(e)$ is minimum; such a tree is called a minimal spanning tree. This has applications such as laying pipe efficiently in order to cover every city in an area.

   In the previous lecture, it was shown that every connected graph has a spanning tree, so there must be a minimal spanning tree as well. Kruskal's algorithm is the standard solution here, but there are others, such as reverse-delete. They are all of the same family, referred to as greedy algorithms, and do the following: at each step, find the cheapest edge that hasn't been used and doesn't form a cycle, and add it to the tree, or from G remove the most expensive edge that doesn't disconnect G.[8]

Another solution is Prim's Algorithm. This builds up a sequence of vertices S such that for each step, one chooses a vertex from $V \setminus S$, and connected it in the cheapest way to a vertex in S. This differs from Kruskal's algorithm in that it requires the tree to always be connected as it is built, whereas Kruskal's doesn't.

Claim. Prim's Algorithm produces an MST.[9]

5. Ramsey Theory: 4/15/13

Ramsey theory can be summarized as: in any sufficiently large system, total disorder is not possible. For example, any sufficiently large graph will have some sort of pattern in it.

Color the complete graph K_n with two colors, red and blue. Is it always possible to find a completely blue K_p or a red K_q for some given p, q?

Theorem 5.1 (Ramsey). *If n is sufficiently large (in terms of p and q), then there always exists either a completely red K_p or a completely blue K_q.*

[7] Alternatively, $t(n; d_1, \dots, d_n)$ can be thought of as a symmetric function of the degree. Thus, it can be assumed they are in order: $d_1 \geq d_2 \geq \cdots \geq d_n$, so $d_n = 1$. Then, repeat the argument as above: vertex n is connected to some other vertex, so

$$t(n, d_1, \dots, d_n) = \sum_{i=1}^{n-1} t(n-1; d_1, \dots, d_{i-1}, d_i - 1, d_{i+1}, \dots, d_n),$$

since removing that vertex and edge leaves behind a tree with the given degrees of its vertices.

[8] This seems like it would be more expensive in terms of running time, but this is a math class, so nobody cares.

[9] There could be many such MSTs, though if the weights are distinct, then there is exactly one.

Example 5.1. Take $p = 2$. This asks if there is a blue edge in the graph. If there is, that's okay, but if not, then the entire graph is red, and all edges are red, so a red K_q can be found if $n \geq q$.

Let $R(p, q)$ be the smallest value of n such that the theorem is true. Then, it was just shown that $R(2, q) = q$.

Example 5.2. $p = 3$ and $q = 3$ is the basic example. The question thus becomes about finding a red triangle or a blue triangle?

Claim. $R(3, 3) = 6$.

Proof. First, it will be necessary to show that five vertices aren't enough. Take the complete graph K_5 and color some pentagon red and every other edge blue, as in Figure 3. Additionally, six vertices are enough, which will be shown by the pidgeonhole principle (like most things in this theory). Choose some vertex: at least three of the vertices coming out of it are the same color (without loss of generality, red). Then, look at the edges between the vertices those edges connect to: if any of these edges is red, there is a red triangle. But if none of them are, then they form a blue triangle. ☒

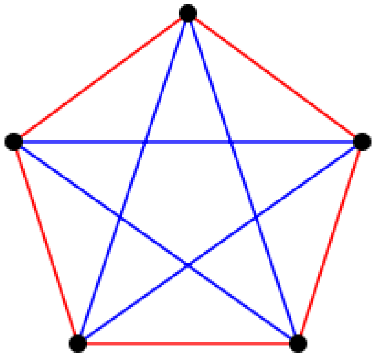


Figure 3. A coloring of K_5 which contains no triangles. Source.

Another easy argument is that $R(p, q) = R(q, p)$ for reasons of symmetry. Now, the general theorem can be tackled:

Proof of Theorem 5.1. Proceed by induction on $p + q$, and bound the number of vertices that are needed. Assume the theorem is false for $n = p + q$.

Pick some vertex v, which is connected to $n - 1$ vertices. Of these edges, suppose b of them are blue and r are red. The b blue edges connect to b vertices forming a complete graph K_b. This can't contain any blue K_{p-1} or red K_q, since these would prove the theorem. Similarly, looking at the r vertices connected to v by red edges, they cannot contain a red K_{q-1} or a blue K_p for the same reasons. Thus, $b \leq R(p-1, q) - 1$ and $r \leq R(p, q-1) - 1$. Thus, $n = b + r + 1 \leq R(p-1, q) + R(p, q-1) - 1$. In other words, this is an upper bound for $R(p, q)$: $R(p, q) \leq R(p-1, q) + R(p, q-1)$. ☒

This proof looks locally at a graph, since nobody knows how to look globally at a graph to make these sorts of proofs. Notice that this recurrence looks just like the one from the Binomial Theorem, and thus:

Corollary 5.2. $R(p, q) \leq \binom{p+q-2}{p-1}$.

Proof. $R(2, q) = \binom{q}{1} = q$ and $R(3, 3) \leq R(2, 3) + R(3, 2) = 6$, so by induction,

$$R(p, q) \leq R(p-1, q) + R(p, q-1) \leq \binom{p+q-3}{p-2} + \binom{p+q-3}{p-1} = \binom{p+q-2}{p-1}.$$ ☒

Beyond these bounds, very little is known about $R(p, q)$ for general p and q:

- $R(3, k)$ is known exactly for $k \leq 9$. For example, $R(3, 9) = 36$.
- $R(4, 4) = 18$, and $R(4, 5) = 25$.

That's it, though lower and upper bounds for other values are known.

> "Imagine an alien force, vastly more powerful than us, landing on Earth and demanding the value of $R(5, 5)$ or they will destroy our planet. In that case, we should marshal all our computers and all our mathematicians and attempt to find the value. But suppose, instead, that they ask for $R(6, 6)$. In that case, we should attempt to destroy the aliens." –Paul Erdös

These are both finite problems, just incredible large ones. But a brute-force calculation of all two-colored graphs of K_{100} would require $2^{\binom{100}{2}}$ calculations, which is impossible.

Looking at the diagonal Ramsey numbers ($p = q$), the upper bound can be given as $R(p, p) \leq \binom{2p-2}{p-1} \leq 2^{p-2}$.[10]

For lower bounds, there is a probabilistic method due to Erdös, which manages to provide a proof without an example. In this finite case, though, probability is just counting. The goal is to find the largest possible n for which there exists some coloring with no red K_p or blue K_p. Do this by summing over all possible 2-colorings of edges of K_n:

$$\sum_{\text{2-colorngs of } K_n} (\#\ \text{blue } K_p + \#\ \text{red } K_p). \tag{4}$$

The goal is to show this is strictly less than the number of possible two-colorings of K_n, which is $2^{\binom{n}{2}}$, since this would imply there is some coloring for which there are no such monochromatic subgraphs K_p.

This bound can be found by double-counting, calculating the sum in two ways to obtain two different kinds of information. For a given red or blue K_p, count the number of 2-colorings of K_n in which it is monochromatic. Choose p vertices from the n, which can be done in $\binom{n}{p}$ ways, and pick either red or blue. Then, there are $2^{\binom{n}{2}-\binom{p}{2}}$ ways to color the rest of the graph, so the sum (4) is equal to $2\binom{2}{p}2^{\binom{n}{2}-\binom{p}{2}}$.

Thus, we want $2\binom{n}{p}2^{\binom{n}{2}-\binom{p}{2}} < 2^{\binom{n}{2}}$. Rearranging, this is $\binom{n}{p} < 2^{\binom{p}{2}-1}$. Since $\binom{n}{p} = \frac{n(n-1)\cdots(n-p+1)}{p!} \leq n^p/p!$, it is sufficient to check if $n^p/p! \leq 2^{\binom{p}{2}-1}$, or that $n < \left(2^{p(p-1)/2-1}p!\right)^{\frac{1}{p}}$. The factorial term dominates, so it's sufficient to have $n < \left(2^{p^2/2+p/2-2}\right)^{1/p}$. It's possible to do better with more care, again, but it's enough to have $n < 2^{p/2}$. Thus, we have the following theorem:

Theorem 5.3. *There exists a coloring of $K_{\lfloor 2^{p/2} \rfloor}$ with no monochromatic K_p. In other words, $R(p, p) \geq 2^{p/2}$.*

Thus, the bounds are $(\sqrt{2})^p \leq R(p, p) \leq 4^p$, so is $R(p, p) \approx c^p$ for some c? Maybe. Nobody knows. A big open problem is to improve these bounds; even changing the 4 to a 3.999 would be a major improvement.

Using Stirling's Formula, which approximates $n! \approx \sqrt{2\pi n}(n/e)^n$, the actual bound for large numbers is something like $R(p, p) \geq p2^{p/2}/(e\sqrt{2})$. If you're willing to work a lot harder, there's a better result (Spencer 1980) that gives a factor of two, which is the best known: $R(p, p) \geq p\sqrt{2}2^{p/2}/e$.

The best known upper bound is due to Conlon in 2011, and shows that $R(p, p) \leq 4^p/p^A$ for any A for sufficiently large p. Nonetheless, this is still nowhere near 3.999^p.

6. More Ramsey Theory: 4/17/13

Recall that Ramsey's Theorem stated that if one chooses p and q, then there exists a complete graph on n vertices for some n such that if every vertex of K_n is colored either red or blue, then there exists either an entirely red K_p or entirely blue K_q as a subgraph of K_n. The minimal such n is called $R(p, q)$; for example, $R(2, q) = q$ and $R(3, 3) = 6$. We showed an upper bound $R(p, q) \leq \binom{p+1-2}{p-1}$, and a lower bound was obtained by the probabilistic method: $R(p, p) \geq 2^{p/2}$.

There are plenty of related questions and generalizations of Ramsey's Theorem:

Theorem 6.1. *For given $p_1, \dots, p_r$, then if n is sufficiently large, then every r-coloring of K_n will contain some K_{p_i} that is monochromatic of color i.*

The smallest n for which this is true can be denoted $R(p_1, \dots, p_i)$.

Proof of Theorem 6.1. The proof will not be much different than the proof of Theorem 5.1: suppose K_n is r-colored but there is no K_{p_i} of color i, where the proof uses induction on n.

Pick some vertex v. Then, there are strictly less than $R(p_1 - 1, p_2 \dots, p_r)$ vertices connected to v such that the connecting edge has color 1, since that would satisfy the theorem (since if there were $p - 1$ of them of color 1, then there would be a monochromatic K_{p_1} created by adjoining v). Similarly, the number of edges of color i must be $R(p_1, \dots, p_{i-1}, p_i - 1, p_{i+1}, \dots, p_r)$. Thus, summing up,

$$n \leq 1 - r + \sum_{i=1}^{r}(p_1, \dots, p_{i-1}, p_i - 1, p_{i+1}, \dots, p_r),$$

so adding one more vertex is sufficient to prove the theorem. ⊠

[10] With a little more calculation, a better bound can be obtained.

Thus, there is an upper bound on this generalized Ramsey number:

$$R(p_1, \dots, p_n) \leq (2 - r) + \sum_{i=1}^{r} (p_1, \dots, p_{i-1}, p_i - 1, p_{i+1}, \dots, p_r).$$

Here, $2 - r$ is nonpositive, so it can be ignored if one wishes.

Exercise 6.1. Prove a generalization of Corollary 5.2: that

$$R(p_1, \dots, p_n) \leq \binom{(p_1 - 1) + \cdots + (p_r - 1)}{p_1 - 1, \dots, p_r - 1}.$$

Consider $R(3, 3, 3)$: if one takes a vertex v, then there must be at most 5 vertices connected to v by an edge of color 1, and similarly for colors 2 and 3, or there would be a triangle because $R(3, 3) = 6$. Thus, $R(3, 3, 3) \leq 17$, and it's known that $R(3, 3, 3) = 17$ as well.

Philosophically, in a sufficiently large set of data, there will be small patterns. This is why there are constellations that look like something in the stars, or something which might look like a secret message if one takes some sequence of letters in a book.

In the last thirty years, there has been a lot of activity in the field of Ramsey theory on natural numbers. Here, the goal is to color $\mathbb{N}$ with r colors.

Theorem 6.2 (Schur)**.** *Given r, if N is sufficiently large and $[1, N]$ is colored with r colors, then there is a monochromatic solution to $x + y = z$.*

Proof. Color K_N and color the edge (i, j) between vertices i and j with the color of $|i - j|$ on $[1, N]$. Then, there is a monochromatic triangle if N is sufficiently large. Label its vertices a, b, and c in descending order. Thus, $a - b$, $b - c$, and $a - c$ are monochromatic, so the equation $a - c = (a - b) + (b - c)$ is monochromatic. Let $z = a - c$, $x = a - b$, and $y = b - c$ to explicitly give the equation. ☒

There are a lot of complicated versions of this result. One of the more beautiful examples:

Theorem 6.3 (Van der Waerden)**.** *In any r-coloring of the integers, there exist arbitrarly long monochromatic arithmetic progressions (i.e. sequences of the form $a, a + d, a + 2d, \dots, a + (k + 1)d$).*

There is a finite version of this theorem, as with Schur's theorem: there is some number $N(r, k)$ such that an arithmetic progression of length k always exists if $[1, N(r, k)]$ is r-colored.

Theorem 6.4 (Gowers)**.** *An upper bound for $N(r, k)$ is*

$$N(n, k) \leq 2^{2^{r^{2^{2^{k+9}}}}}.$$

Gowers got a Fields medal in part because of this theorem. It leads to lots of useful things, such as the theorem that the primes contain arbitrarily long arithmetic progressions.

Another example involves playing Tic-Tac-Toe in high dimensions. Suppose one has a d-dimensional board of size $k \times k \times \cdots \times k$. Denote the set $\{1, \dots, k\}^d$ as k-tuples of coordinates $(k_1, \dots, k_d)$. Then, define a *combinatorial line* to be a function of these k-tuples where each coordinate is either constant or varies with the same value (e.g. $(1, 1, *, *, 3, 5)$, where $*$ is always the same in both components; $*$ can be thought of as a wildcard). This sort of line is more restrictive than the lines allowed in Tic-Tac-Toe, and is sometimes known as a Hales-Jewett line.

Theorem 6.5 (Hales-Jewett)**.** *If d is large enough (depending on k and r)and $\{1, \dots, k\}^d$ is r-colored, then there exists a monochromatic combinatorial line.*

This means that for sufficiently large board, there is always a winner in Tic-Tac-Toe of many dimensions. Note that this is explicitly not true when $r = 2$ and $d = 2$. A statement called the density version of this theorem also deals with how much of $\{1, \dots, k\}^d$ is colored before the statement holds.

It is possible to restate Ramsey's theorem in a different way: for any graph G on n vertices, one can color K_n by taking an edge to be red if it's in G, and blue if it's not. Then,

Definition. An independent set on a simple graph G is a set of vertices such that no edge connects any of the vertices. A clique is the opposite: it is a set of vertices where all possible edges between them exist.

Thus, Ramsey's theorem states that if $n \geq R(p,q)$, then any simple graph G on n vertices either contains p independent vertices or a q-clique.

Thus, one can ask how many edges a simple graph on n vertices contains given that it has no triangles. For example, one can have a bipartite graph, with the two sets of sizes a and $(n-a)$. Thus, one can obtain a graph of $a(n-a)$ vertices without triangles, which can be maximized as about $a^2/4$, when $a \approx n/2$. Specifically, one has $\lfloor n/2 \rfloor (n - \lfloor n/2 \rfloor)$ edges, which is almost half as many as in a complete graph. However, this is a highly nonrandom graph, since it's so structured, so one could have some theory as to whether a graph looks random or looks structured, or whether these patterns are true in random graphs as well.

Theorem 6.6 (Turán)**.** *If G is a simple graph that has more than $\lfloor n/2 \rfloor (n - \lfloor n/2 \rfloor)$ edges, then G contains a triangle.*

Proof. Let $K(n)$ be the maximum example of edges that a simple, triangle-free graph can have. A bipartite graph gives a lower bound of $K(n) \geq \lfloor n/2 \rfloor (n - \lfloor n/2 \rfloor)$ edges.

For some small values: $K(1) = 1$, $K(2) = 1$, $K(3) = 2$, etc. By induction, an upper bound can be discovered.

Choose an edge $e = \{1, 2\}$ and remove it and its vertices. Then, there are at most $K(n-2)$ edges. Then, back in the original graph G, each vertex can be connected to at most on of vertices 1 or 2, so $K(n) \leq K(n-2) + n - 1$. Applying this repeatedly, one can check by induction that $K(n) \leq \lfloor n/2 \rfloor (n - \lfloor n/2 \rfloor)$. ⊠

It's actually possible to prove a stronger result, which is that the only graph with $K(n)$ edges is the bipartite one mentioned above.

7. Turán's Theorem: 4/22/13

Recall Turán's theorem from the previous lecture: it shows that if a graph has no triangles, then it can have the maximum number of edges if it is a complete bipartite graph.

More generally, let G be a simple graph on n vertices, such that G contains no K_p. Let $M(n,p)$ denote the maximum number of edges that G can have. Then, if G has $M(n,p)$ edges, then adding an edge to G means it contains a K_p, so G must contain a K_{p-1}, and there are $n - (p-1)$ vertices not in this K_{p-1}. Thus, the maximum number of edges such that G doesn't contain a K_p can be found: among the $n - (p-1)$ vertices, there must be at most $M(n-(p-1), p)$ edges to ensure there is no K_p. There are $\binom{p-1}{2}$ edges connecting the vertices in the K_p, and in order to avoid a K_p coming from the K_{p-1}, the number of vertices joining the two parts must be at most $(p-2)(n-(p-1))$, since for each vertex in $n-(p-1)$, there must be at most $p-2$ connections to the K_{p-1}. Thus, one obtains the bound of

$$M(n,p) \leq M(n-(p-1),p) + \binom{p-1}{2} + (p-2)(n-(p-1)).$$

If $n \leq p-1$, then $M(n,p) = \binom{n}{2}$, since there aren't enough vertices to have a K_p. Thus, write $n = t(p-1) + r$, where t is the quotient and $r < p-1$ is the remainder. Thus,

$$M(n,p) \leq M((t-1)(p-1)+r, p) + \binom{p-1}{2} + (p-2)((t-1)(p-1)+r).$$

Exercise 7.1. Check that

$$M(t(p-1)+r, p) \leq \frac{(p-2)r^2}{2(p-1)} - \frac{r(p-1-r)}{2(p-1)}.$$

In fact, this bound is optimal, which can be shown using a $(p-1)$-partite graph: divide the vertex set "evenly" into $p-1$ sets $V_1, \dots, V_{p-1}$, such that $V_1, \dots, V_r$ each contain $t+1$ points, and $V_{r+1}, \dots, V_{p-1}$ each contain t points. Draw edges from any $v \in V_i$ to all vertices not in V_i. This doesn't contain a K_{p-1}, since that would imply there are connections within one of the V_i, and one can check that the number of edges matches the upper bound.

Of course, there's a variant on the problem: if instead of 3-cycles on considers 4-cycles, how many edges can such a graph have? This recalls the notion of girth from the first week, which is the length of the smallest cycle in a graph. If a graph is triangle-free and quadrilateral-free, then its girth is at least 5.

Theorem 7.1. *If G is simple and has no 3- or 4-cycles, then G has at most $O(n^{3/2})$ edges.*

This is substantially smaller than the triangle case, and as the girth increases, the exponent on n decreases.

Proof of Theorem 7.1. Proof by double-counting: pick a vertex u and look at the set $\{\{v,w\} \mid uv \in E, uw \in E\}$. If $d(u) = \deg u$, then the size of this set is $\binom{d(u)}{2} = d(u)(d(u)-1)/2$. Then, for any edge $\{v,w\}$ there is at most one u that they come from, or else there would be a 4-cycle. Thus, the number of "tents" of edges $v - u - w$ is $\sum_{u \in V} \binom{d(u)}{2} \leq \binom{n}{2}$ by above, so by the Handshaking Lemma, so

$$\sum_{u \in V} \frac{d(u)^2 - d(u)}{2} = \frac{1}{2} \sum_{u \in V} d(u)^2 - |E|.$$

This requires a little bit of analysis; specifically, take the Cauchy-Schwarz Inequality

$$\left| \sum_{i=1}^{n} x_i y_i \right|^2 \leq \left(\sum_{i=1}^{n} x_i^2 \right) \left(\sum_{i=1}^{n} y_i^2 \right).$$

Then,

$$4|E|^2 = \left(\sum_{u \in V} d(u) \cdot 1 \right)^2 \leq \left(\sum_{u \in V} 1 \right) \left(\sum_{u \in V} d(u)^2 \right) = n \left(\sum_{u \in V} d(u)^2 \right).$$

Thus, $1/2 \sum d(u)^2 \geq 2|E|^2/n$, so

$$\begin{aligned} \frac{2|E|^2}{n} - |E| &\leq \frac{n(n-1)}{2} \\ \implies 4|E|^2 - 2n|E| &\leq n^2(n-1). \end{aligned}$$

Intuitively, the use of Cauchy-Schwarz is due to considering the most evenly distributed case, or a pidgeonhole argument. But now, the quadratic formula can be used to find the maximum number of edges, giving

$$|E| \leq \frac{2n + \sqrt{4n^2 + 16n^2(n-1)}}{8} = \frac{n + n\sqrt{4n-3}}{4}. \qquad \boxtimes$$

This also gives a more concrete upper bound: G must have at most $\lfloor (n(1 + n\sqrt{4n-3})/4 \rfloor$ edges. This is not a tight bound, but no tight bound is known.

In some sense, the Cauchy-Schwarz Inequality is a statement that the largest number of edges occurs when they are evenly distributed.

Definition. If G is a simple graph, a Hamiltonian circuit on G is a closed path that travels through each vertex in the graph exactly once.

This is different from the Eulerian circuit, which travels every edge exactly once.

Theorem 7.2 (G. Dirac[11]). *Suppose that G is a simple graph on n vertices and each vertex has degree at least $n/2$. Then, G contains a Hamiltonian cycle.*

Proof. Take the maximal counterexample (i.e. one such that if any edge is added, then there is a Hamiltonian cycle). Thus, G doesn't contain any edge uv such that $G \cup uv$ contains a Hamiltonian cycle (that contains uv). Suppose $u = u_1$ is connected to u_{i+1} and $v = u_n$ is connected to u_i, where $u_1, \dots, u_n$ is the cycle. Then, there exist such i such that $u = u_1$ is connected to at least $n/2$ vertices u_{i+1} and $v = u_n$ is connected to at least $n/2$ vertices u_i by the pidgeonhole principle. $\boxtimes$

Definition. The chromatic number of a graph G is the minimum number of colors necessary to color the vertices of G such that no two adjacent vertices share the same color.

To think about the definition, here's a nice toy unsolved problem: color all of the points on the plane with χ colors. Are there exactly two points exactly distance 1 apart that have the same color? Somewhere between $\chi = 4$ and $\chi = 7$, inclusive, there aren't.

For $\chi = 3$ with colors R, B, and W, take some equilateral triangle and color each vertex a different color. Then, take the reflection of that triangle through the line between the red and blue vertices; the remaining vertex must be white. Thus, all points in a circle of radius $\sqrt{3}$ must be white, and take any two points on that circle which are one inch apart, so any 3-coloring of the plane has points of the same color of distance 1 apart.

[11] Not the famous Dirac.

8. Coloring the Vertices of a Graph: 4/24/13

A lot of the problems in this class can be difficult: they tend to require some sort of clever trick. There are various ways to think about this, but one good one is due to Pólya: if you can't solve, then there is a smaller problem you can't solve. Solving the smaller problem (or applying the same heuristic) will give insight into the larger one.

A big question in vertex coloring is: given a graph G, color the vertices such that no adjacent vertices have the same color. How many colors are necessary? As discussed in the previous lecture, take the set of points in the plane; then, if they are 3-colored, then there always exist 2 points of distance 1 apart with the same color. Thus, at least four colors are necessary to ensure there are no two points of distance 1 with the same color, but not much more than that is known.

However, it is possible to prove that an upper bound exists? Imagine tiling the plane with squares with diagonal distance slightly smaller than 1. Take each 3×3 set of these squares and color each one with a different color, and then repeat this. Thus, any two points with the same color are either within the same square (so they have distance less than 1) or different squares of more than distance 1 away. Thus, the problem has an upper bound of 9, and there's a more clever way of using hexagons to show that 7 colors is also sufficient. Thus, the answer to the problem is one of 4, 5, 6, or 7, but nobody knows which.

Definition. The chromatic number $\chi(G)$ of a finite graph G is the smallest number r such that the vertices of G can be r-colored without two adjacent vertices being the same color.

It can be assumed that G is simple: the existence of loops would cause issues with the statement of the problem, so they won't be considered, and any additional edges don't change the adjacency of the problem, and therefore don't change the chromatic number.

$\chi(K_n) = n$, since all pairs of vertices are adjacent. Thus, for any G with n vertices, $\chi(G) \leq n$, since each vertex can be given a different color. If $\chi(G) = 1$, then G has no edges. If $\chi(G) = 2$, then let V_1 be the set of vertices of color 1 and V_2 be the set of vertices of color 2. Then, there are no edges in V_1 or in V_2, so G is bipartite.

More generally, if $\chi(G) = r$, then the set V of vertices of G can be written as $V = \bigcup_{i=1}^{r} V_i$, where the V_i are disjoint, and there are no vertices entirely within any given V_i. In general, if there are more edges, the chromatic number increases. Thanks to Turán's theorem, there exist triangle-free graphs with arbitrarily large chromatic number.

Is it possible to do better than this?

Proposition 8.1. *Suppose* $\Delta = \max_{v \in V} \deg v$. *Then,* $\chi(G) \leq \Delta + 1$.

Proof. Use a greedy algorithm: build up the graph vertex-by-vertex. When any particular vertex is added, it must be connected to at most Δ vertices, so it can be given a color that is distinct from the color of each of these vertices, since there are $\Delta + 1$ colors. ⊠

In general, this is only a modest strengthening of the theorem, and it's possible to do a bit better.

Theorem 8.2 (Brooks). *If G is a connected simple graph, then $\chi(G) \leq \Delta$ except in two cases:*

- $G = K_{\Delta+1}$, *or*
- $\Delta = 2$ *and G is an odd-length cycle.*

It's also possible to state a lower bound.

Definition. Define the clique number of G to be the largest ℓ such that G contains an ℓ-clique (i.e. a K_ℓ).

All of the vertices in a clique must be colored different colors, so χ is at least the clique number. Additionally, if G can be split into independent sets $V_1, \dots, V_r$, then there is some V_j such that $|V_j| > n/r$. Thus, another lower bound is found: $\chi(G) > n/|V_j|$. This may be better or worse than the previous lower bound. Combining, $\chi(G) \geq \max(\ell, n/|V_j|)$.

Definition. The chromatic function $\chi(G, k)$ is the number of different ways in which one can k-color the graph G.

In this scheme, two colorings are the same if each vertex has the same color in the colorings. For example, a triangle can be k-colored in $k(k-1)(k-2)$ ways. In general, $\chi(G, k) = 0$ for $0 \leq k < \chi(G)$.

If $G = K_n$, then there are k ways to color the first vertex, $k - 1$ ways to color the second, etc. Thus, $\chi(K_n, k) = \binom{k}{n}$, which also neatly implies that $\chi(K_n, k) = 0$ when $0 \leq k < n$.

Theorem 8.3. *For any graph G, the chromatic function is an n^{th}-degree polynomial in k.*

Proof. Divide V into r independent sets $V_1, \dots, V_r$. Then, given such a partition, one can color all of the points in each V_i with the same color, and use different colors for different regions. There are $\binom{k}{r}$ such ways of doing this. Then, this must be summed over all ways of partitioning V into indepdent sets:

$$\chi(G,k) = \sum_{\substack{\text{partions of } G\\ \text{into independent sets}}} k(k-1)\cdots(k-r+1).$$

Thus, it is a finite linear combination of polynomials, so it is a polynomial. Its degree comes from the partition into sets of single points, since otherwise there are fewer permutations. This gives $k(k-1)\cdots(k-n+1)$, so the degree is at most n. If the degree were $n-1$, then there is a doubleton V_1 and some singletons $V_2, \dots, V_{n-2}$, so the number of options is $k(k-1)\cdots(k-n+2)\binom{n}{2}$, so the product is

$$k(k-1)\cdots(k-n+1) + \left(\binom{n}{2} - |E|\right)(k(k-1)\cdots(k-n+2)) + \cdots,$$

where everything not written has degree at most $n-2$. Thus, the coefficient of $k^n = 1$, of k^{n-1} is $-\binom{n}{2} + \binom{n}{2} - |E| = -|E|$.

This proof will be continued in the next lecture, but notice that the constant term is 0 (since a no-coloring of a graph has problems), the leading coefficient is 1, the degree is n, and the second coefficient is $-|E|$ in a simple graph. Additionally, this polynomial is 0 for $0 \le k < \chi$.

It will also be possible to show that the coefficients of the polynomial alternate in sign.

Probabilistic Constructions: The Legacy of Paul Erdös: 4/25/13

This lecture wasn't technically part of the class, as it was part of this quarter's SUMO Speaker Series, given by Professor Amir Dembo. However, it was relevant to Math 108, so it has been included in these notes.

Paul Erdös initiated the idea of the probabilistic method in 1947, and gave lots of nice examples and uses of it as some of his over 1500 papers. This is such a large number of papers that mathematicians are often measured based on their graph distance to Erdös, where connectedness is given by coauthorship of a paper. Professor Dembo's Erdös number is 2, and they tend to be small natural numbers.

A relevant book on this subject is Alon and Spencer's *The Probabilistic Method*. This lecture will touch only on the first two chapters. Sometimes, a course called Math 159 is offered, which discusses this in more detail.

The general goal is to use probability in ways that don't seem like they would need probability. On some finite set S let $f : S \to \{0,1\}$ (corresponding to some property: $f(x) = 1$ if x has the property, and $f(x) = 0$ if it doesn't). Then, the goal is to find an $x \in S$ such that $f(x) = 1$. To do so, put a probability distribution P on S, $X \sim P$, such that $\mathbb{E}[f(X) > 0] > 0$ (i.e. the expected value that $f(x)$ is positive is itself positive). If this is the case, there must be an x such that $f(x) = 1$. This isn't constructive, but this is precisely its power: it turns some nightmarish constructions into simple proofs of existence. However, it will require some cleverness.

Example (Erdös, 1947)**.** Take the complete graph K_n on n vertices (so that there are $\binom{n}{2}$ edges) and color its edges with 2 colors, red and blue. The goal is to find a monochromatic k-clique (i.e. a $K_k \subset K_n$ whose edges are either all red or all blue). Then, the Ramsey number $R(k,k)$ is the smallest number n such that this holds true for any coloring of K_n. The key here is that of course such a K_k can be created, but the condition is that it is present in all colorings.

Theorem (Erdös)**.** *If $\binom{n}{2}2^{1-\binom{k}{2}} < 1$, then $R(k,k) > n$.*

Proof. One way to find a lower bound is to find a coloring of K_n without any monochromatic K_k. But this is hard, so colorings will be chosen at random. Let S be the set of colorings of K_n, so that $|S| = 2^{\binom{n}{2}}$, and let P be the uniform distribution on S. Let $f(x) = 1$ is there exists no monochromatic K_k in the coloring and $f(x) = 0$ otherwise. Then, the goal is to show that $\mathbb{P}(f(x) = 1) > 0$, or equivalently that $\mathbb{P}(f(x) = 0) < 1$.

There are $\binom{n}{k}$ choices of k vertices in K_n, and each such choice is equivalent to a subgraph K_k of K_n. The union bound of probability states that $\mathbb{P}(A \cup B) \le \mathbb{P}(A) + \mathbb{P}(B)$, so

$$\mathbb{P}(f(x) = 0) \le \binom{n}{k}\mathbb{P}(\text{a specific } K_k \text{ is monochromatic}) = \binom{n}{k}2^{1-\binom{k}{2}},$$

since there are $\binom{k}{2}$ ways to color K_k and exactly two of them are monochromatic. ⊠

Thus, $R(k,k) \ge 2^{k/2}$ for all $k \ge 3$, and thus the Ramsey numbers grow at least exponentially.

Notice how easy this is compared to a construction. If one wants an algorithm, there's the possibility of an exhaustive search. which takes exponential time and is unrealistic for large n, or a smarter way: if $n = \lfloor 2^{k/2} \rfloor$, then $\binom{n}{k}2^{1-\binom{k}{2}} \ll 1$, so there is a very small probability that any given graph doesn't have the property. Since it's easy to check, one can just randomly guess a graph, and guess again in the unlikely event that the graph is bad. Thus, this nonconstructive proof provides an algorithm, albeit a probabilistic one.

In some ways, this is elementary, but not trivial: the proof is very easy to follow but hard to create. In some sense, it's really nice: someone else does all of the hard work.

Definition. A tournament on $V = \{1, \dots, n\}$ is a directed graph $T = (V, E)$ of n vertices such that for any distinct $x, y \in V$, either $(x, y) \in E$ or $(y, x) \in E$.[12]

Tournament graphs can be used to represent some competition, where the direction of the edge between x and y indicates the winner of a match between them.

A tournament has a property S_k such that if for any subset of k players there exists some other player who beat all of them. Clarly, this requires $k < n$, and the "for any subset" aspect makes it difficult unless $n \gg k$. What is the minimum value of n such that there exists a tournament on T with S_k? The probabilistic method can provide an upper bound by showing that a randomly chosen tournament T will have S_k with positive probability.

Claim (Erdös, 1963)**.** If $\binom{n}{k}(1-2^{-k})^{n-k} < 1$, then there exists a tournament on n vertices with S_k.

Proof. It will be shown that a random tournament will have probability of not satisfying S_k with probability strictly less than 1. For all fixed subsets K of size k, let A_k be the complement to S_k. Thus, $\mathbb{P}(A_k) = (1-2^{-k})^{n-k}$: iterate over all $j \notin K$ (there are $n-k$ such j) and check if each j beat every $k \in K$. Since the tournament was chosen uniformly at random, this has probability $1 - 2^{-k}$, so

$$\mathbb{P}(T \text{ doesn't have } S_k) = \mathbb{P}\left(\bigcup_{\substack{K \subseteq V \\ |K|=k}} A_K\right) = \sum_{\substack{K \subseteq V \\ |K|=k}} \mathbb{P}(A_k) = \binom{n}{k}(1-2^{-k})^{n-k},$$

and if this is less than 1, then there must be a T with probability S_k. ⊠

This bound was chosen to be easy to prove, as with the probability distribution. A better bound can certainly be found, but it requires much more work and is more elaborate.

Remark.

(1) The bound guaranteed by the theorem is a chore to calculate, but ends up being $O\left(k2^k\right)$).
(2) There is a construction of such a T when $n \geq (1+\delta)^k$ for some $\delta > 0$, which is a little better.

Definition. A family of sets F is called intersecting (also a family of intersecting sets) if for any $A, B \in F$, $A \cap B \neq \emptyset$.

Theorem (Erdös-Ko-Rado)**.** *If $\mathcal{F}$ is a family of intersecting k-subsets*[13]*, of $\{0, 1, \dots, n-1\}$ and $n \geq 2k$, then $|\mathcal{F}| \leq \binom{n-1}{k-1}$.*

Proof. This proof is due to Kentona in 1972. Let $A_s = \{s, s+1 \bmod n, \dots, (s+k-1) \bmod n\}$, so that there are n of them.

Lemma. *$\mathcal{F}$ contains at most k of these sets.*

Assuming the lemma (whose proof is skipped, but very easy), let σ be a uniformly chosen permutation of $\{0, \dots, n-1\}$ and i uniformly chosen in $\{0, \dots, n-1\}$ independently of σ. Then, let $A = \{\sigma(i), \sigma(i+1 \bmod n), \dots, \sigma(i+k-1 \bmod n)\}$. Clearly, $\mathbb{P}(A \in \mathcal{F} \mid \sigma) \leq k/n$ since there are n choices for i, and only k can be present by the lemma, so uncondition: $\mathbb{P}(A \in \mathcal{F}) \leq k/n$. But since σ and i were chosen at random, A is uniformly distributed among all k-subsets of $\{0, \dots, n-1\}$, so $P(A \in \mathcal{F}) = |\mathcal{F}|/\binom{n}{k}$. ⊠

This upper bound is actually tight: take $F = \{A \subset \{0, \dots, n-1\} \mid 0 \in A\}$, so there are $\binom{n-1}{k-1}$ such sets that all intersect.

The beauty of the probabilistic method is that there is no probability in the questions it tackles. Thus, it can be used to say a lot of very deep things about how probability relates to the rest of mathematics, and in more ways than just counting. This relates probability to group theory, topology, etc., and is the most far-reaching use of probability in mathematics.

[12] When this graph is taken as an undirected graph, one obtains a complete graph.

[13] i.e. they are all of size k.

9. Chromatic Polynomials: 4/29/13

In the proof of Theorem 8.3, the key was to write the vertex set V of a graph G as $V = V_1 \cup \cdots \cup V_r$, were the sets $V_1, \ldots, V_r$ are disjoint and independent. Then, enumerating the possible colorings was fairly easy: there are $k(k-1)\cdots(k-r+1)$ of them. Thus, the chromatic function is a polynomial of degree n and its second-order term is $-|E|$, where E is the edge set of G. Some other things were shown; see the previous lecture for details.

A recurrence can be used to compute $\chi(G, k)$: let $e \in E$, and denote $G - e$ be the graph obtained by removing e and leaving the vertices unchanged. Clearly, $\chi(G - e, k) \geq \chi(G, k)$, since every coloring still works when an edge is removed. If the endpoints of e have different colors, then a coloring of $G - e$ implies a coloring of G, and the leftover case is when the endpoints have the same color. Let $G \cdot e$ be the graph formed by removing e from G and identifying its endpoints. This has one fewer vertex and one fewer edge than G. Note that $G \cdot e$ might not be simple, even if G is. Then, $\chi(G - e, k) = \chi(G, k) + \chi(G \cdot e, k)$, where $\chi(G \cdot e, k)$ is calculated by removing the extraneous edges (since they have no effect on the coloring). This can be rewritten as $\chi(G, k) = \chi(G - e, k) - \chi(G \cdot e, k)$, which is useful because both graphs on the right-hand side are smaller. It can be used to compute the chromatic polynomial of a graph, though at some base case it is necessary to just compute it straightforwardly. This argument can be used to show that if T is a tree, then $\chi(T, k) = k(k-1)^{n-1}$.

Remark. Using a similar contraction-deletion argument, one can obtain a formula for $\tau(G)$, the number of spanning trees of G. From Theorem 2.2, we know $\tau(K_n) = n^{n-2}$, but more generally $\tau(G) = \tau(G - e) + \tau(G \cdot e)$: choose an edge and ask whether it belongs in the spanning tree (in which case we go to the right) or not (left, or $\tau(G - e)$).

Claim. The coefficients of $\chi(G, k)$ alternate in sign.

Proof. This can be seen by the recurrence: if G has no edges, then $\chi(G, k) = k^n$, which trivially alternates in sign. Then, using induction, $\chi(G, k) = \chi(G - e, k) - \chi(G \cdot e, k)$; the first term on the right-hand side has degree n, so its $(n-1)^{\text{st}}$ term is negative and its terms alternate, and $\deg \chi(G \cdot e, k))$ has degree $n - 1$, so its terms alternate in the opposite way. Thus, taking $-\chi(G \cdot e, k))$ makes all of the signs align, so to speak. ⊠

If G is not connected, so that $G = G_1 \cup G_2$, where G_1, G_2 are its connected components, then $\chi(G, k) = \chi(G_1, k)\chi(G_2, k)$.

Theorem 9.1 (Erdös). *There exist simple connected graphs with arbitrarily large girth and arbitrary large chromatic number.*

The probabilistic method is also useful here.

Definition. A planar graph is a graph that can be drawn in the plane such that no edges cross (they only meet at vertices).

There are many ways of drawing a graph; K_4 is planar, even though the intuitive representation of it is not planar, since there exists a drawing of it that is planar. Technically, in order to speak of the "inside" and "outside" of curves (the edges), one needs to prove something complicated like the Jordan Curve Theorem, but here, any edge can be represented by a piecewise linear curve, for which the theorem isn't necessary. Thus,

(1) there are no topological problems,
(2) and even if there were, we wouldn't care, anyway.

A planar graph thus separates the plane into regions, called faces. The number of faces isn't obviously well-defined, since it depends on the way that the graph is drawn. However, the following is known.

Theorem 9.2 (Euler). *Suppose a planar graph G has V vertices, E edges, and divides the plane into F faces. Then, $V - E + F = 2$.*

Example 9.1. If G is a tree, then there is one face (since there are no cycles), so $V - E + 1 = 2$, or $V = E + 1$.

There are hundreds of proofs of Euler's theorem, and here's one:

Proof of Theorem 9.2. Proceed by induction on F. The base case is due to Example 9.1, so suppose that G is not a tree, or that there exists a cycle in G. Take some edge in the cycle and delete it; then, the nuber of faces is smaller, so by induction $V - (E - 1) + (F - 1) = 2$, so $V - E + F = 2$. ⊠

Thus, the number of faces for a planar graph is actually well-defined. The 2 in Euler's theorem relates to the surface, and is called the Euler characteristic of a surface: for example, on a torus there would be a different constant.

Another useful fact is that a connected, simple planar graph doensn't have "too many" edges, which allows for a nice bound on the chromatic number.

Theorem 9.3. *Every planar graph can be colored using five colors.*

Actually, only four are necessary:

Theorem 9.4 (Four-Color Conjecture)**.** *Every planar graph can be colored using four colors.*

This was proven using a computer, which sounds scary until one realizes that the 1970s didn't have a lot of computing power compared to today. There were philosophical concerns about using a computer-generated proof, but 40 years later most people accept it. However, a better proof was given in 1997 by Seymour et. al., which only required 5 minutes of computing time.

After all, we have computers, so why not use them? A computer-aided proof was also used to verify the Kepler conjecture, which stated that a hexagonal arrangement is optimal for packing spheres.

The statement that a planar graph doesn't have too many edges can be clarified: each face gives rise to at least 3 edges, but each edge could be counted twice (or not at all for some edges), so $E \geq 3F/2$. Using Euler's formula, $2 = V - E + F \leq V - E + 2E/3$, so $E/3 \leq V - 2$, or $E \leq 3V - 6$. This is relatively low given how many edges are possible.

10. Planar Graphs and Matchings: 5/1/13

As was shown last lecture, for a planar graph, $E \leq 3(V-2)$ and $2 = V - E + F$. Thus, K_5 is not planar: it has 5 vertices and 10 edges, but to be planar it would need $E \leq 3(5-2) = 9$. Similarly, $K_{3,3}$ isn't planar: it has 6 vertices and 9 edges. In a bipartite graph, each face must be bounded by at least four edges (since there are no triangles), so $E \geq 4F/2 = 2F$, so $2 = V - E + F \leq V - E/2$, so $E \leq 2(V-2)$ if a bipartite graph is planar, which is a problem.

Claim. If G is a planar graph, then G has a vertex of degree at most 5.

Proof. If $\deg(v) \geq 6$ for all $v \in V$, then $2E = \sum \deg v \geq 6V$, so $E \geq 3V$, but it is necessary that $E \leq 3(V-2)$. ☒

This leads to a (not strict) upper bound on the chromatic number of a planar graph:

Theorem 10.1 (Six-Color Theorem)**.** *Any planar graph can be 6-colored, or if G is a planar graph, then $\chi(G) \leq 6$.*

Proof. Proceed by induction on the number of vertices. If $V < 6$, then of course G is 6-colorable, because each vertex can be assigned a different color.

In general, let x be a vertex of degree at most 5, as per the claim. Then, by induction $G - x$ is 6-colored, because it is also planar and has strictly fewer vertices. Then, bring that coloring to G; at most five different colors are adjacent to x, so it can be given the remaining color. ☒

Six colors are actually too many:

Proof of Theorem 9.3. Proceed by induction again, with the base case essentially the same: a graph with fewer than 5 vertices can of course be 5-colored.

In a general planar graph, pick a vertex x of degree at most 5 and by induction, 5-color $G - x$. It can be assumed that $\deg(x) = 5$, and furthermore that each of its neighbors $x_1, \dots, x_5$ have different colors (so color x_i with color i), because otherwise we would be done. Thus, it will be necessary to recolor some of the neighbors of x.

Look at the subgraph of vertices colored only with colors 1 and 3, called $H_{1,3}$. This is a planar bipartite graph, and isn't necessarily connected.

- If x_1 and x_3 are in different connected components of $H_{1,3}$, take the connected component of x_1 in $H_{1,3}$ and switch all coloings of 1 and 3 in that component. This is still a valid coloring of G, since there wasn't any conflict beforehand and switching the color in a connected component doesn't create one. Thus, x_1 can be given color 3 while x_3 still has color 3, so x can be given color 1 and G has a valid 5-coloring.
- If x_1 and x_3 lie in the same connected component of $H_{1,3}$, then take the same argument with colors 2 and 4 and the analogously defined subgraph $H_{2,4}$. If they're in different connected components, then the previous item allows $G - x$ to be recolored and G to be 5-colored, or they are in the same connected component. Then, there is a path from x_1 to x_3 containing vertices only of vertices colored with colors 1 and 3, and there is a path from x_2 to x_4 in the same manner, but they must intersect, because G is planar.[14] Thus, this can't be possible. ☒

[14] This requires that x_2 or x_4 be between x_1 and x_3 in the plane, but strictly speaking this isn't always the case. However, the colors can be rearranged such that x_2 does lie between x_1 and x_3, since some colors must lie between other colors around x.

This proof doesn't use x_5 at all, which made people think there was some clever trick that would allow it to be reduced to four colors. In fact, the discoverer of the proof, Kempe, published it as a proof of the Four-Color theorem, and it was accepted as such for ten years. This was not the only such false proof, so much that when the real proof was reported, people were wary of reporting it!

Moving to matchings of graphs, Gale and Shapley's solution to the stable marriage problem recently won a Nobel Prize in Ecomonics.[15] The paper itself was called *College Admissions and the Stability of Marriage*, one of the most flavorful names for a math paper in a long tine, and one that illustrates some of the many applications of this problem (there are lots of them, in fields as diverse as organ transplants).

Definition. A matching of a graph G is a subset of the edge set such that no vertex appears in more than one edge of the matching. A perfect matching is when all vertices appear exactly once.

One can consider a bipartite graph of a set M of men and a set W of women, where it is guaranteed that there exists a matching such that every vertex in M is in an edge (so that $|W| \geq |M|$).[16] Additionally, for any $m \in M$, $\deg(m) \geq 1$, or else no suitable matching would exist. More generally, if $I \subseteq M$, then $|\bigcup_{i\in I} W(i)| \geq |I|$, where $W(i)$ is the set of vertices neighboring I. Clearly, these are necessary conditions.

Theorem 10.2 (Hall's Marriage Theorem)**.** *These conditions are sufficient; if for any $I \subseteq M$, $|\bigcup_{i\in I} W(i)| \geq |I|$, then G admits a matching in which all elements of M are paired up.*

Corollary 10.3. *If $G = M \cup W$ is a k-regular bipartite graph and $|M| = |W|$, then G has a perfect matching.*

Proof. Exercise; it is necessary to check that this condition implies the one in the theorem.

Theorem 10.2 can be reformulated as: if $A_1, \dots, A_N$ are some finite sets such that for any $I \subseteq \{1, \dots, N\}$ we have $|\bigcup_{i\in I} A_i| \geq |I|$, then it is possible to find distinct elements $a_1, \dots, a_N$ with $a_i \in A_i$. This is called a system of distinct representatives (SDR).

This has applications to group theory: suppose G is a finite group and $H \leq G$. Then, there exist $g_1, \dots, g_m \in G$ (where $m = [G : H]$) such that $G = \bigcup_{i=1}^m g_1 H = \bigcup_{i=1}^m H g_1$.

Proof of Theorem 10.2. Proceed by induction on N. If $N = 1$, then of course an SDR exists for one set; just choose any element of the set.

Call a collection of sets $\{A_i : i \in I\}$ for $1 \leq |I| \leq N$ critical if $\left|\bigcup_{i\in I} A_i\right| = |I|$ (e.g. ten men who among them know exactly ten women). Thus, there are two cases:

- Suppose there are no critical collections. Then, pick some $a_N \in A_N$, and let $\tilde{A}_i = A_i - a_N$, and each such set is nonempty. Thus, for any collection $\{A_i\}$, $\left|\bigcup_{i\in I} A_i\right| \geq |I| + 1 \geq |I|$, so $\tilde{A}_1, \dots, \tilde{A}_{N-1}$ satisfies the hypothesis by induction. Thus, there is an SDR $\tilde{a}_1, \dots, \tilde{a}_{N-1}$, and each element is distinct from a_N, Thus, $\tilde{a}_1, \dots, \tilde{a}_{N-1}, a_N$ is an SDR for the whole collection.

11. The Gale-Shapley Theorem and Network Flow: 5/6/13

- Continuing with the proof of Theorem 10.2, suppose there exists a critical collection $A_1, \dots, A_k$ for $1 \leq k \leq N - 1$ (by relabeling if necessary). Then, $\left|\bigcup_{i=1}^n A_i\right| = k$. By the inductive hypothesis, there are distinct $a_1 \in A_1, \dots, a_k \in A_k$. Let $\tilde{A}_j = A_j - \{a_1, \dots, a_k\}$, for $k + 1 \leq j \leq n$. These sets are nonempty: if one was, then (re)label it A_{k+1}, so that $|A_1 \cup \dots \cup A_{k+1}| = k$, which contradicts the initial assumption.

  Claim. For any $J \subseteq \{k + 1, \dots, n\}$, $\left|\bigcup_{j\in J} \tilde{A}_j\right| \geq |J|$.

  Proof.
$$\left|\bigcup_{i=1}^k A_i \cup \bigcup_{j\in J} A_j\right| = \left|\bigcup_{j\in J} \tilde{A}_j\right| + k \geq k + |J|$$
  by the original assumption. ☒

  Thus, one can find $a_{k+1} \in \tilde{A}_{k+1}, \dots, a_n \in \tilde{A}_N$ distinct from each other and the $a_1, \dots, a_k$. ☒

[15] Since Gale was dead, Shapley won the prize, along with Roth, a Stanford business professor.

[16] If one wishes to have a perfect matching, it is necessary that $|M| = |W|$.

This is useful in lots of places; for example, the 2012 Putnam B3 can be solved by using this, and the trick is seeing exactly how.

A related problem is that of stable marriage, in which each man and woman has a list of preferences, and each person prefers to be married to someone than to be unmarried. Thus, the goal is to pair the men and women, but to avoid unstable marriages, defined as pairs m_1, w_1 and m_2, w_2 if m_1 prefers w_2 to w_1 and w_2 prefers m_1 to m_2. A stable pairing is a pairing such that there are none of these sorts of marriages.

Theorem 11.1 (Gale-Shapley)**.** *For any such choices of preferences, there exists a stable matching.*

There are lots and lots of variations, such as finding stable pairs of roomates (in which case the graph isn't bipartite), finding optimal preferences for colleges, etc. This can even help in terms of organ transplants. Apparently, this is used by Stanford Housing to assign staff to dorms.

Proof of Theorem 11.1. The proof is essentially the "old-fashioned marriage proposal." To wit:

(1) Each man proposes to the woman he likes most.[17] Each man proposes to only one woman, and keeps the proposal until it is rejected.
(2) If all women get a proposal, they all accept, and a stable matching is found.
(3) Otherwise, some women get more than one proposal. Each such woman accepts the one she prefers most, and rejects the rest. A woman who has only one proposal still waits; after all, someone better might come along.
(4) The men who were rejected propose to the woman they most prefer who have not yet turned them down. Then, return to step 2.

This algorithm terminates, because each time a man is rejected, the number of choices he has dwindles, but since everyone would prefer to be married, then eventually someone accepts. The matching found is stable, because if it weren't, then suppose m_1, w_1 and m_2, w_2 were unstable with m_1 and w_2 preferring each other. Then, m_1 must have proposed to w_2 before w_1, and therefore rejected by her, which means that w_2 prefers m_2 to m_1, which generates stability.[18] ⊠

Note that this stable matching is not necessarily unique: one might obtain a different pairing by letting the woman propose and the men choose. It turns out, though, that the algorithm discussed above is most optimal for the men: each man gets the best of all possible women in a stable pairing, and each woman gets the worst of all possible men such that the pairing is still stable.

The last graph-theoretic topic of this class will be network flow. Here, one takes a directed graph and two distinguished nodes s and t to try to get something from s to t.

Definition. A flow on a directed graph is a real-valued positive function on the edges of the graph, such that the net flow of all values at a vertex (signed by whether they're coming in or going out) is zero. Formally, if $v \in V$, then $\sum_x f(v, x) - \sum_y f(y, v) = 0$ (where the sums are over all x, y where those edges exist, or such that the flow is defined).

Note that negative weights could be used instead of directed edges, but one is more intuitive than the other.

Then, there can be a capacity on each edge $c(x, y) \geq 0$, representing the maximum possible flow through the pipe x to y, so the only flows considered are those for which $f(x, y) \leq c(x, y)$. Define the volume of a flow f to be $|f| = \sum_v f(s, v) - \sum_u f(u, s)$. Then, what is the flow with the maximum volume?

Claim. Then, since there are no points of accumulation, the amount flowing out of s must be equal to the amount flowing into t.

Proof.

$$\sum_{v \neq s,t} \left(\sum_w f(v, w) - \sum_u f(u, v) \right) = 0,$$

and

$$\sum_v \left(\sum_w f(v, w) - \sum_u f(u, v) \right) = 0,$$

so

$$\sum_{v = s,t} \left(\sum_w f(v, w) - \sum_u f(u, v) \right) = 0,$$

which is just the difference in the net flow out of s and the net flow into t. ⊠

[17] Of course, you can switch the women and the men in this scenario, which is better for the women and worse for the men. . .

[18] Notice the lack of formalism in this proof. This is okay, because the original paper didn't have any either!

More generally, one could take some set $S \subseteq V$, and look at the flow from things in S to things not in S.

Claim.

$$\sum_{\substack{u\in S\\ v\notin S}} f(u,v) - \sum_{\substack{u\notin S\\ v\in S}} f(u,v) = \sum_{u\in S}\sum_{v} f(u,v) - \sum_{v\in S}\sum_{u} f(u,v).$$

That is, if the things in S are included in the flow calculation, the result doesn't change: this is just the sum of the net flows at u for each $u \in S$.

If $S \subseteq V - \{s,t\}$, then this is zero, and if $s \in S$ but $t \notin S$, then this is $\operatorname{vol}(f) = |f|$.

Definition. A cut is a collection of vertices S with $s \in S$ but $t \in S^c$, or a partition such that s is in one half, but t is in the other.

Thus, for any cut, we have

$$|f| = \sum_{\substack{u\in S\\ v\in S^c}} f(u,v) - \sum_{\substack{u\in S^c\\ v\in S}} f(u,v).$$

12. Network Flow II: 5/13/13

Recall that on a directed graph G, a flow is a nonnegative function $f(x,y)$ on the edges of G (i.e. $f(x,y) = 0$ if there is no edge from x to y), such that for any $v \neq s,t$, $\sum_w f(v,w) - \sum_w f(w,v) = 0$ (intuitively, there are no leaks in the graph). If the maximum possible flow is $c(x,y)$, which is some other nonnegative function, one considers only flows such that $0 \le f(x,y) \le c(x,y)$. In some situations, it will be assumed that the capacity is finite, which doesn't make too much of a difference if it's sufficiently large.

Suppose S is any subset of the vertex set V, and $\overline{S} = V - S$. Then,

$$\sum_{\substack{x\in S\\ y\in \overline{S}}} f(x,y) - \sum_{\substack{x\in S\\ y\in \overline{S}}} f(y,x) = \sum_{x\in S} f(x,y) - \sum_{x\in S} f(x,y).$$

This is zero if $S \subseteq V - \{s,t\}$, and if $s \in S$ but $t \notin S$, then this becomes $|f|$. Recall that a cut is a partition of G such that s is on one side and t is on the other. Then, the capacity of a cut is $\sum_{x\in S, y\in\overline{S}} c(x,y)$, which represents the amount that can flow at the boundary. Then, since $f(x,y) \le c(x,y)$ for all x and y, then

$$|f| = \sum_{\substack{x\in S\\ y\in \overline{S}}} f(x,y) - \sum_{\substack{x\in S\\ y\in \overline{S}}} f(y,x) \le \sum_{\substack{x\in S\\ y\in \overline{S}}} c(x,y).$$

Thus, the capacity of any cut is at least the volume of the flow. This leads to something a little more interesting: if the goal is to find a flow of maximum volume, then the maximum volume is less than the capacity of any cut, or the maximum volume is at most the minimum capacity of a cut. Since there are a finite number of options, then clearly the minimum capacity exists, but the maximum flow is more nuanced: just because it is bounded above doesn't mean it has a maximum, as in the sequence $\{x \in \mathbb{Q}, x^2 < 2\}$. This requies a little bit of elementary analysis to address fully: specifically. using the Bolzano-Weierstrauss theorem, it is possible to show there is a maximal flow on each edge and therefore on the entire graph by taking subsequences that converge on each edge.

Theorem 12.1. *In fact, these quantities are equal: the maximum flow is equal to the minimum cut.*

This leads to some number of algorithms for finding the maximum flow given the capacities. The maximum flow might not be unique, however; several different flows could lead to the same total flow. The following proof is due to Ford and Fullerson.

Proof of Theorem 12.1. Take a maximal flow f. Then, the goal is to construct a cut S such that the capacity of S is equal to $|f|$. First, put $s \in S$. Then, pick any $v \notin S$, and check if there is an $x \in S$ such that $f(x,v) < c(x,v)$ or $f(v,x) > 0$. If so, then add v to S. Then, repeat as long as there are vertices to add. Then, there are two things to show: first, that S is a cut (so that $t \notin S$) and that the capacity of S is $|f|$.

Suppose $t \in S$. Then, there is a sequence $x_0 = s, x_1, \dots, x_n = t$ such that for all j, either $c(x_j, x_{j+1}) - f(x_j, x_{j+1}) > 0$ (so there's more room) or $f(x_{j+1}, x_j) > 0$ (so there is backflow). Let $\varepsilon_j = \max(c(x_j, x_{j+1}) - f(x_j, x_{j+1}), f(x_{j+1}, x_j))$, so that $\varepsilon_j > 0$, and $\varepsilon = \min \varepsilon_j$. Then, adjust the flow: if ε is the first case, then increase the flow through that point by ε, and if the second case happens, reduce the backflow by ε.

This is a valid flow that satisfies the capacity bounds, and its volume is $|f| + \varepsilon$. This is very much like an Euler trail, with something going in and something coming out. Then, since the volume is increased, f isn't a maximal flow, which is a problem. Thus, it is necessary that $t \notin S$, so S is a cut.

The capacity of this cut is

$$\sum_{\substack{x\in S\\ y\in\overline{S}}} c(x,y) = \sum_{\substack{x\in S\\ y\in\overline{S}}} f(x,y) - \sum_{\substack{x\in\overline{S}\\ y\in S}} f(x,y) = |f|. \qquad \boxtimes$$

13. Network Flow III: 5/15/13

An algorithm for finding the maximum flow can be given if it is assumed that the capacities are all nonnegative integers:

(1) Start with a flow of 0 on every edge.
(2) Construct S as in the proof of Theorem 12.1, and in particular, find some $\varepsilon > 0$. Then, it is possible to increase the flow by 1 (since all of the capacities are integers, so this won't overflow it).
(3) Repeat until S is a cut.

This proof clearly also works for rational numbers, because the capacities can just be scaled to some common denominator. However, for irrational numbers, this is not so easy, even though the theorem guarantees that there at least is a solution. In the real world, this doesn't make too much of a difference, since everything can be approximated by a rational number.

This has a nice application to the proof of Hall's theorem: let G be a bipartite graph $G = V_1 \cup V_2$, with $|V_1| = |V_2| = n$.

Proof of Theorem 10.2. Add to G two vertices s, connected to every vertex in V_1, and t, connected to every vertex in t_2. Then, give capacity 1 to all edges adjacent to s or t, and some integer large capacity to every other edge, such as n. Then, the minimum flow must be equal to the maximum cut, so suppose S is a cut that contains s but not t. Then, if $S \cap V_1$ and $y \in \overline{S} \cap V_2$ and (x, y) is an edge in G, then this is fine, because $c(x, y) = n$. Then, the flow from s to t passes through the neighbors of $S \cap V_1$ in V_2, of which there are at least $|S \cap V_1|$ by the precondition, so each they are also in S. Thus, the capacity of the flow is at least n, so the theorem follows. $\boxtimes$

There are lots of variants on Menger's theorem, and here's one related to network flow and edge connectivity:

Theorem 13.1 (Menger)**.** *Suppose G is an undirected graph and $s, t \in V(G)$. Then, let p be the maximum number of paths from s to t such that all used edges are disjoint, and let p' be the minimum number of edges required to disconnect s and t. Then, $p = p'$.*

Proof. Consider a directed graph in which every edge $\{x, y\}$ in G corresponds to two edges $\overrightarrow{xy}$ and $\overrightarrow{yx}$, and give each edge a capacity of 1. Then, the max flow with $f(x, y) = 0$ or 1 is the number of edge-disjoint paths from s to t, which makes sense by thinking about how the flow works, and the min cut is the smallest number of edges necessary to disconnect s and t, so since the max-flow and min-cut are equal, then $p = p'$. $\boxtimes$

Turning to enumerative combinatorics, consider some famous theorems: first, Hall's theorem, Theorem 10.2 on SDRs, can be considered enumerative.

Definition. A poset, or partially ordered set, is a set S with an order relation $<$ such that not all elements are comparable, but if $x, y \in S$ are comparable, then $x < y$, $y < x$, or $x = y$, and if $x < y$ and $y < z$, then $x < z$.

Example 13.1. Let X be a finite set and S be its power set, ordered by inclusion, where if $A, B \subseteq S$, then $A < B$ iff $A \subsetneq B$ (equivalently, $A \leq B$ iff $A \subset B$). This is clearly a partial order, but there exist $A \neq B$ such that $A \not< B$ and $B \not< A$.

Definition. A chain in a poset is a sequence $x_1 < \cdots < x_n$, and an anti-chain is a set $\{x_1, \ldots, x_n\}$ such that no two elements are comparable.

Theorem 13.2 (Sperner)**.** *If X is a finite set of set n and S is its power set ordered by inclusion, then the size of a maximal antichain in S is $\binom{n}{\lfloor n/2\rfloor}$.*

Some examples of antichains are sets of singleton sets, such as $\{x_1\}, \ldots, \{x_n\}$, and note that $\max_k \binom{n}{k} = \binom{n}{\lfloor n/2\rfloor}$.

Proof of Theorem 13.2. Consider chains of n elements $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_n = X$, where $|A_i| = i$. Each A_i corresponds to adding one element to A_{i-1}, so this chain corresponds to a permutation of $\{1, \ldots, n\}$ specifying the order the numbers are added in, so there are $n!$ such chains.

Each chain contains at most one element of a given maximal antichain, so it has at most $n!$ elements. However, they can be double-counted: look at the antichains. If $|A| = k$, then A is in $k!(n-k)!$ chains, because the first k elements can be permuted while preserving A, and similarly the last $n - k$ ordered can be permuted. Then, $\sum_{|A|=k} k!(n-k)! \leq n!$

14. Sperner's Theorem: 5/20/13

Recall that a partial order on a set S (a poset) is an order $<$ such that exactly one of the following must hold: $a < b$, $b < a$, $a = b$. or a and b are incomparable. If the fourth option doesn't happen for any $a, b \in S$, then $<$ is a total order on S (e.g. $(\mathbb{R}, <)$).

Continuing the proof of Theorem 13.2, since $\sum_{|A|=k} n!/\binom{n}{k} = \sum_{|A|=k} k!(n-k)! \leq n!$, so if $\mathcal{A}$ is any antichain, then

$$\sum_{A\in\mathcal{A}} 1 \leq \sum_{|A|=k} \frac{\binom{n}{\lfloor n/2\rfloor}}{\binom{n}{k}} \leq \binom{n}{\lfloor n/2\rfloor},$$

so $|\mathcal{A}| \leq \binom{n}{\lfloor n/2\rfloor}$. ☒

Theorem 14.1 (Dilworth)**.** *If S is any finite poset, then the minimal number of chains in a partition of S into a union over all such partitions is equal to the size of the maximal antichain.*

Showing $\geq$ is considerably easier than $\leq$, but that is beyond the scope of this class.

In a set X, a family of k-subsets $\mathcal{A} = \{A_1, \dots, A_\ell\}$ is called intersecting if $A_i \cap A_j \neq \emptyset$ for all $1 \leq i, j \leq \ell$. If $|X| = n$, then the size of the largest intersecting family is at most $\binom{n}{k}$, and if $k > n/2$ then all k-element subsets intersect by the pigeonhole principle.

Theorem 14.2 (Erdös-Ko-Rado)**.** *Suppose $X = \{1, \dots, n\}$. Then, if $n \geq 2k$, then any intersecting family of k-subsets has at most $\binom{n-1}{k-1}$ elements.*

Proof. Arrange $\{1, \dots, n\}$ around a circle in some order. Since the circle can be rotated, there are $(n-1)!$ ways of arranging the elements. Given some k-family of intersecting sets $\mathcal{A}$, count the number of sets in $\mathcal{A}$ that appear as k consecutive elements as the circle is traversed clockwise (out of n possible).

It happens that there are at most k such sets, which we will go back and prove in a moment. Assuming it, there are at most $k(n-1)!$ sets over all such arrangements on the circle. Thus, we have

$$\begin{aligned} k(n-1)! &= \sum_{\substack{\text{circular}\\ \text{arrangements}}} \#\{k\text{-consecutive sets in } \mathcal{A}\} \\ &= \sum_{A\in\mathcal{A}} \#\{\text{circular arrangements in which } A \text{ appears } k\text{-consecutively}\}, \end{aligned}$$

so $|\mathcal{A}|k!(n-k)! \leq k(n-1)!$, so $|\mathcal{A}| \leq (n-1)!/(k-1)!(n-k)! = \binom{n-1}{k-1}$.

This is good, but now the fact still needs to be proven: suppose $A, B \in \mathcal{A}$ are k-consective. Then, they must intersect, and since $n \geq 2k$, then A and B must intersect, so each element of A must give rise to at most one element in $\mathcal{A}$; therefore, there are at most k of them in this arrangement. ☒

This leads into various problems in enumeration. For example, one might count functions from n-element sets to k-element sets.

Let S be a set of n elements and $A_1, \dots, A_k$ be a collection of subsets of S. Then, one can use the principle of inclusion and exclusion to calculate the number of elements of $\bigcup_{i=1}^k A_i = \bigcap_{i=1}^k (S - A_i)$ given $\bigcap_{i\in I} A_i$ for $I \subseteq \{1, \dots, k\}$.

Theorem 14.3 (Principle of Inclusion and Exclusion)**.**

$$\begin{aligned} \left| S - \bigcup_{i=1}^k A_i \right| &= n - \sum_{i=1}^k |A_i| + \sum_{1\leq i<j\leq k} |A_i \cap A_j| - \sum_{1\leq i<j<\ell\leq k} |A_i \cap A_j \cap A_\ell| + \cdots \\ &= n + \sum_{\substack{I\subseteq\{1,\dots,k\}\\ I\neq\emptyset}} (-1)^{|I|} \left| \bigcap_{i\in I} A_i \right|. \end{aligned} \tag{5}$$

In some sense, one starts by overcounting, then undercounting, then overcounting, and so on.

A neat application of this is that if $S = \{1, \dots, N\}$ and $p \leq \sqrt{n}$ is prime, then let $A_p = \{n \leq N \mid p \mid n\}$. If $\pi(x)$ represents the number of primes less than x, then then $S - \bigcup_{p\leq\sqrt{N}} A_p$ is the set of primes between $\sqrt{N}$ and N. Thus, the principle says that

$$1 + \pi(N) - \pi(\sqrt{N}) = N - \sum_{p\leq\sqrt{N}} |A_p| + \sum_{p,q} |A_p \cap A_q| - \cdots$$

More generally, for d not necessarily prime, let $A_d = \{n \leq N \mid d \mid n\}$, so that $|A_d| = \lfloor N/d \rfloor$. The numbers that are interesting here are products of distinct primes $d = p_1 \cdots p_\ell$ (i.e. are square-free). This number d appears in the formula with sign $(-1)^\ell$.

In number theory, the Möbius function is an important concept:

$$\mu(d) = \begin{cases} 1, & d = 1 \\ (-1)^\ell, & d = p_1 \cdots p_\ell \text{ are distinct} \\ 0, & p^2 \mid d \text{ for some prime } p \end{cases}$$

The formula then becomes

$$1 + \pi(N) - \pi(\sqrt{N}) = \sum_{\substack{1 \leq d \leq N \\ p \mid d \implies p \leq \sqrt{N}}} \mu(d)|A_d|.$$

This is an example of a sieve identity, called the Sieve of Eratosthenes.[19]

15. The Principle of Inclusion and Exclusion: 5/22/13

First, it will be helpful to prove the principle of inclusion and exclusion:

Proof of Theorem 14.3. The formula for $\left|\bigcup_{i=1}^k A_k\right|$ will be found: pick some $a \in S$ and suppose that it appears in r of the $A_1, \ldots, A_k$, so that $0 \leq r \leq k$. Then, the left-hand side of the formula (5) counts a once if $r \geq 1$ and zero times if $r = 0$. The right-hand side counts a zero times if $r = 0$, and if $r \geq 1$, then suppose $a \in A_{i_1}, \ldots, A_{i_r}$, and in the right-hand side of (5) a is counted

$$\sum_{\emptyset \neq I \subseteq \{i_1, \ldots, i_r\}} (-1)^{|I|-1} \cdot 1 = \sum_{\ell=1}^{r} (-1)^{\ell-1} \binom{r}{\ell},$$

times, where $|I| = \ell$. By the Binomial theorem,

$$0 = \sum_{\ell=0}^{r} \binom{r}{\ell} (-1)^{r-\ell} = 1 - \sum_{\ell=1}^{r} \binom{r}{\ell} (-1)^{\ell-1},$$

so the quantity calculated must be 1. The second formula is the complement of this one, and its formula follows as a result. $\boxtimes$

Consider the set of permutations on $\{1, \ldots, n\}$, which are required to be bijections.[20] There are $n!$ permutations, but consider the number of derangements: permutations which have no fixed points. Imagine putting n letters in n envelopes such that no letter goes in the correct envelope.[21] How many derangements are there? Let A_ℓ be the set of permutations that fix ℓ; then, none of the derangements belong to any of these sets, so the total number of derangements is

$$\left|\bigcap_{\ell=1}^{n} (S_n - A_\ell)\right| = n! + \sum_{\substack{I \subseteq \{1, \ldots, n\} \\ I \neq \emptyset}} \left|\bigcap_{i \in I} A_i\right|.$$

In $\left|\bigcap_{i \in I} A_i\right|$, the elements of I are fixed and the rest are permuted, so there are $(n - |I|)!$ such permutations, so this becomes

$$\begin{aligned} n\left|\bigcap_{\ell=1}^{n} (S_n - A_\ell)\right| &= n! + \sum_{\substack{|I|=\ell \\ \ell=1}}^{n} (-1)^\ell (n-\ell)! \binom{n}{\ell} = n! + \sum_{\ell=1}^{n} \frac{n!}{\ell!} \\ &= n! \left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^n}{n!}\right), \end{aligned}$$

which approaches $n!/e$ as $n \to \infty$. Thus, most permutations are derangements for large n. There is literature on what a random permutation might look like (which has applications, such as shuffling cards). The probability of having k fixed points is approximately $1/ek!$, which forms a Poisson distribution. There are lots of similar problems: imagine a set of 100 prisoners who are given numbers on their backs (which they cannot see). There are 100 boxes, and each prisoner

[19] There is a sculpture near the Cantor Arts Center that is relevant to this sieve.

[20] Recall that a function $f : A \to B$ is injective if $f(a) = f(b)$ for $a, b \in A$, then $a = b$. A function is surjective if it is onto: $f : A \to B$ is surjective if every $b \in B$ has an $a \in A$ such that $f(a) = b$. A bijection is a function that is both injective and surjective.

[21] Apparently this is a homework question in one of the professor's other classes. Oops.

can look at 50. If all of them see their numbers, they can go free. There's a reasonably good strategy that they can employ; can you find it?

A probem called the twelve-fold way asks for the number of functions $f : N \to K$, where $N = \{1, \dots, n\}$ and $K = \{1, \dots, k\}$. One may wish to count injections or surjections, or arbitrary functions, giving bijections. Then, there are four ways of thinking of the functions as the same (which leads to the number twelve): there may be no restricton, or one may wish to consider them up to a permutation of N, or of K, or of both.

This can be imagined as placing n balls in k boxes, such that there are no restrictions on the number of balls in each box, at most one ball in each box, or at least one ball in each box. One could have the balls colore with n colors, or that the balls have the same color and the boxes have different labels, or the balls have different colors and the boxes look identical, or the balls and boxes are indistinguishable among themselves. This forms a natural collection of enumeration problems, five of which are difficut and seven of which are easy.

Case i. There are k^n functions $N \to K$, where all balls and boxes are distinct.

Case ii. How many injective functions are there? There are k choices for the location for the first ball, and $k-1$ for the second, $k-2$ choices for the third, etc. If $k < n$, then there are zero choices, so the total number of functions is $k(k-1)(k-2)\cdots(k-n+1)$, a number called the falling factorial.

Case iii. How many surjective functions are there? We will return to this next week, using the principle of inclusion or exclusion or setting up a recurrence. This would be a good thing to think about. Let A_i be the set of balls in box i; since the function is surjective, then there are all nonempty sets. Thus, the goal is to partition $\{1, \dots, n\}$ into k nonempty sets. This is different than writing n as the sum of k numbers because the elements of the set are distinct.

Case iv. Suppose the balls are indistinguishable, but the boxes aren't. Then, the goal is to consider all such functions, so if box i contains a_i balls, then $a_1 + \cdots + a_k = n$, and it is possible to permute the a_i. There are two ways to approach this:

Imagine all of the balls are red. Put some partitions, represented by $k-1$ green balls, between the red ones, such that green balls may be adjacent. Then, everything to the left of the first green ball goes in box 1, everything between the first and the second goes to box 2, etc. Thus, the goal is to choose locations for the $k-1$ green balls among the total of $n+k-1$ slots (after which the red balls can be added), giving a total of $\binom{n+k-1}{k-1}$.

The second proof uses a very useful idea called a generating function. For each box i, take a sequence $(1 + x + x^2 + x^3 + \cdots)$. Then, the number of balls in box i is one of these coefficients, and the coefficient of x^n in the product of all of these terms is the product we want: it includes all distinguishable partitions of the n balls into k boxes. The geometric series converges to $1/(1-x)$, so if $F(x) = 1/(1-x)^k$ represents the product, one can compute the n^{th} derivative at zero, which once divided by $n!$ is the coefficient of x^n. This becomes $\binom{n+k-1}{k-1}$.

16. The Twelvefold Way: 5/29/13

Case v. Consider the set of injective functions in which the boxes, but not the balls, are distinct. Then, each box contains at most one ball, so choose n boxes out of the k, giving $\binom{k}{n}$ options. This is 0 if $k > n$.

Case vi. If the functions are requred to be surjective, the goal is to put n identical balls into k boxes such that each box has at least one ball. After removing one ball from each box, the goal is to put $n-k$ balls in k boxes with no conditions, which reduces to Case iv, giving $\binom{n-k+k-1}{k-1} = \binom{n-1}{k-1}$, and there are of course no ways if $k < n$. This illustrates another way to approach Case iv: one chooses ℓ boxes for some $1 \le \ell \le k$ that have at least one ball, so the total number is

$$\sum_{\ell=1}^{k} \binom{k}{\ell}\binom{n-1}{\ell-1} = \binom{n+k-1}{k-1}.$$

This itself is a useful combinatorial identity.

This can again be done with generating functions, by taking $(x + x^2 + \cdots)^k = x^k/(1-x)^k$ and looking at the coefficient of x^n.

Case vii. So now we suppose that the balls are distinguishable, but the boxes aren't. Then, the balls go in ℓ boxes for some ℓ, leading to the sum $\sum_{\ell=1}^{k} S(n,\ell)$, where $S(n,\ell)$ is as given below.

Case viii. For the injective functions, if $k \ge n$ then there is exactly one way to do this (since the boxes can be shuffled around), and otherwise there are none.

Case ix. For the surjective case, this asks the number of ways to partition $\{1, \dots, n\}$ into exactly k nonempty subsets (provided $n \geq k$; otherwise, it is zero). This number is known as the Stirling number of the second kind, $S(n, k)$.

This leads to a solution for Case iii: since there are $k!$ ways to permute the boxes, the number is $k!S(n, k)$.

Case x. Here, neither the balls nor the boxes have labels on them so the goal is to determine the number of ways of writing $n = a_1 + \cdots + a_k$, where $a_1 \geq a_2 \geq \cdots \geq a_k$.

Case xi. If these functions are injective, then each of the a_i above must be at most 1 and therefore there is one such way to order if $k \geq n$ and none if $k < n$. Using the notation defined below, suppose the balls end up in ℓ boxes, and after the summation one has $\sum_{\ell=1}^{k} p_\ell(n)$.

Case xii. If they must be surjective, then each of the a_i must be strictly positive. Thus, one obtains the number of ways of partitioning the number n into exactly k parts, which is similar to the Stirling number, but not the same, and it is denoted $p_k(n)$. Of course, this only makes sense if $n \geq k$; otherwise, it is zero.

Thus, it will be useful to obtain a formula for the Stirling number. There are several ways to obtain it. First, consider the number of cases where $\{n\}$ appears by itself. Thus, there are $S(n-1, k-1)$ ways for this to happen. Alternatively, if n appears in some set with at least one other element, then there are $kS(n-1, k)$ choices, so $S(n, k) = S(n-1, k-1) + kS(n-1, k)$. Then, one has the base cases $S(n, 1) = 1$ for $n \geq 1$ and $S(k, k) = 1$. One often adopts the convention that $S(0, 0) = 1$.

One could also use the principle of inclusion and exclusion, since we know there are $k!S(n, k)$ functions in one of the above cases: there are k^n functions from $\{1, \dots, n\}$ to $\{1, \dots, k\}$, so splitting up based on the subset of $\{1, \dots, k\}$ is the image, one obtains

$$k!S(n, k) = \sum_{\ell=0}^{k} (-1)^\ell \binom{k}{\ell} (k - \ell)^n,$$

which isn't very illuminating, but is at least explicit.

One could also consider coloring the graph on n vertices with no edges and x colors, so that there are x^n options. Then, the chromatic polynomial can be obtained by partitioning the vertices into k sets, and multiply by the number of ways to color these sets: $x(x-1)(x-2)\cdots(x-k+1) = (x)_k$, which is the falling factorial function. Thus,

$$x^n = \sum_{k=1}^{n} S(n, k)(x)_k.$$

It will also be nice to know $p_k(n)$, the number of partitions of n into k parts. A recursive formula is given by noticing that if one wishes to write $n = a_1 + \cdots + a_n$ such that $a_1 \geq \cdots \geq a_k \geq 1$, then either $a_k = 1$, whih gives $p_{k-1}(n-1)$ options, or $a_k \geq 2$, in which case all of the a_i are, so there is an expression for $n - k = (a_1 - 1) + \cdots + (a_k - 1) \geq 1$, contributing a factor of $p_k(n-k)$. Thus, $p_k(n) = p_{k-1}(n-1) + p_k(n-k)$.

17. Combinatorial Functions: 6/3/13

Recall the definitions of some functions given in the Twelvefold Way: the binomial coefficients are probably familiar, but also $S(n, k)$, the Stirling number, which is the number of set partitions of $\{1, \dots, n\}$ into k nonempty sets, which obeys the relation $S(n, k) = kS(n-1, k) + S(n-1, k-1)$. Additionally, there is the function $p_k(n)$, which is the number of partitions of n as $n = a_1 + \cdots + a_k$ such that $a_1 \geq a_2 \geq \cdots \geq a_k \geq 1$. This obeys $p_k(n) = p_{k-1}(n-1) + p_k(n-k)$, and there are other relations given in the textbook.

Consider bijections of $\{1, \dots, n\}$. Each bijection π can be written as a cycle decomposition by seeing where 1 goes, then $\pi(1)$, etc., until 1 is reached again. This is written as $(1\ \pi(1)\ \pi(\pi(1))\ \cdots)$. Then, repeat with a number that wasn't in the first cycle, and so on. Some numbers are fixed points: if $\pi(i) = i$, then the cycle is (i). The Stirling number of the first kind is the number of permutations of $\{1, \dots, n\}$ which have exactly k cycles. Then, summing over k, one would obtain all $n!$ permutations. Notice that the 100 prisoners problem presented earlier has a nice solution in terms of cycles: what is the probability there is a cycle of size at least 50?

This comes up in "nature:" 52! is the number of posible shuffles for a deck of cards. One could use this to determine whether a deck has been shuffled correctly: does it "look like" a random permutation? Markers indicating too many cards in the same position are a red flag.

In all of the $n!$ permutations, how many n-cycles are there? There are $(n-1)!$, because some cycle decompositions are identical, such as (1 2 3) and (2 3 1).

One can also use the principle of inclusion and exclusion: for example, how many permutations of $\{1, \dots, n\}$ have no cycle of length greater than $n/2$? By size constraints, each permutation must have at most one such cycle, so the answer is

$$n! - \sum_{\substack{n\geq k>n/2\\ \pi \text{ has a cycle of length } k}} 1 = n! - \sum_{n\geq k>n/2} \binom{n}{k}\prod_{i=1}^{k}(k-i)! = n! - \sum_{n\geq k>n/2}\frac{n!}{k}.$$

Notice that if $k \geq n/2$ this wouldn't work, because some things would be double-counted.

Recall the prisoners problem: one solution is to take prisoner n to look at box n and get a number, then look at the box with that number, and so on. The goal is to look up the cycles, and the prisoners can go free if all cycles are of length at most 50 (otherwise, the prisoners might not see the boxes with their own numbers). This is optimal, though it's not easy to show it. As shown above, the number of permutations for which this holds is $100!\left(1-\sum_{k=51}^{100}\frac{1}{k}\right)$. An approximation can be given as

$$\sum_{i=51}^{100}\int_i^{i+1}\frac{\mathrm{d}t}{t} < \sum_{k=51}^{100}\frac{1}{k} < \sum_{i=51}^{100}\int_{i-1}^{i}\frac{\mathrm{d}t}{t},$$

showing that the chance is at least $\log(101/51) \approx 1/3$ and at most $\log 2 \approx 0.693$, which is pretty nice. Notice that if this fails, then at least 50 people fail to see their number.

Further questions can be asked: what happens if fifty boxes are replaced by 35? In general, if they must inspect n/u out of n boxes, then as n gets large one sees the Dickman-de Bruijn function $\rho(u)$: we saw that $\rho(2) = \ln 2$, and this number is always positive. This actually has to do with unpublished work of Ramanujan (of course). This function ocurs in lots of ways: if one factors a random n (whatever random means here) into primes $n = p_1^{\alpha_1}\cdots p_k^{\alpha_k}$, then what is the chance that all of these prime factors are less than $\sqrt{n}$? This ends up being $1-\ln 2$, and there are all sorts of interesting connections between what a random number looks like, a random permutation looks like, and a random polynomial looks like, and there's a lot of structure here (Don Knuth did some work here, too). This function satisfies a differential difference equation, funnily enough: $u\rho'(u) = -\rho(u-1)$, and $\rho(u) = 1$ for $0 \leq u \leq 1$ (the prisoners always succeed). Ramanujan wrote down what the principle of inclusion and exclusion would give for this in cases for $u \leq 6$, and then said "and so on."

These sorts of things are useful in case one doesn't care as much about exact formulas, but instead the asympotics of how these combinatorial functions behave over the long run. Generally, one would use tools of analysis or calculus to determine this. This also relates to the philsophical idea of a good formula, which is an expression for a function that is ideally easy to compute or easy to understand.

One of the most basic and useful identities is Stirling's formula for $n!$ We don't know how to compute factorials exactly faster than $O(n)$ (just multiplying things together). Since there doesn't seem to be a great formula, there's at least one for an approximation:

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

where $\sim$ means that

$$\lim_{n\to\infty}\frac{n!}{\sqrt{2\pi n}\,(n/e)^n} = 1.$$

This can be used to understand how $p_k(n)$ or $S(n,k)$ behave as $n \to \infty$. One could also consider the horrible Bell numbers $B(n) = \sum_{k=1}^{n} S(n,k)$ (all possible partitions of $\{1,\ldots,n\}$ into nonempty sets), which have no good formula, but asymptotically are better to understand, and similarly $p(n) = \sum_{k=1}^{n} p_k(n)$. One can count trees on n vertices up to isomorphism, for example, but this is a headache for general n, and the best case is an asymptotic understanding.

Now for some formulas: one useful technique is to compare a sum with an integral of the expression being summed.

$$\log N! = \sum_{k=1}^{N}\log k \leq \sum_{k=1}^{N}\left(\int_k^{k+1}\log t\,\mathrm{d}t\right) = \int_1^{N+1}\log t\,\mathrm{d}t.$$

For a lower bound, we have

$$\log N! = \sum_{k=2}^{N}\log k \geq \sum_{k=2}^{N}\left(\int_{k-1}^{k}\log t\,\mathrm{d}t\right) = [t\log t - t]_1^N = N\log N - N + 1.$$

Expontntiating, $N! \geq (N/e)^N e$, which is nice to know. The integral for the upper bound can also be evaluated, giving $(N+1)\log(N+1) - N$. Since $\log(N+1) - \log N = \log(1+1/N)$, this can be expanded out using a Taylor series:

$$\log(1+x) = \int_0^x \frac{\mathrm{d}t}{1+t} = \int_0^x (1 - t + t^2 - t^3 + \cdots)\,\mathrm{d}t = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots,$$

so the difference can be found and simplified to $\log N! \leq (N+1)(\log N + 1/N) - N$. In summary,

$$\left(\frac{N}{e}\right)^N e \leq N! \leq \frac{N^{N+1}}{e^N} e^{1+1/N},$$

and the difference is a factor of $Ne^{1/N}$. Stirling's formula indicates that this difference is split, up to constant, as the $\sqrt{x}$ term.

In general, converting to an integral is a good way to approach a solution; for example, the harmonic series $\sum_{k=1}^N 1/k$ should be approximately $\log n$. Similarly,

$$\sum_{k=1}^N n^k \approx \int_1^N x^k \, \mathrm{d}x \approx \frac{N^{k+1}}{k+1},$$

and this approximation can be made precise with the upper and lower bounds, as seen in the previous examples.

Another approximation can be made: since $\log k \approx \int_{k-1}^k \log t \, \mathrm{d}t$, then for $k \geq 2$,

$$\int_{k-1}^k \log k \, \mathrm{d}t - \int_{k-1}^k \log t \, \mathrm{d}t = \int_{k-1}^k \log\left(\frac{k}{t}\right) \mathrm{d}t,$$

If $t = k - y$, then $0 \leq y \leq 1$ is an appropriate change of variables, yielding

$$= \int_0^1 \log\left(\frac{1}{1-y/k}\right) \mathrm{d}y.$$

The Taylor series $\log(1/(1-x)) = x + x^2/2 + x^3/3 + \cdots$ gives

$$= \int_0^1 \sum_{\ell=1}^\infty \left(\frac{y}{k}\right)^\ell \frac{1}{\ell} \, \mathrm{d}y = \sum_{\ell=1}^\infty \frac{1}{k^\ell} \cdot \frac{1}{\ell} \int_0^1 y^\ell \, \mathrm{d}y = \sum_{\ell=1}^\infty \frac{1}{k^\ell(\ell)(\ell+1)}.$$

Thus,

$$\log N! = \sum_{k=2}^N \log k = \int_1^N \log t \, \mathrm{d}t + \sum_{k=2}^N \left(\log k - \int_{k-1}^k \log t \, \mathrm{d}t\right).$$

A lot of things cancel out, as shown above, with the remainder

$$\sum_{k=2}^N \left(\sum_{\ell=1}^\infty \frac{1}{\ell(\ell+1)k^\ell}\right) = \frac{1}{2k} = \sum_{\ell \geq 2} \frac{1}{\ell(\ell+1)k^\ell},$$

where the sum is now constant as $N \to \infty$, and $\sum_{k=2}^N 1/(2k) \to (1/2)\log N + C$ for some constant C. Thus, the square root occurs in Stirling's formula, and the constant will lead to the rest of the term. The point is to see just how much can be done with single-variable calculus and some elbow grease.

18. More Techniques for Asymptotics: 6/5/13

Recall that Stirling's formula claims that $n! \sim \sqrt{2\pi n}(n/e)^n$, where $\sim$ means that the limit as $n \to \infty$ of their ratio is 1. We showed that $n! \sim C\sqrt{n}(n/e)^n$ last time, and can be more specific.

We know that $\binom{2n}{n}$ is the largest of all binomial coefficients $\binom{2n}{k}$ where n is fixed. One can ask how big it is, and how close nearby coefficients (e.g. $\binom{2n}{n+1}$, or more generally $\binom{2n}{n+\ell}$ for a "small" ℓ) are to it, at least in orders of magnitude. Using Stirling's formula,

$$\binom{2n}{n+\ell} = \frac{(2n)!}{(n+\ell)!(n-\ell)!} \sim \frac{C\sqrt{n}}{C\sqrt{n+\ell}C\sqrt{n-\ell}} \frac{(2n/e)^{2n}}{((n+\ell)/e)^{n+\ell}((n-\ell)/e)^{n-\ell}}$$

A couple of things quickly cancel out, showing $2^{2n}/(2n+1) \leq \binom{2n}{n} \leq 2^{2n}$:

$$\begin{aligned} &\sim \frac{\sqrt{2}}{C} \frac{1}{\sqrt{n}} \frac{(2n)^{2n}}{(n+\ell)^{n+\ell}(n-\ell)^{n-\ell}} \\ &\sim \frac{\sqrt{2}}{C} \frac{2^{2n}}{\sqrt{n}} \frac{n^{n+\ell}n^{n-\ell}}{(n+\ell)^{n+\ell}(n-\ell)^{n-\ell}} = \frac{\sqrt{2}}{C} \frac{2^{2n}}{\sqrt{n}} \left(\frac{n}{n+\ell}\right)^{n+\ell} \left(\frac{n}{n-\ell}\right)^{n-\ell}. \end{aligned}$$

We can take the logarithm of this quantity:

$$\log\left(\left(\frac{n}{n+\ell}\right)^{n+\ell} \left(\frac{n}{n-\ell}\right)^{n-\ell}\right) = (n+\ell)\log\left(\frac{n}{n+\ell}\right) + (n-\ell)\log\left(\frac{n}{n-\ell}\right).$$

When x is small, $\log(1+x) = x - x^2/2 + x^3/3 - \cdots$ and $\log(1-x)^{-1} = x + x^2/2 + x^3/3 + \cdots$, so

$$\begin{aligned}
(n+\ell)\log\left(\frac{1}{1+\ell/n}\right) &= -(n+\ell)\log\left(1+\frac{\ell}{n}\right)\\
&= -(n+\ell)\left(\frac{\ell}{n} - \frac{1}{2}\left(\frac{\ell}{n}\right)^2 + \cdots\right)\\
(n-\ell)\log\left(\frac{1}{1-\ell/n}\right) &= (n-\ell)\left(\frac{\ell}{n} + \frac{1}{2}\left(\frac{\ell}{n}\right)^2 + \cdots\right)\\
\implies (n+\ell)\log\left(\frac{1}{1+\ell/n}\right) + (n-\ell)\log\left(\frac{1}{1-\ell/n}\right) &= -2\ell\left(\frac{\ell}{n}\right) + \frac{1}{2}\left(\frac{\ell^2}{n}\right)(2n) + \text{smaller terms}\\
&= -\frac{\ell^2}{n} + \text{smaller terms.}
\end{aligned}$$

Thus, the approximation for the binomial coefficient is

$$\binom{2n}{n+\ell} \approx \frac{\sqrt{2}}{C\sqrt{n}} 2^{2n} e^{-\ell^2/n} + \text{smaller terms.}$$

Thus, this is largest when $\ell = 0$, which we already knew, but the nearby coefficients are almoat as large when ℓ^2/n is small (i.e. ℓ is on the scale of $\sqrt{n}$). This can be applied to the binomial distribution in probability, where one flips n coins and receives ℓ more heads than tails. Then, it's most likely that the number of heads and tails will be about the same (to about $\sqrt{n}$ difference). Thus, we can approximate something else:

$$\sum_{\ell=-n}^{n} \binom{2n}{n+\ell} = 2^{2n} \sim \frac{\sqrt{2}}{C\sqrt{n}} 2^{2n} \sum_{\ell=-n}^{n} e^{-\ell^2/n}.$$

Thus, we can approximate with an integral:

$$\sum_{\ell=-n}^{n} e^{-\ell^2/n} \approx \int_{-n}^{n} e^{-x^2/n}\,\mathrm{d}x \approx \int_{-\infty}^{\infty} e^{-x^2/n}\,\mathrm{d}x,$$

because when $x > n$, the value of the function is very small. Setting $y = x/\sqrt{n}$,

$$2^{2n} \sim \frac{\sqrt{2}}{C} 2^{2n} \int_{-\infty}^{\infty} e^{-y^2}\,\mathrm{d}y,$$

so $C = \sqrt{2}\int_{-\infty}^{\infty} e^{-y^2}\,\mathrm{d}y$. This integral is a common joke, if you're into that kind of humor, and it evaluates to $\sqrt{\pi}$, so $C = \sqrt{2\pi}$.

Moving to a different function, recall $p_k(n)$ is the number of ways to write $n = a_1 + \cdots + a_k$ such that $a_1 \ge a_2 \ge \cdots \ge a_k \ge 1$. We want to understand this when k is fixed and n is large. Since some of the a_i may be the same, then computing all of the permutations of them would be overcounting them, so $k!p_k(n) \ge \#\{n = a_1 + \cdots + a_k, a_i \ge 1\} = \binom{n-1}{k-1}$ (i.e. the number of compositions of n into k parts). Thus,

$$p_k(n) \ge \frac{1}{k!}\binom{n-1}{k-1} \approx \frac{(n-1)(n-2)\cdots(n-(k-1))}{k!(k-1)!} \approx \frac{n^{k-1}}{k!(k-1)!}$$

if k is fixed as $n \to \infty$. Notice that this approximation isn't valid if k also grows with n. Then, there is a nice trick to obtain the upper bound: suppose $n = a_1 + \cdots + a_k$, written in descending order, and let $b_i = a_i + (k-i)$. Then, $n + k(k-1)/2 = (a_1 + (k-1)) + \cdots + (a_k + (k-k)) = b_1 + \cdots + b_k$. In fact, every permutation of the b_i gives a set of a_i that works, so the number of such b_i provides an upper bound: $k!p_k(n) \le \#\{\text{compositions of } n + k(k-1)/2 \text{ into } k \text{ parts}\}$. Thus,

$$p_k(n) \le \frac{1}{k!}\binom{n + k(k-1)/2 - 1}{k-1}.$$

If k is fixed as $n \to \infty$, then this is approximately $n^{k-1}/(k!(k-1)!)$. Thus, we have shown the following theorem:

Theorem 18.1. $p_k(n) \sim \dfrac{1}{k!}\dfrac{n^{k-1}}{(k-1)!}$.

This means there are lots and lots of such partitions: it grows faster than any polynomial.

One can also use generting functions to analyze these: the convergence of a generating function yields information about the coefficients of a recurrence.

Example 18.1. Consider the Fibonacci numbers: $F_0 = 0$, $_1 = 1$, and $F_{n+2} = F_n + f_{n+1}$. The generating function is $f(x) = \sum_{n=0}^\infty F_n x^n$, so

$$\begin{aligned} xf(x) &= \sum_{n=0}^\infty F_n x^{n+1} = \sum_{n=1}^\infty F_{n-1}x^n, \text{ and} \\ x^2 f(x) &= \sum_{n=0}^\infty F_n x^{n+2} = \sum_{n=2}^\infty F_{n-2}x^n. \end{aligned}$$

Thus, $f(x) - xf(x) - x^2 f(x) = x$, using the recurrence $F_n = F_{n-1} = F_{n-2}$. Thus, one can solve to get $f(x) = x/(1-x-x^2)$. No discussion of convergence has been mentioned here; these are just formal power series. Using a partial fraction decomposition,

$$\begin{aligned} 1 - x - x^2 &= -\left(x - \frac{\sqrt{5}-1}{2}\right)\left(x + \frac{1+\sqrt{5}}{2}\right) \\ &= \left(\frac{\sqrt{5}+1}{2}\right)\left(1 + \frac{x}{(\sqrt{5}+1)/2}\right)\left(\frac{\sqrt{5}-1}{2}\right)\left(1 - \frac{x}{(\sqrt{5}-1)/2}\right) \\ &= \left(1 + \frac{x}{(\sqrt{5}+1)/2}\right)\left(1 + \frac{x}{(\sqrt{5}-1)/2}\right). \end{aligned}$$

Thus, we have

$$\frac{x}{1-x-x^2} = \frac{A}{1 + x(1+\sqrt{5})/2} + \frac{B}{1 - x(\sqrt{5}-1)/2},$$

with

$$\frac{1}{1 - x(1+\sqrt{5})/2} = \sum_{n=0}^\infty \left(\frac{1+\sqrt{5}}{2}\right)^n x^n, \text{ and } \frac{1}{1 - x(1-\sqrt{5})/2} = \sum_{n=0}^\infty \left(\frac{1-\sqrt{5}}{2}\right)^n x^n.$$

Thus,

$$F_n = A\left(\frac{\sqrt{5}-1}{2}\right)^n + B\left(\frac{\sqrt{5}+1}{2}\right)^n,$$

so asympotically, they are exponential, behaving like a constant times the exponential in the Golden Ratio. This is nice because it's not obvious from the recurrence.

Example 18.2. The generating function for $p(n)$ is $P(x) = \sum_{n=0}^\infty p(n)x^n$. The geometric series gives us that

$$\prod_{n=1}^\infty (1-x^n)^{-1} = \prod_{n=1}^\infty \sum_{j=0}^\infty x^{jn}.$$

In this infinite product, the coefficient of x^n is the number of ways to write n as a sum of other numbers, or $p(n)$.[22] This product $\prod(1-x^n)^{-1}$ converges when $\sum \log(1-x^n)^{-1}$ does, and after a little bit of work this works when $|x| < 1$.

The general technique here is for a generating function $\sum_{n=0}^\infty a(n)x^n = A(x)$, assume $a(n) \geq 0$, so that $a(n) \leq x^{-n}A(x)$. Then, the minimum value of $x^{-n}A(x)$, which is just a single-variable calculus problem, is a reasonable choice for an upper bound of the function. Thus, in this specific case, $p(n) \leq x^{-n}P(x)$ for all $0 < x < 1$, so we want $p(n) \leq \min_{0<x<1}(x^{-n}P(x))$. Thus, use logarithms:

$$\log(x^{-n}P(x)) = \log P(x) - n\log x = \sum_{k=1}^\infty \log(1-x^k)^{-1} - n\log x,$$

so differentiating,

$$\sum_{k=1}^\infty \frac{kx^{k-1}}{1-x^k} = \frac{n}{k}, \text{ so } n = \sum_{k=1}^\infty \frac{kx^k}{1-x^k}.^{23}$$

Of course, it is a great pain to figure out what this exactly means. One will need to solve for x_0 such that this is maximized, so the bound is obtained. This looked complicated, but of course that's because it is. A rough idea for x_0 is that it should go to 1, but how can this be quantified in terms of n? Crudely, if $x \to 1$, then $1 - x^k \to 0$, but a little more

[22] A lot of this may feel handwavy, especially in terms of convergence, but it's possible to work in the ring of formal power series over a field, in which the goal is to only consider formal expressions, and ignore questions of convergence. Then, two power series are the same if their coefficients are termwise the same. However, one could also argue convergence on some interval, which is an equally valid, but alternate, approach.

[23] At this point, you're supposed to figure out if it is a minimum or maximum, but we know here it must be a minimum.

precisely, $1 - x^k = (1-x)(1 + x + \cdots + x^{k-1}) \approx k(1-x)$. Thus, $n \approx x/(1-x)^2$, so $1 - x \approx 1/\sqrt{n}$. This calculation can be made more precise, of course, and once the details are worked out, then one sees that

$$p(n) < \frac{Ce^{\pi\sqrt{2n/3}}}{\sqrt{n}}.$$

The true bound is a beautiful theorem:

Theorem 18.2 (Hardy-Ramanujan)**.**

$$p(n) \sim \frac{e^{\pi\sqrt{2n/3}}}{2n\sqrt{3}}.$$