15-251: Great Theoretical Ideas in Computer Science 15251 Staff
Notes on Graph Theory (draft!!) January 29, 2012

1 Notation

A *graph* is a set of *vertices* or *nodes* V, and a set of *edges* E, where each edge is a pair of vertices. In an *undirected* graph an edge is just a set of two vertices $\{u, v\}$ (order does not matter), whereas in a *directed* graph an edge is an ordered pair (u, v) with the edge pointing from u to v. In this course we will mostly deal with undirected graphs. The graph G is thus the tuple (V, E).

> Some jargon alert: *You should get used to the phrase "Let $G = (V, E)$ be a graph…". We use "nodes" and "vertex" completely interchangeably. Sometimes we will even just say "Let G be a graph…", and you should imagine $G = (V, E)$ in there.*

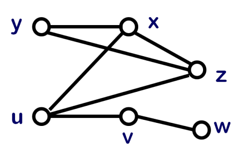
As an example: the graph $G = (V, E)$ with

$$V = \{u, v, w, x, y, z\}$$

and

$$E = \{\{u, x\}, \{u, v\}, \{v, w\}, \{x, y\}, \{x, z\}, \{y, z\}, \{u, z\}\}$$

is drawn below.



We also allow E to be a multiset of edges: the graph can have multiple edges connecting the same two nodes. These are called *parallel edges*. We also allow an edge to be a multiset of two copies of the same vertex $\{u, u\}$—such an edge is called a *self-loop*. But these are exceptional cases. A graph is *simple* if it does not have any self-loops, and no parallel edges.

> An important note: *We assume our graphs are simple, unless otherwise stated. So if you see a problem saying "Given a graph G, …", imagine the word "simple" in there.*

The edge $e = \{u, v\}$ is said to be *incident* to its two nodes u and v: these are also called its *endpoints*. A vertex u is *adjacent* to v is there is an edge $\{u, v\}$ "connecting them".

The *degree* of a node v is the number of edges incident on v, and is written as $\deg(v)$. (One exception: a self-loop $\{u, u\}$ contributes 2 to the degree of u.) Since each edge is incident on two nodes, it contributes to the degree of two nodes, and

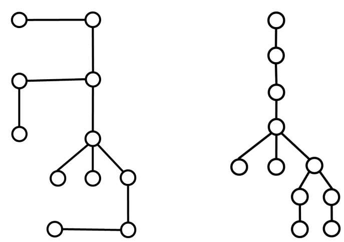
$$\sum_{v \in V} \deg(v) = 2|E|.$$

A *path* in G is a sequence of vertices $(v_1, v_2, \ldots, v_k)$ such that each consecutive pair $\{v_i, v_{i+1}\}$ is an edge in G. The path is *simple* if all the vertices are distinct. A *cycle* is a path where the start vertex v_1 is the same as the end vertex v_k, but all other vertices are distinct. For example, in the graph above: x, y, z, x, u, v, u is a path, but is not simple, whereas x, y, z, u, v is a simple path. And both x, y, z, x and x, y, z, u, x are cycles.

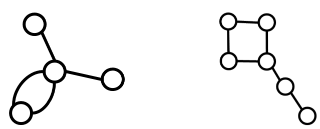
A graph $G = (V, E)$ is *connected* if every pair of nodes in G has a path between them. If the graph is not connected, each maximal connected piece is called a *component*.

2 Trees

A *tree* is a connected graph without any cycles (it is "acyclic"). The following two graphs are trees:



whereas the following two are not:



But there are many equivalent definitions of trees:

Theorem 2.1 *Let G be a graph on n nodes and e edges. The following statements are equivalent:*

1. *G is a tree (connected, acyclic)*
2. *Every two nodes in G are connected by a unique path*
3. *G is connected and $n = e + 1$*
4. *G is acyclic and $n = e + 1$*

Proofs like this often proceed by showing (1) $\Longrightarrow$ (2) $\Longrightarrow$ (3) $\Longrightarrow$ (4) $\Longrightarrow$ (1), and in this case we will do exactly this.

Proof: (1) $\Longrightarrow$ (2). Clearly, each pair u, v is connected by at least one path, since G is connected. Now suppose u and v have two different paths $P = (u = v_0 - v_1 - \ldots - v_k - v_{k+1} = v)$ and $Q = (u = w_0 - w_1 - \ldots - w_l - w_{l+1} = v)$. Let i be the smallest value such that $v_i \neq w_i$; hence $i \geq 1$. Let j be the smallest value $j > i$ such that v_p also lies on Q, say it is the same as w_q. Then the paths $w_{i-1} - w_i - \ldots - w_q$ and $v_{i-1} - v_i - \ldots - v_p$ are disjoint apart from the start and end, and form a cycle. This contradicts the acyclic property of G.

(2) $\implies$ (3). By induction. The base case is when $n = 2$, then the unique path implies a single edge, hence $n = 2 = e + 1$. Suppose the implication is true for all graphs with $< n$ nodes. Now suppose G has n nodes, let x, y be adjacent. Since there is a unique path between x, y there cannot be another path connecting x, y. If we delete the edge x, y the graph falls apart into two pieces G_1 and G_2. By induction $n_1 = e_1 + 1$ and $n_2 = e_2 + 1$, and so $n = n_1 + n_2 = e_1 + e_2 + 2 = e + 1$ (since $e = e_1 + e_2 + 1$, to account for the edge $\{x, y\}$).

(3) $\implies$ (4). By assumption, G is connected with $n = e + 1$. Suppose G has a cycle, with k vertices. Then this cycle also has k edges. Let S be this cycle. Since G is connected, there must be some vertex with an edge to S. Add this vertex to S along with this edge connecting it to S. Repeat until all nodes in G have been added to S. Since we add one edge for each vertex, we will end up with n vertices and n edges in S, which contradicts the fact that G has $n - 1$ edges.

(4) $\implies$ (1). Let G be acyclic and $n = e + 1$, and suppose G is not connected. Suppose there are k components. Each component is acyclic and connected, so if it has n_i nodes it must have $n_i - 1$ edges (from part (1) $\implies$ (2)). Hence $e = \sum_i e_i = \sum_i n_i - k$. So $k = 1$, which means the graph is connected. ■

These multiple definitions allow us to prove many facts about trees. For example, the following.

Fact 2.2 *A* leaf *is a vertex in a tree with degree* 1*. Any tree has at least two leaves.*

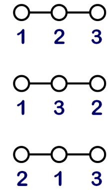
Proof: Suppose not. Then at least $n - 1$ nodes have degree at least 2. Also, since the graph is connected, the last node has degree at least one. So, $\sum_v \deg(v) \geq 2(n-1) + 1 = 2n - 1$, and hence $|E|$ is at least $n - 1/2$. But it is an integer, so $|E| \geq \lceil n - 1/2 \rceil = n$. But this is impossible for a tree. ■

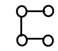
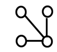
> **Exercise:** Give an example of a tree with only two leaves.
>
> **Exercise:** For a tree T, let S be the set of nodes with degree 1 or 2. Show that $|S| \geq n/2$.

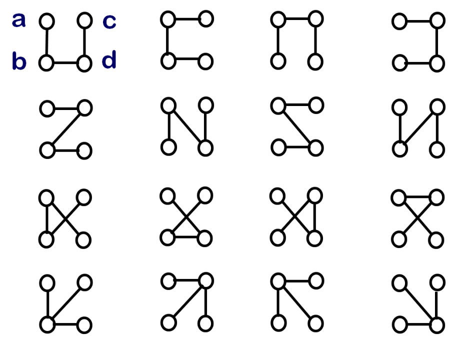
2.1 Counting Labeled Trees on n Nodes

In this section, we will count the number of labeled trees on n nodes, where the nodes are numbered $\{1, 2, \ldots, n\}$.

- There is only one tree on the two nodes $\{1, 2\}$: a single edge.
- On three nodes, each tree looks like a path with two edges: it comes down to choosing which node is the middle — this can be done in 3 ways. Hence $3 = 3^1$ trees.



- On four nodes, either the tree looks like a path with three edges like this: 
  or like a "3-leaf star", which has a vertex with three neighbors like this: 
  For the former, there are $4!/2 = 12$ ways to label it, and for the latter, there are 4 ways of choosing which node is the middle. Hence a total of $16 = 4^2$ trees.



The following theorem proves the pattern behind these examples:

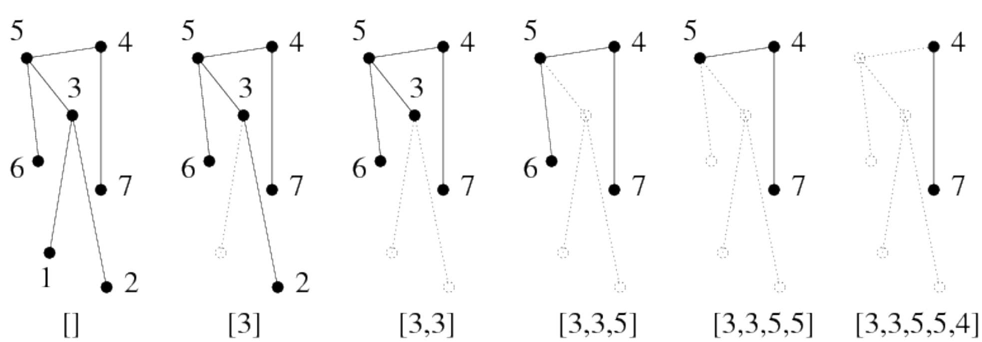
Theorem 2.3 (Cayley's Formula) *For each $n \geq 2$, there are n^{n-2} distinct labeled trees on n nodes.*

There are many proofs of this fact: in fact, the first proof was due to a guy called Borchardt in 1860, and even though Cayley's paper clearly cites Borchardt, it was Cayley's name that got attached to the formula. The proof we give here is due to Heinz Prüfer.

Proof: The proof will show a bijection between the set of labeled trees on n vertices, and the set $\{1, 2, \ldots, n\}^{n-2}$ consisting of sequences of length $n-2$ where each number is between 1 and n.

Given a tree T, you generate its Prüfer code thus: repeatedly find the leaf with the lowest label, delete this leaf and the incident edge, and and output the label of its unique neighbor. Stop when only two vertices remain.

Note that when $n = 2$, there is a single tree, and it is encoded by the empty sequence. For $n = 3$, each tree is a path, and the encoding just outputs the name of the center vertex of this path. Here is an example on a 7-node graph:



This shows a way to get from a tree to a sequence of $n-2$ numbers from the set $\{1, 2, \ldots, n\}$. There are n^{n-2} of them. Note that the name of each non-leaf node in the tree T is output at least once in this sequence, and the vertices not appearing in the sequence are precisely the leaves of the original tree T.

Now, to go the other way from the sequence $(a_1, a_2, \ldots, a_{n-2})$ to a tree, this suggests the following algorithm. Start with all the numbers in $\{1, 2, \ldots, n\}$ that do *not* appear in the sequence. These are the leaves of the tree. Call this set L. Now remove the least numbered node l in L, and add an edge from this node l to the node numbered a_1. Remove a_1 from the sequence, and l from the set L. If the number a_1 does not appear as any other a_j, add a_1 to L. And repeat. If the sequence becomes empty, there are two nodes in L: connect them by an edge.

Hence, for $n = 7$ and the sequence $[3, 3, 5, 5, 4]$ we got above,

- We start off with the set of leaves $L = \{1, 2, 6, 7\}$.
- Now we add an edge $(1, 3)$, L becomes $\{2, 6, 7\}$ (since 1 is deleted) and the sequence becomes $[3, 5, 5, 4]$ (since we remove 3 from the front of the sequence).
- Next we add edge $(2, 3)$, the sequence becomes $[5, 5, 4]$, and 2 is deleted but 3 is added to L to get $\{3, 6, 7\}$.
- Add edge $(3, 5)$, sequence becomes $[5, 4]$, L is $\{6, 7\}$.
- Add edge $(6, 5)$, sequence becomes $[4]$, L is now $\{5, 7\}$.
- Add edge $(5, 4)$, sequence becomes $[]$, L is now $\{4, 7\}$.
- Add the last edge $(4, 7)$.

Note that the edges we created above are precisely the ones we deleted in going from T to the code, and in the same order.

To prove this, we need an inductive argument formalizing precisely this, which shows that we get a bijection. We'll omit the proof for now: for this iteration of the course, you should know how to move from a tree to its Prüfer code, and back. ■

Exercise: If the Prüfer code for a tree consists of a single number repeated n times, prove that the tree must be a star. (A star is a tree with a single node with degree $n - 1$, and all other nodes having degree 1.)

Exercise: If the Prüfer code for a tree has no repeated numbers, then show that the tree must be a path.

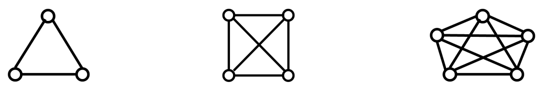
Exercise*: The number of labeled trees on n vertices such that vertex i has degree d_i (with $\sum_i d_i = 2n - 2$)) is given by the multinomial:

$$\binom{n-2}{d_1 - 1 \; ; \; d_2 - 1 \; ; \; \ldots \; ; \; d_n - 1}$$

2.2 Spanning Trees

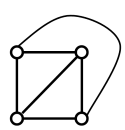
Given a graph $G = (V, E_G)$, a spanning tree T of G is a connected acyclic subgraph of G with the same vertex set V. I.e., $T = (V, E_T)$ for some subset of edges $E_T \subseteq E_G$, and is a tree.

Note that if G is not itself connected, it has no spanning trees. Else it has at least one spanning tree. To see this, just take the connected graph, and if dropping some edge from it does not cause the graph to become disconnected, drop it and continue. What remains must be a spanning tree.

> **Exercise:** Let K_n denote the "complete graph" on n vertices labeled $\{1, 2, \ldots, n\}$: for every pair of distinct vertices, it has an edge connecting that pair. (Hence there are $\binom{n}{2}$ edges in K_n.) E.g., here are the graphs K_3, K_4, K_5:
>
> 
>
> How many distinct spanning trees does K_n have?

3 Planar Graphs

A graph is *planar* if it can be drawn in the plane without any crossing edges. (Such a drawing is called a planar drawing.) E.g., the graph K_4 is planar, since it admits the following planar (i.e., non-crossing) drawing:



Note that any such planar drawing splits the infinite 2-dimensional plane into disjoint areas, and these will be called *faces*. E.g. the drawing above creates 4 faces. (Remember to count the outside, infinite face!)

3.1 Euler's Formula

Euler's formula says that if a connected planar graph has n vertices, e edges and f faces, then $n - e + f = 2$. In fact, we don't even need the graph to be simple for this formula: it could have parallel edges. Let's check this on two graphs: for K_3 we have $n = 3$ nodes, $e = 3$ edges, and $f = 2$ faces. And for K_4, we have $n = 4$, $e = 6$ edges and $f = 4$ faces. Both of which check out: perhaps Euler is correct after all. And indeed he is: here's the proof.

Theorem 3.1 (Euler's Formula) *If a connected (arbitrary) planar graph has n vertices, e edges and f faces, then*

$$n - e + f = 2.$$

Proof: The proof is by induction. Let's build up the graph by adding edges one at a time, always preserving the Euler formula.

Start with a single edge and 2 vertices. Note that $n = 2, f = 1, e = 1$. (Check!) Add the edges in some order so that what we've added so far is connected. There are two cases to consider.

- The edge connects two vertices already there in the graph. In this case we add a new edge, and also split an existing face in two. Hence `e++` and `f++`, so $n-e+f$ is preserved, and we're OK.

- The edge connects the current graph to a new vertex. Now `n++` (since we added a new vertex) and `e++`. Moreover, since the new vertex is degree-1, it does not create a new face, so again $n - e + f$ is preserved.

Since we started off with $n - e + f = 2$ and maintained that over the course of the process, we get that at the end. ■

Given Euler's formula, we get a number of cool results as simple corollaries:

Theorem 3.2 *A (simple) planar graph on n nodes has at most $3n - 6$ edges, as long as $n \geq 3$.*

Proof: Let G be a simple planar graph with n nodes, e edges, and f faces, and let $n \geq 3$. Suppose G is not connected, then we can add in edges maintaining planarity until G is connected: this only gives us more edges.

Also, suppose there is an edge that has the same face on both sides of it. (This means deleting this edge will disconnect the graph into two components A and $V \setminus A$.) Then since $n \geq 3$, at least one of these components has anoter vertex apart from the endpoints of e, and hence there must be at least one edge we can add that maintains planarity. Adding this new edge gives us another planar graph with even more edges, so it suffices to prove the theorem for graphs where every edge has two different faces on either side of it.

Every face of G has at least three edges around it, since we have at least three vertices and G is simple and connected. (A face having a single edge would be a self-loop, and a face having two edges around it can only be made by parallel edges. And each edge lies on exactly two faces, so $2e \geq 3f$. Applying this inequality to Euler's formula gives

$$\begin{aligned} n - e + f &= 2 \\ \Rightarrow \quad 3n - 3e + 3f &= 6 \\ \Rightarrow \quad 3n - 3e + 2e &\geq 6 \\ \Rightarrow \quad 3n - 6 &\geq e \end{aligned}$$

And we're done.

This theorem is very useful: it shows that the number of edges in a planar (simple) graph can never be too many more than the number of vertices in it. Also note that while Euler's formula works even for multigraphs, this works only for simple graphs. ■

> **Exercise:** Suppose G is a connected simple planar graph that has no cycle of length 3, and $n \geq 3$. Modify the proof above to show that the number edges in G is at most $2n - 4$.

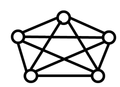
Corollary 3.3 *Any (simple) planar graph contains a vertex with degree at most 5.*

Proof: Let G be a simple planar graph with n nodes, e edges, and f faces. The average degree of G is $\sum_v \deg(v)/n = 2e/n$. By Theorem 3.2, $e \leq 3n - 6$, so the average degree of G is at most $2(3n-6)/n = 6 - 12/n < 6$. If every vertex in G had degree at least 6, then the average degree of G would be at least 6. Therefore G must contain a vertex of degree at most 5. ■

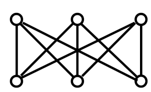
3.2 Characterizing Planar Graphs

Given a graph G, a *minor* of G is a graph you can obtain by deleting some edges of G, and then contracting some other edges. (If you contract an edge, you merge the two nodes that are its endpoints.) E.g., the graph on the left is a minor of the graph on the right if you delete the blue edge, and contract the edges incident to the red nodes.

You've already seen the graph K_5 with $n = 5$ and $e = \binom{5}{2} = 10$.



Let's introduce the graph $K_{3,3}$. In this graph each of the top three vertices are connected to all of the bottom vertices, giving $n = 6$ and $e = 9$.



Theorem 3.4 (Kuratowski's Theorem) *A graph is planar if and only if it does not contain K_5 or $K_{3,3}$ as a minor.*

The proof that if a graph is planar, then it must contain either $K_{3,3}$ or K_5, is the hard direction of the proof. (We won't prove it here.) But the other direction is simple: as the following exercise shows, $K_{3,3}$ and K_5 are not planar, and hence neither is any graph that contains one of these as a minor.

> **Exercise:** Use Theorem 3.2 to show that K_5 is not planar. Can you do the same for $K_{3,3}$? Show that $K_{3,3}$ is not planar using a similar proof based on Euler's formula.

3.3 Colorings in Planar Graphs

We often study *colorings* of graphs: these are maps that assign colors to the vertices of the graph. We say a coloring is *proper* if no two vertices that are connected by an edge have the same color. In other words, the coloring does not color the endpoints of any edge the same. [1]

[1] The study of *edge-colorings* of graphs is also an interesting topic: here, we assign colors to the edges, such that no two edges incident to the same vertex have the same color. However, when we talk about graph colorings, it's usually vertex colorings.

A graph G is k-colorable if there exists a proper coloring for G that uses at most k colors. Planar graphs happen to be graphs which can be colored using very few colors.

Theorem 3.5 (The Six Color Theorem) *Any planar graph is* 6*-colorable.*

Proof: The proof is by induction. For the base case, the graph with a single vertex is clearly 1-colorable.

Now let $G = (V, E)$ be a planar graph with n nodes. By Corollary 3.3, it has a vertex v with degree at most 5. Remove v and all edges incident on it to get H. This is another planar graph, and hence can be colored using at most 6 colors. Now add back v and the edges. Since v had at most 5 neighbors, at most 5 colors are used by its neighbors. So we can color v using the remaining (sixth) color. ■

A similar proof can be used to show that any graph with maximum degree Δ can be $(\Delta+1)$-colored, regardless of whether it is planar or not.

But this is not the best we can do for planar graphs:

Theorem 3.6 (The Five Color Theorem) *Any planar graph is* 5*-colorable.*

Proof: Start with a planar embedding of G. Use Corollary 3.3 to find a vertex v with degree at most 5. If v's degree was 4, then we can inductively color the graph $G - \{v\}$, and use the fifth color to color v. So suppose v has degree 5.

Now remove v and all its incident edges—in the planar map, this will create a region bounded by the five neighbors of v. If all the $\binom{5}{2}$ possible edges between these five neighbors exist, we have found a K_5 in the graph, which is not planar, and hence gives us a contradiction. So there exist two neighbors u, w of v that are not connected by an edge.

Figure here

Create a new graph H by merging together u and w as shown in the figure. Note that H is also planar! So inductively find a coloring for H, unmerge u and w—giving them the same color is OK since they don't have an edge between them. But you've used only 4 colors, since u and w have the same color, so use the fifth color for v. This completes the induction. ■

And you can do even better!

Theorem 3.7 (The Four Color Theorem) *Any planar graph is* 4*-colorable.*

A proof of theorem had been claimed many times since 1852, but the first correct proof was only given by Kenneth Appel and Wolfgang Haken in 1976. The proof was computer-assisted, in that 1,936 special graphs had to be checked for a certain property. While recent proofs have reduced the number of cases to be checked to about 600, and a proof using an automated theorem prover has been given, no "simple" proofs are known. There are several good online articles on the subject that give a sense of the techniques used in these computer-assisted proofs.

Can we do even better? Can all planar graphs be 3-colored? No!

> **Exercise:** Give an example of a planar graph that requires 4 colors to be properly colored.

4 Adjacency Matrix and Lists

The *adjacency matrix* A for a graph $G = (V, E)$ is defined as follows:

$$A_{ij} = \begin{cases} 0 & \text{if } (i,j) \notin E, \\ 1 & \text{if } (i,j) \in E. \end{cases}$$

example here

Theorem 4.1 *The number of paths of length k between i and j is given by $(A^k)_{ij}$.*

The proof is a simple induction; we omit it here.