Chapter 16

Trees

16.1 Basic properties of trees

Trees in mathematics are graphs of a certain kind. In a sense, trees are the simplest interesting graphs, in that they have a very simple structure, but possess a rich variety of nontrivial properties. Trees have innumerable applications throughout computer science.

Definition 16.1.1. *A* tree *is a connected graph with no cycles. A vertex of degree one in a tree is called a* leaf*.*

An extensive theory of trees has been developed, and we give the tip of the iceberg below: Four additional characterizations that each could have been used to define trees.

Theorem 16.1.2. *Given a graph $G = (V, E)$, the following conditions are equivalent:*

(a) G is a connected graph with no cycles. (Thus G is a tree by the above definition.)

(b) For every two vertices $u, v \in V$, there exists exactly one path from u to v.

(c) G is connected, and removing any edge from G disconnects it. (Thus G is a minimal connected graph*.)*

(d) G has no cycles, and adding any edge to G gives rise to a cycle. (Thus G is a maximal acyclic graph*.)*

(e) G is connected and $|E| = |V| - 1$.

Proof. We will prove for each of the conditions (b)–(e) in turn that it is equivalent to condition (a). This implies the equivalence of all the conditions. The proof proceeds by induction on the number of vertices $|V|$ in G, and we relate a tree with $n + 1$ vertices to a tree with n vertices in the inductive step by "tearing off" a leaf. We begin by proving two lemmas that will be useful in this process.

Lemma 16.1.3. *Each tree with at least 2 vertices contains at least 2 leaves.*

Proof. Given a tree $T = (V, E)$, consider a path P of maximum length in T. We claim that the two end-points of P are leaves of T. Indeed, assume for the sake of contradiction that an end-vertex u of P has degree greater than 1 in T. Thus there exists an edge $\{u, u'\} \in E$ that is not part of P. If u' belongs to P then T contains a cycle. Otherwise we can extend P by the edge $\{u, u'\}$ and P is not a longest path in T. This contradiction proves the lemma. □

Lemma 16.1.4. *Given a graph $G = (V, E)$ and a leaf $v \in V$ that is incident to an edge $e = \{v, v'\} \in E$, the graph G is a tree if and only if $G' = (V \setminus \{v\}, E \setminus \{e\})$ is a tree.*

Proof. Assume that G is a tree and consider two vertices $u, w \in V \setminus \{v\}$. u and w are connected by a path P in G. Every vertex of P other than u and w has degree at least 2, and thus v cannot be a vertex of P. Therefore P is a path in G', which proves that G' is connected. Since G does not contain a cycle, G' cannot contain a cycle and is thus a tree.

For the other direction, assume that G' is a tree. Since a cycle only contains vertices with degree at least 2, a cycle in G must also be a cycle in G'. Therefore there are no cycles in G. Also, any two vertices of G other than v can be connected by the same path as in G', and v can be connected to any vertex u in G by a path that consists of the edge e and a path in G' between v' and u. Thus G is a tree. □

We are now ready to employ induction to prove that condition (a) implies each of (b)–(e). For the induction basis, all five conditions hold for the graph with a single vertex. Consider a graph $G = (V, E)$ with $|V| = n \geq 2$ and assume that (a) holds for G. By Lemma 16.1.3, G has a leaf $v \in V$ that is incident to an edge $e = \{v, v'\} \in E$. By Lemma 16.1.4, condition (a) holds for G'. The inductive hypothesis states that condition (a) implies conditions (b)–(e) for the graph $G' = (V \setminus \{v\}, E \setminus \{e\})$. We now need to prove that conditions (b)–(e) also hold for G.

Condition (b) holds for G by a similar argument to the one employed in the proof of Lemma 16.1.4. Condition (c) holds for G since removing any edge other than e disconnects G by the induction hypothesis, and removing e disconnects the vertex v from the rest of the graph. Condition (d) holds since G cannot have cycles by an argument similar to the proof of Lemma 16.1.4; adding an edge that is not incident to v creates a cycle by the inductive hypothesis, and adding an edge $\{v, u\}$, for some $u \in V \setminus \{v\}$ creates a cycle that consists of the edge $\{v, u\}$, the path from u to v', and the edge e. Finally, condition (e) holds since G is obtained from G' by adding one vertex and one edge.

We now prove that each of (b)–(d) imply (a). Conditions (b) and (c) on G each imply connectedness of G. By contrapositive, assume that G contains a cycle. Then taking two distinct vertices u, w on the cycle, there are two paths from u to w along the cycle, which implies $(b) \Rightarrow (a)$. Furthermore, removing one edge of the cycle does not disconnect G, which implies $(c) \Rightarrow (a)$. Condition (d) implies that G does not contain a cycle. By contrapositive, assume that G is disconnected. Then there are two vertices u and w that have no path connecting them and we can add the edge $\{u, w\}$ to G without creating a cycle. This implies $(d) \Rightarrow (a)$.

To prove $(e) \Rightarrow (a)$ we use induction on the number of vertices of G. The induction basis is the graph with one vertex and the claim trivially holds. For the induction hypothesis, assume that the claim holds for all graphs with $|V| - 1$ vertices. For the inductive step, assume that condition (e) holds for G and hence $|E| = |V| - 1$. Therefore the sum of the degrees of the vertices of G is $2|V| - 2$, and thus there is some vertex $v \in V$ of degree 1. The graph $G' = (V \setminus \{v\}, E \setminus \{e\})$ is connected and satisfies $|E \setminus \{e\}| = |V \setminus \{v\}| - 1$. By the induction hypothesis, G' is a tree. Lemma 16.1.4 now implies that G is a tree, which completes the proof. $\square$

16.2 Spanning trees

One of the reasons that trees are so pervasive is that every connected graph G contains a subgraph that is a tree on all of the vertices of G. Such a subgraph is called a *spanning tree* of G.

Definition 16.2.1. *Consider a graph $G = (V, E)$. A tree of the form (V, E'), where $E' \subseteq E$ is called a spanning tree of G.*

Proposition 16.2.2. *Every connected graph $G = (V, E)$ contains a spanning tree.*

Proof. Consider a tree subgraph $T = (V', E')$ of G with the largest number of vertices. Suppose for the sake of contradiction that $V' \neq V$, and thus there exists $v \in V \setminus V'$. Take an arbitrary vertex $u \in V'$ and consider a path P between v and u. Let u' be the first vertex along P that belongs to V', and let v' be the vertex that immediately precedes u' in P. Consider the graph $T' = (V' \cup \{v'\}, E' \cup \{v', u'\})$. Lemma 16.1.4 implies that T' is a tree in G with a greater number of vertices than T, which is a contradiction. $\square$

Chapter 17

Planar Graphs

17.1 Drawing graphs in the plane

As we have seen in class, graphs are often visualized by drawing them in the plane—vertices are drawn as points, and edges as curved segments (called *arcs*) that connect the corresponding points. A graph together with a drawing of it in the plane is called a *topological graph*.

A graph is called *planar* if there exists a drawing of it in which the interior of any arc does not touch or intersect any other arc. That is, two distinct arcs are either disjoint or touch at endpoints that they share. A planar graph together with a planar drawing of it is called a *plane graph*.

It is easy to verify that paths, cycles and trees of any size are planar. Transportation networks often provide examples of planar graphs, and graph planarity became important in computer science due to a connection with VLSI circuit design. Planar drawings are often considered superior when visualizing graphs, as they have no edge crossings that can be mistaken for vertices. In fact, a whole subfield of computer science called *graph drawing* is devoted to the study of various kinds of drawings of graphs.

It might not be obvious at first that there are any nonplanar graphs at all. There are, but we'll have to do some work to prove this, and we'll need two preliminary steps just to approach this issue. The first is to define the *faces* of a plane graph and the second is to mention the (in)famous Jordan curve theorem.

Let us begin with faces. Define an equivalence relation on the plane as follows: Two points $a, b \in \mathbb{R}^2$ are equivalent if they can be connected by an arc that does not intersect the edges of a given plane graph G. Then the set of all points that belong to a particular equivalence class of this relation are said to be a *face* of G. Intuitively, if we draw G on a white sheet of paper with a black pencil, the faces are the white regions; alternatively, if we cut the paper along the edges of the drawing, the faces are the resulting pieces. Note that faces are defined for plane graphs, but not for planar graphs without a drawing: Different drawings of the same graph can produce different sets of faces!

The second piece of mathematical equipment we'll need to study planar graphs is the Jordan curve theorem.[1] It is a classical example of a mathematical statement that is intuitively obvious, but exceedingly difficult to prove. (Related specimens that arguably fall into this category are Kepler's conjecture and the Kneser-Poulsen conjecture.)

Theorem 17.1.1 (Jordan curve theorem)**.** *Every closed non-self-intersecting curve γ in the plane separates the plane into exactly two regions, one bounded and one unbounded, such that γ is the boundary of both. Alternatively, a plane drawing of any cycle C_i, for $i \geq 3$, has exactly two faces.*

To see why the Jordan curve theorem is not so easy to prove recall that there are some crazy curves out there—just think about fractals like the *Koch snowflake*. How would you go about proving that such monsters have "interior" and "exterior"?

The following corollary follows from the Jordan curve theorem by routine arguments.

Corollary 17.1.2. *Consider a plane graph G and an edge e that is part of a cycle in G. Then e lies on the boundary of exactly two faces of G.*

17.2 Euler's formula

The fundamental tool in the study of planar graphs is Euler's formula, presented by Euler in 1752.[2]

Theorem 17.2.1 (Euler's formula)**.** *Let G be a connected plane graph with n vertices, e edges, and f faces. Then*

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

Note that the theorem need not hold if the graph is not connected—Just think of a collection of isolated vertices. On the other hand, the formula remains true even for (non-simple) graphs with multiple edges and self-loops.

Proof. The proof proceeds by induction on the number of edges. If there are none, the graph consists of a single vertex, the drawing has one face, and the formula holds as $1 - 0 + 1 = 2$. Assume that the formula holds for all plane graphs having k edges. Consider a plane graph $G = (V, E)$ with n vertices, f faces, and $k + 1$ edges. We distinguish between two cases:

G is a tree. In this case $n = k + 2$, due to a tree characterization we have seen previously, and $f = 1$ since any planar drawing of a tree has exactly one face. Then the formula holds as $(k + 2) - (k + 1) + 1 = 2$.

[1] Jordan gets all the press even though his proof of the theorem was wrong, and it took almost 20 years until Veblen found a correct one in 1905.

[2] Caution: Due to Euler's prodigious output, there are multiple "Euler's formulae", "Euler's theorems", etc.

G **has a cycle** C**.** Take an edge e that lies on C and consider a plane graph $G' = (V, E \setminus \{e\})$, whose vertices and edges are drawn as in G. By Corollary 17.1.2, the edge e is adjacent to two faces of G, and these faces "merge" into one in G'. Thus G' has n vertices, $f - 1$ faces, and k edges. By the induction hypothesis, $n - k + (f - 1) = 2$, hence $n - (k + 1) + f = 2$.

This completes the proof by induction. □

Euler's formula implies that the number of faces of a plane graph does not depend on the drawing, so even though the faces themselves are only defined for a particular drawing, their number is fixed *a priori* for any planar graph! The formula has a number of other consequences that are frequently used in theoretical computer science. These consequences say that planar graphs have few edges, and always have at least one low-degree vertex. As they make abundantly clear, not only are not all graphs planar, but *most* graphs aren't. (Do you understand the sense in which the theorem below implies this?)

Theorem 17.2.2. *For any simple planar graph G with n vertices and e edges:*

(a) If $n \geq 3$ then $e \leq 3n - 6$. If $e = 3n - 6$ then every face of G is a 3-cycle (a "triangle") and G is called a triangulation*.*

(b) There is a vertex of G that has degree at most 5.

Proof. The proofs of the two parts are similar in their clever use of Euler's formula:

(a) If G is not connected, we can add edges to connect G while maintaining its planarity. Assume therefore that G is connected. Let f be the number of faces of G. For such a face t, let $\alpha(t)$ be the number of edges adjacent to t and consider the sum $\sum_t \alpha(t)$ that ranges over all faces t of G. As each edge is adjacent to at most two faces, a particular edge is counted at most twice in the above sum. Hence
$$\sum_t \alpha(t) \leq 2e.$$
On the other hand, each face has at least three edges on its boundary, so
$$\sum_t \alpha(t) \geq 3f.$$
We get $3f \leq 2e$, and, using Euler's formula, $3(2 - n + e) \leq 2e$ and
$$e \leq 3n - 6.$$
Finally, if $e = 3n - 6$ then $3f = 2e$ and it must be that every face has exactly three edges on its boundary.

(b) If the graph is disconnected we consider one particular connected component of it, so assume that G is connected. If G has two vertices or less the result is immediate, so assume that $n \geq 3$. Recall that $d_G(x)$ denotes the degree of a vertex x in G. The sum $\sum_x d_G(x)$, ranging over the vertices x of G, counts every edge twice, so

$$\sum_x d_G(x) = 2e.$$

As we have seen, $e \leq 3n - 6$, so

$$\sum_x d_G(x) \leq 6n - 12.$$

If the degree of every vertex is at least 6, we get

$$6n \leq 6n - 12,$$

which is a contradiction. Therefore, there must be a vertex with degree at most 5.

□

An intuitive way to think about Theorem 17.2.2(a) is that once a graph has too many edges, there is no more room for them in the plane and they start intersecting. This gives a way to prove that a particular graph is not planar. Take K_5, for example. It has 5 vertices and 10 edges, and $10 > 3 \cdot 5 - 6$. Thus K_5 is not planar! In fact, no K_n is planar for $n \geq 5$, since they all contain K_5 as a subgraph. On the other hand, K_n is planar for $n \leq 4$, as can be demonstrated by their simple planar drawings. This illustrates a point that might be obvious by now: proving a graph to be planar is often easier than proving the opposite. (Just draw it!™)

How about complete bipartite graphs? It is easy to verify that $K_{i,j}$ is planar when $i \leq 2$ or $j \leq 2$. The smallest remaining suspect is $K_{3,3}$. Playing around with drawings doesn't help: There seems to be no way to draw $K_{3,3}$ without intersections. Let's try the trick that worked for K_5: The number of vertices of $K_{3,3}$ is 6, its number of edges is 9, and $9 \leq 3 \cdot 6 - 6$. No luck. We need a stronger tool, and here it is:

Proposition 17.2.3. *For any simple planar graph G with n vertices and e edges, if G does not contain a cycle of length 3 then $e \leq 2n - 4$.*

Proof. We can assume that G is connected as in Theorem 17.2.2. Let f be the number of faces of G and let $\alpha(t)$ be the number of edges adjacent to a face t. These edges make up a cycle in G, and thus their number is at least 4, implying $\alpha(t) \geq 4$. Consider the sum $\sum_t \alpha(t)$, over all faces t of G. Each edge is adjacent to at most two faces, thus

$$4f \leq \sum_t \alpha(t) \leq 2e.$$

Using Euler's formula, we get $4(2 - n + e) \leq 2e$ and $e \leq 2n - 4$. □

With this result we're finally in business: $K_{3,3}$ does not contain an odd cycle since it is bipartite, thus every cycle in the graph has length at least 4. Since $9 > 2 \cdot 6 - 4$, $K_{3,3}$ is not planar. Let's summarize what we've learned.

Theorem 17.2.4. *K_n is planar if and only if $n \leq 4$ and $K_{i,j}$ is planar if and only if $i \leq 2$ or $j \leq 2$.*

At this point we have laid the groundwork for one of the most striking results concerning planar graphs, known as Kuratowski's theorem. To state it we need the following definition:

Definition 17.2.5. *Given a graph $G = (V, E)$, an* edge subdivision *operation on an edge $\{u, v\}$ of G results in the graph $(V \cup \{x\}, (E \setminus \{\{u, v\}\}) \cup \{\{u, x\}, \{x, v\}\})$, where $x \notin V$ is a new vertex. A graph G' is said to be a* subdivision *of G if it can be obtained from G by successive applications of edge subdivision.*

Kuratowski's theorem says that not only are K_5 and $K_{3,3}$ non-planar, but every non-planar graph contains either a subdivision of K_5 or a subdivision of $K_{3,3}$. That is, the graphs K_5 and $K_{3,3}$ characterize the whole family of non-planar graphs!

Theorem 17.2.6 (Kuratowski's theorem)**.** *A graph is planar if and only if it does not contain a subdivision of K_5 or a subdivision of $K_{3,3}$ as a subgraph.*

17.3 Coloring planar graphs

You might have heard of the four-color problem. It was posed in the mid-19th century and occupied some of the best discrete mathematicians since that time. The original formulation is in terms of political maps. In such maps, neighboring countries are drawn with different colors. The question is how many colors are needed. It is easy to construct simple examples of maps that need at least four colors. The four color problem asks whether four colors always suffice, for any political map. (We require that every country is *connected*, unlike, say, the US.)

This problem is equivalent to whether every planar graph can be colored with four colors. (To see this, construct a graph whose vertices correspond to countries and whose edges connect neighbors through border segments.) It took over a century until Appel and Haken found a proof that four colors always suffice, and even that was possible only by using computers to conduct extensive case enumeration and analysis. To this date no proof of the four color theorem is known that does not rely on computers. On the other hand, in 1890 Heawood discovered a beautiful proof that *five* colors always suffice. To prepare for his proof, let us warm up by showing that every planar graph can be colored with 6 colors. The proof is surprisingly simple.

Theorem 17.3.1. *The chromatic number of a planar graph G is at most six.*

Proof. By induction on the number n of vertices of G. If $n \leq 6$ the claim is trivial. Assume every planar graph with at most k vertices can be colored with 6 colors or less, and consider a graph $G = (V, E)$ with $k + 1$ vertices. By Theorem 17.2.2(b), there is a vertex v of G with degree at most 5. Let G' be the induced subgraph of G on the vertices $V \setminus \{v\}$. By the induction hypothesis, G' can be colored with five colors or less. Color the vertices $V \setminus \{v\}$ of G with the colors that they are assigned in the coloring of G'. Assign to v the color that is not used by its neighbors. Since the degree of v is at most five, such a color exists. This specifies a valid coloring of G. □

We are now ready for Heawood's five color theorem.

Theorem 17.3.2. *The chromatic number of a planar graph $G = (V, E)$ is at most five.*

Proof. The proof proceeds by induction on the number n of vertices of G. The base case is trivial. Assume every planar graph with at most k vertices can be colored with 5 colors or less, and consider a graph $G = (V, E)$ with $k+1$ vertices. Let v be a vertex of G with degree at most 5. If $d_G(v) < 5$ we can produce a 5-coloring of G as in the proof of Theorem 17.3.1. Assume $d_G(v) = 5$ and let $c : (V \setminus \{v\}) \to \{1, 2, 3, 4, 5\}$ be a 5-coloring of the induced subgraph G' of G on the vertices $V \setminus \{v\}$. This coloring exists by the induction hypothesis.

We consider a particular drawing of G in the plane and henceforth regard G as a plane graph. Let v_1, v_2, v_3, v_4, v_5 be the neighbors of v in the order they appear around v in G. (That is, according to one of the circular orders in which the corresponding edges emanate from v in G.) Without loss of generality, assume that $c(v_i) = i$ for $1 \leq i \leq 5$. (Note that if some color is unused by v_1, v_2, v_3, v_4, v_5, we can simply assign that color to v.) We distinguish between two cases: Either there does not exist a path between v_1 and v_3 in G that uses only vertices of colors 1 and 3, or there does.

There is no such path. In this case consider the subgraph G'' of G that is the union of all paths that begin at v_1 and use only vertices with colors 1 and 3. Note that neither v_3 nor its neighbors belong to G''. We produce a 5-coloring of G as follows: All the vertices of G'' of color 1 are assigned the color 3, all the vertices of G'' of color 3 are assigned the color 1, the vertex v is assigned the color 1, and all other vertices of G keep the color assigned by c. No monochromatic edges are created by this assignment and the coloring is valid.

There is such a path. Consider a path P from v_1 to v_3 that uses only vertices with colors 1 and 3. Together with the edges $\{v, v_1\}$ and $\{v, v_3\}$ this forms a cycle. The vertices v_2 and v_4 lie on different sides of this cycle. (Here we use the Jordan curve theorem.) Therefore there is no path between v_2 and v_4 that uses only vertices with colors 2 and 4, and we can apply the reasoning of the previous case.

□

17.4 Concluding remarks

There are many more amazing results associated with planar graphs. Two of the most striking are Fáry's theorem and Koebe's theorem. Fáry's theorem states that every planar graph can be drawn in the plane without edge crossings, such that all the arcs are *straight line segments*. Koebe's theorem says that every planar graph is in fact isomorphic to an "incidence graph" of a collection of nonoverlapping discs in the plane. (The vertices of this graph correspond to the discs, and two vertices are adjacent if and only if the corresponding disks are tangent.) Fáry's theorem is an immediate consequence of Koebe's theorem, although they were discovered independently. Both are remarkable.