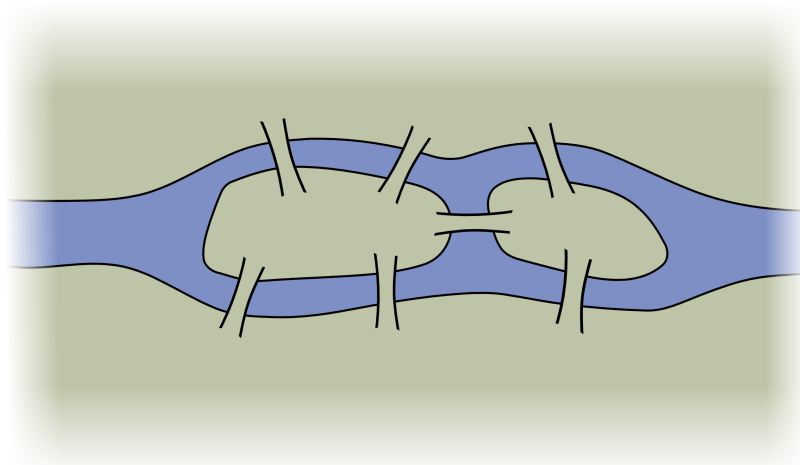


# Introduction to Graph Theory

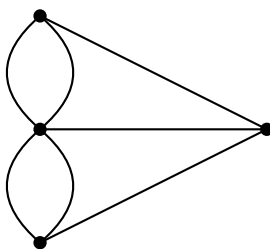
Graph Theory is a relatively new area of mathematics, first studied by the super famous mathematician Leonhard Euler in 1735. Since then it has blossomed in to a powerful tool used in nearly every branch of science and is currently an active area of mathematics research. We will begin our study with the problem that started it all: The Seven Bridges of Königsberg.

In the time of Euler, in the town of Königsberg in Prussia, there was a river containing two islands. The islands were connected to the banks of the river by seven bridges (as seen below). The bridges were very beautiful, and on their days off, townspeople would spend time walking over the bridges. As time passed, a question arose: was it possible to plan a walk so that you cross each bridge once and only once? Euler was able to answer this question. Are you?



Try finding a path which uses each bridge exactly once.

Here is another problem: below is a drawing of four dots connected by some lines. Is it possible to trace over each line once and only once (without lifting up your pencil)? You must start and end on one of the dots.



There is an obvious connection between these two problems - any path through the dot and line drawing corresponds exactly to a path over the bridges of Königsberg.

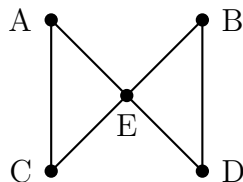
Pictures like the dot and line drawing are called *graphs*. Graphs are made up of a collection of dots (called *vertices*) and lines connecting those dots (called *edges*). The nice thing about looking at graphs instead of pictures of rivers, islands and bridges is that we now have a mathematical object to study. A *set* of vertices and a set of edges (in fact, we can take the set of edges to be a set of ordered pairs from the set of vertices). We have

distilled the “important” parts of the bridge picture for the purposes of the problem - it does not matter how big the islands are, what the bridges are made out of, if the river contains alligators, etc. All that matters is which land masses are connected to which other land masses, and how many times. This was the great insight that Euler had.

We will return to the question of finding paths through graphs later. But first, here are a few other situations you can represent with graphs.

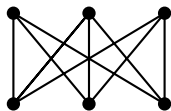
**Example:** Al, Bob, Cam, Dan, and Euclid are all members of the social networking website *Facebook*. The site allows members to be “friends” with each other. It turns out that Al and Cam are friends, as are Bob and Dan. Euclid is friends with everyone. Represent this situation with a graph.

*Solution:* Each person will be represented by a vertex and each friendship will be represented by an edge - that is, there will be an edge between two vertices if and only if the people represented by those vertices are friends. We get the following graph:



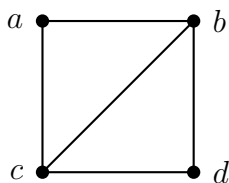
**Example:** Each of three houses must be connected to each of three utilities. Is it possible to do this without any of the utility lines crossing?

*Solution:* We will answer this question later. For now, notice how we would ask this question in the context of graph theory. We are really asking whether it is possible to redraw the graph below without any edges crossing (except at vertices).



## 1 Basics

While we almost always think of graphs as pictures, these are really just visual representations of mathematical objects. In fact, all a graph is is a set of vertices some pairs of which are “related” by an edge. For example, we can describe one graph like this: the vertices are the letters  $\{a, b, c, d\}$  and the edges are the pairs  $\{(a, b), (a, c), (b, c), (b, d), (c, d)\}$ . In other words, we have four vertices,  $a$ ,  $b$ ,  $c$ , and  $d$ , and  $a$  is connected (by an edge) to  $b$  and  $c$ ,  $b$  is connected to both  $c$  and  $d$ , and  $c$  is also connected to  $d$ . One way to draw this graph is this:



However we could also have drawn the graph differently. For example either of these:



Viewed as pictures, the graphs above are the same whether or not the vertices are labeled as they are, or at all. In fact, two graphs are equal precisely when there is a way to label the vertices so that all the pairs of vertices which have edges in one graph also have edges in the other graph, and *vice versa*. Two vertices which are connected by an edge are called *adjacent*.

The graph above is also particularly nice in that no pair of vertices is connected more than once, and no vertex is connected to itself. The technical name for graphs like this (no double edges or single edge loops) is called *simple*. The graph is also *connected* - you can get from any vertex to any other vertex by following some path of edges ( $a$  and  $d$  are connected by a path two edges long). A graph that is not connected can be thought of as two separate graphs drawn close together. Unless otherwise stated, we will assume all graphs are connected. However, our graphs might or might not be simple (the graph for the Königsberg bridge problem is *not* simple).

The graph above does not have an edge between  $a$  and  $d$ . Thus it is possible to add an edge to the graph (and keep it simple). If we add all possible edges (keeping the graph simple), then the resulting graph is said to be *complete*. That is, a graph is complete if every pair of vertices is connected by an edge. Since a graph is determined completely by which vertices are adjacent to which other vertices, there is only one complete graph with a given number of vertices. We give these a special name:  $K_n$  is the complete graph on  $n$  vertices.

Each vertex in  $K_n$  is adjacent to  $n - 1$  other vertices. We call the number of edges emanating from a given vertex the *degree* of that vertex. So every vertex in  $K_n$  has degree  $n - 1$ . How many edges does  $K_n$  have? One might think the answer should be  $n(n - 1)$ , since we count  $n - 1$  edges  $n$  times (once for each vertex). However, each edge is adjacent to 2 vertices, so we count every edge exactly twice in this way. Thus there are  $n(n - 1)/2$  edges in  $K_n$ .

In general, if we know the degrees of all the vertices in a graph we can find the number of edges. The sum of the degrees of all vertices will always be twice the number of edges, since each edge adds to the degree of two vertices. Notice this means that the sum of the degrees of all vertices in any graph must be even!

**Example:** At a recent math seminar, 9 mathematicians greeted each other by shaking hands. Is it possible that each mathematician shook hands with exactly 7 people at the seminar?

*Solution:* It seems like this should be possible - each mathematician chooses one person to not shake hands with. But it is impossible. We are asking whether a graph with 9 vertices can have each vertex have degree 7. If such a graph existed, the sum of the degrees of the vertices would be  $9 \cdot 7 = 63$ . This would be twice the number of edges (handshakes) so this says that the graph would have 31.5 edges. That is impossible - edges only come in whole numbers. Thus at least one (in fact an odd number) of the mathematicians must have shaken hands with an *even* number of people at the seminar.

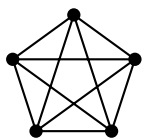
One final definition: we say a graph is *bipartite* if the vertices can be divided into two sets,  $A$  and  $B$ , with no two vertices in  $A$  adjacent and no two vertices in  $B$  adjacent. Of course the vertices in  $A$  can be adjacent to some or all of the vertices in  $B$ . If all vertices in  $A$  are adjacent to all the vertices in  $B$ , then the graph is a *complete bipartite graph*, and gets a special name:  $K_{m,n}$ , where  $|A| = m$  and  $|B| = n$ . The graph in the houses and utilities puzzle is  $K_{3,3}$ .

### Named Graphs

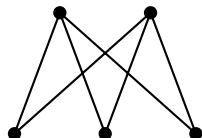
Some graphs are used more than others, and get special names.

- $K_n$ : the complete graph on  $n$  vertices.
- $K_{m,n}$ : the complete bipartite graph with sets of  $m$  and  $n$  vertices.
- $C_n$ : the cycle graph on  $n$  vertices - just one big loop.
- $P_n$ : the path graph on  $n$  vertices - just one long path.

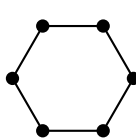
Here are some typical examples:



$K_5$



$K_{2,3}$



$C_6$



$P_6$

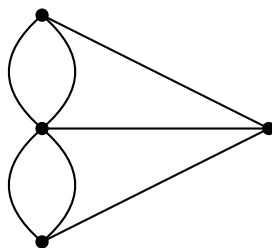
## Graph Theory Definitions

- **Graph:** A collection of *vertices*, some of which are connected by *edges*.
- **Adjacent:** Two vertices are *adjacent* if they are connected by an edge.
- **Bipartite graph:** A graph for which it is possible to divide the vertices into two disjoint sets such that there are no edges between any two vertices in the same set.
- **Complete bipartite graph:** A bipartite graph for which every vertex in the first set is adjacent to every vertex in the second set.
- **Complete graph:** A simple graph with edges connecting every pair of vertices.
- **Connected:** A graph is *connected* if there is a path from any vertex to any other vertex.
- **Degree of a vertex:** The number of edges connected to a vertex is called the *degree* of the vertex.
- **Euler path:** A path which uses each edge exactly once
- **Euler circuit:** An Euler path which starts and stops at the same vertex
- **Planar:** A graph is planar if it is possible to draw it without any edges crossing.
- **Simple:** A graph is *simple* if it does not contain any double edges or single edge loops (that is an edge from a vertex to itself).
- **Tree:** A graph with no cycles.
- **Vertex coloring:** An assignment of colors to each of the vertices of a graph.
- **Proper vertex coloring:** A vertex coloring is proper if adjacent vertices are always colored differently.
- **Chromatic number:** The minimum number of colors required in a proper vertex coloring of the graph.

## 2 Euler Paths and Circuits

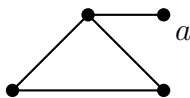
If we start at a vertex and trace along edges to get to other vertices, we create a *path* on the graph. If the path travels along every edge exactly once, then the path is called an *Euler path* (or *Eulerian path*). If in addition, the starting and ending vertices are the same (so you trace along every edge exactly once and end up where you started) then the path is called an *Euler circuit*. Of course if a graph is not connected, there is no hope of finding such a path or circuit. For the rest of this section, assume all graphs are connected.

The bridges of Königsberg problem is really a question about the existence of Euler paths. There will be a route that crosses every bridge exactly once if and only if the graph below has an Euler path:



This graph is small enough that we could actually check every possible path and in doing so convince ourselves that there is no Euler path (let alone an Euler circuit). On small graphs which do have an Euler path, it is usually not difficult to find one. Our goal is to find a quick way to check whether a graph has an Euler path or circuit, even if the graph is quite large.

If you have not already, be sure to try the worksheet on Euler paths. It is quite instructive to build graphs which do and do not have Euler paths. One way to guarantee that a graph does *not* have an Euler circuit is to include a “spike” - a vertex of degree 1.



The vertex  $a$  has degree 1, and if you try to make an Euler circuit, you see that you will get stuck at the vertex. It is a dead end. That is, unless you start there. But then there is no way to return, so there is no hope of finding an Euler circuit. There is however an Euler path - you can start at the vertex  $a$ , then loop around the triangle. You will end at the vertex of degree 3.

You run into a similar problem whenever you have a vertex of odd degree. If you start at such a vertex, you will not be able to end there (after visiting every edge exactly once). After using one edge to leave the starting vertex, you will be left with an even number of edges emanating from the vertex. Half of these could be used for returning to the vertex, the other half for leaving. So you return, then leave. Return, then leave. The only way to use up all the edges is to use the last one by leaving the vertex. On the other hand, if you have a vertex with odd degree that you do not start a path at, then you will eventually get stuck at that vertex. The path will use pairs of edges connected to the vertex to arrive and leave again. Eventually all but one of these edges will be used up, leaving only an edge to arrive by, and none to leave again.

What all this says is that if a graph has an Euler path and two vertices with odd degree, then the Euler path must start at one of the odd degree vertices and end at the other. In such a situation, every other vertex *must* have an even degree - since we need an equal number of edges to get to those vertices as to leave them. How could we have an Euler circuit? The graph could not have an odd degree vertex - an Euler path would have to start there or end there, but not both. Thus for a graph to have an Euler circuit, all vertices must have even degree.

The converse is also true: if all the vertices of a graph have even degree, then the graph has an Euler circuit, and if there are exactly two vertices with odd degree, the graph has an Euler path. To prove this is a little tricky, but the basic idea is that you will never get stuck because there is an “outbound” edge for every “inbound” edge at every vertex. If you try to make an Euler path and miss some edges, you will always be able to “splice in” a circuit using the edges you previously missed.

### Euler Paths and Circuits

- A graph has an Euler circuit if and only if the degree of every vertex is even.
- A graph has an Euler path if and only if there are at most two vertices with odd degree.

Since the bridges of Königsberg graph has all four vertices with odd degree, there is no Euler path through the graph.

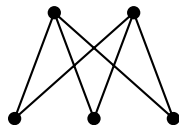
### Hamilton paths

Suppose you wanted to tour Königsberg in such a way where you visit each land mass (the two islands and both banks) exactly once. This can be done. In graph theory terms we are asking whether there is a path which visits every vertex exactly once. Such a path is called a *Hamilton path* (or Hamiltonian path). It appears that finding Hamilton paths would be easier - graphs often have more edges than vertices, so there are fewer requirements to be met. However, nobody knows whether this is true. There is no known simple test for whether a graph has a Hamilton path. For small graphs this is not a problem, but as the size of the graph grows, it gets harder and harder to check whether there is a Hamilton path. In fact, this is an example of a question which as far as we know is too difficult for computers to solve - it is an example of a problem which is NP-complete.

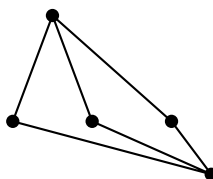
## 3 Planar Graphs

We leave the question of finding paths for a completely different type of question: when is it possible to draw a graph so that none of the edges cross, except for at vertices? If this is possible, we say the graph is *planar* (since you can draw it on the *plane*).

Notice that the definition of planar includes the phrase “it is possible to.” This means that even if a graph does not look like it is planar, it still might be. Perhaps you can redraw it in a way in which no edges cross. For example, this is a planar graph:



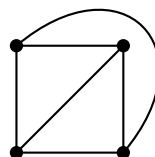
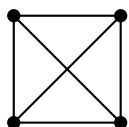
That is because we can redraw it like this:



The graphs are the same graph, so if one is planar, the other must be too. The original drawing of the graph was not a *planar representation* of the graph.

Now when a planar graph is drawn without edges crossing, the graph divides the plane into regions. We will call each region a *face*. The graph above has 3 faces (yes, we **do** include the “outside” region as a face). The number of faces does not change no matter how you draw the graph (as long as you do so without the edges crossing), so it makes sense to ascribe the number of faces as a property of the planar graph.

A warning: you can only count faces when the graph is drawn in a planar way. For example, consider these two representations of the same graph:



If you try to count faces using the graph on the left, you might say there are 5 faces (including the outside). But drawing the graph with a planar representation shows that in fact there are only 4 faces.

There is a connection between the number of vertices ( $V$ ), the number of edges ( $E$ ) and the number of faces ( $F$ ) in any connected planar graph. This relationship is called Euler's Formula.

### Euler's Formula for Planar Graphs

For any (connected) planar graph with  $V$  vertices,  $E$  edges and  $F$  faces, we have

$$V - E + F = 2$$



Why is Euler's formula true? One way to convince yourself of its validity is to draw a planar graph step by step. Start with the graph  $P_2$ :



Any connected graph (besides just a single isolated vertex) must contain this subgraph. Now build up to your graph by adding edges and vertices. Each step will consist of either adding a new vertex connected by a new edge to part of your graph (so creating a new “spike”) or by connecting two vertices already in the graph with a new edge (completing a circuit).

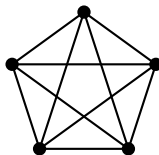


What do these “moves” do? When adding the spike, the number of edges increases by 1, the number of vertices increases by one, and the number of faces remains the same. But this means that  $V - E + F$  does not change. Completing a circuit adds one edge, adds one face, and keeps the number of vertices the same. So again,  $V - E + F$  does not change.

Since we can build any graph using a combination of these two moves, and doing so never changes the quantity  $V - E + F$ , that quantity will be the same for all graphs. But notice that our starting graph  $P_2$  has  $V = 2$ ,  $E = 1$  and  $F = 1$ , so  $V - E + F = 2$ .

## Non-planar graphs

Not all graphs are planar. If there are too many edges and too few vertices, then some of the edges will need to intersect. The first time this happens is in  $K_5$ .



If you try to redraw this without edges crossing, you quickly get into trouble. There seems to be one edge too many. In fact, we can prove that no matter how you draw it,  $K_5$  will always have edges crossing.

**Theorem.**  $K_5$  is not planar.

*Proof.* The proof is by contradiction. So assume that  $K_5$  were planar. Then the graph would satisfy Euler's formula for planar graphs.  $K_5$  has 5 vertices and 10 edges, so we get

$$5 - 10 + F = 2$$

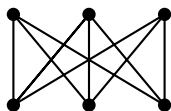
which says that if the graph were drawn without any edges crossing, there would be  $F = 7$  faces.

Now consider how many edges surround each face. Each face must be surrounded by at least 3 edges (since  $K_5$  is simple - it contains no double edges or loops). Let  $B$  be the total number of *boundaries* around all the faces in the graph. Thus we have that  $B \geq 3F$ . But also  $B = 2E$ , since each edge is used as a boundary exactly twice. Putting this together we get

$$3F \leq 2E$$

But this is impossible, since we have already determined that  $F = 7$  and  $E = 10$ , and  $21 \not\leq 20$ . This is a contradiction so in fact  $K_5$  is not planar. QED

The other simplest graph which is not planar is  $K_{3,3}$



Proving that  $K_{3,3}$  is not planar answers the houses and utilities puzzle - it is not possible to connect each of three houses to each of three utilities without the lines crossing.

**Theorem.**  $K_{3,3}$  is not planar.

*Proof.* Again, we proceed by contradiction. Suppose  $K_{3,3}$  were planar. Then by Euler's formula there will be 5 faces, since  $V = 6$ ,  $E = 9$ , and  $6 - 9 + F = 2$ .

How many boundaries surround these 5 faces? Let  $B$  be this number. Since each edge is used as a boundary twice, we have  $B = 2E$ . Also,  $B \geq 4F$  since each face is surrounded by 4 or more boundaries. We know this is true because the graph is simple (so there are no faces surrounded by 1 or 2 boundaries) and because  $K_{3,3}$  is bipartite, so does not contain any 3-edge cycles. Thus

$$4F \leq 2E$$

But this would say that  $20 \leq 18$ , which is clearly false. Thus  $K_{3,3}$  is not planar. QED

Note the similarities and differences in these proofs. Both are proofs by contradiction, and both start with using Euler's formula to derive the (supposed) number of faces in the graph. Then we find a relationship between the number of faces and the number of edges based on how many edges surround each face. This is the part that changes - in the proof for  $K_5$ , we got  $3F \leq 2E$  and for  $K_{3,3}$  we go  $4F \leq 2E$ . The coefficient of  $F$  is the key. It is the smallest number of edges which could surround any face. If some number of edges surround a face, then these edges form a circuit. So that number is the size of the smallest circuit in the graph.

In general, if we let  $g$  be the size of the smallest cycle in a graph ( $g$  stands for *girth*, which is the technical term for this) then for any planar graph we have  $gF \leq 2E$ . When this disagrees with Euler's formula, we know for sure that the graph cannot be planar.

## 4 Coloring

We conclude with perhaps the most famous graph theory problem - how to color maps.

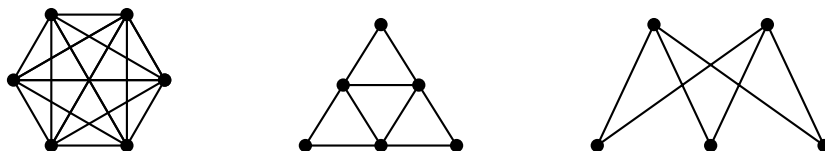
**Question:** Given any map of countries, states, counties, etc. how many colors are needed to color each region on the map so that neighboring regions are colored differently?

Actual map makers usually use around seven colors - for one thing, they require watery regions to be a specific color, and with a lot of colors it is easier to find a permissible coloring. But we want to know whether there is a smaller palette that will work for any map.

How is this related to graph theory? Well, if we place a vertex in the center of each region (say in the capital of each state) and then connect two vertices if their states share a border, we get a graph. The coloring regions on the map corresponds to coloring the vertices of the graph. Since neighboring regions cannot be colored the same, our graph cannot have vertices colored the same when those vertices are adjacent (connected by an edge).

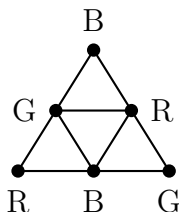
In general, given any graph, a coloring of the vertices is called (not surprisingly) a *vertex coloring*. If the vertex coloring has the property that adjacent vertices are colored differently, then the coloring is called *proper*. Every graph has a proper vertex coloring - for example, you could color every vertex with a different color. But often you can do better. The smallest number of colors needed to get a proper vertex coloring is called the *chromatic number* of the graph.

**Example:** Find the chromatic number of the graphs below.



*Solution:* The graph on the left is  $K_6$ . The only way to properly color the graph is to give every vertex a different color (since every vertex is adjacent to every other vertex). Thus the chromatic number is 6.

The middle graph can be properly colored with just 3 colors (Red, Blue, and Green). For example:



There is no way to color it with just two colors, since there are three vertices mutually adjacent (i.e., a triangle). Thus the chromatic number is 3.

The graph on the right is just  $K_{2,3}$ . As with all bipartite graphs, this graph has chromatic number 2 - color the vertices in the top row red and the vertices on the bottom row blue.

It appears that graphs can have any chromatic number. It should not come as a surprise that  $K_n$  has chromatic number  $n$ . So how could there possibly be an answer to the original map coloring question? If the chromatic number of graph can be arbitrarily large, then it seems like there would be no upper bound to the number of colors needed for any map. But there is.

The key observation is that while it is true that for any number  $n$ , there is a graph with chromatic number  $n$ , only some graphs arrive as representations of maps. If you convert a map to a graph, the edges between vertices correspond to borders between the countries. So you should be able to connect vertices in such a way where the edges do not cross. In other words, the graphs representing maps are all *planar*!

So the question is, what is the largest chromatic number of any planar graph? The answer is one of the best known theorems of mathematics:

**Theorem.** *The Four Color Theorem If  $G$  is a planar graph, then the chromatic number of  $G$  is less than or equal to 4. Thus any map can be colored with 4 or fewer colors.*

We will not prove this theorem. Really. Even though the theorem is easy to state and understand, the proof is not. In fact, there is currently no “easy” known proof of the theorem. The current best proof still requires powerful computers to check an *unavoidable set* of 633 *reducible configurations*. The idea is that every graph must contain one of these reducible configurations (this fact also needs to be checked by a computer) and that reducible configurations can in fact be colored in 4 or fewer colors.

The chromatic number of a graph tells us about coloring vertices - but we could also ask about coloring edges. What if we colored every edge of a graph either red or blue. Can we do so without, say, creating a triangle of same like colored edges (i.e., an all red or all blue triangle - we say the triangle is *mono-chromatic*)? Certainly for some graphs the answer is yes - try doing so for  $K_4$ . What about  $K_5$ ?  $K_6$ ? How far can we go?

The answer the above problem is known - I encourage you to try to solve it. We could extend the question though - what if we had three colors? What if we were trying to avoid other graphs. The surprising fact is that very little is known about these questions. For example, we know that you need to go up to  $K_{17}$  in order to force a mono-chromatic triangle using three colors, but nobody knows how big you need to go with more colors. Similarly, we know that using two colors  $K_{18}$  is the smallest graph that forces a mono-chromatic copy of  $K_4$ , but the best we have to force a mono-chromatic  $K_5$  is a range - somewhere from  $K_{43}$  to  $K_{49}$ .