Steve Gredell

Brian Witt

Graph Theory

A graph is a mathematical structure used to represent relations between objects. A graph consists of a number of "nodes" or vertices, which are connected by edges. These edges may be undirected, like a two way street, or directed, like a one way street.

Graph theory originated from topology and map making, specifically a paper written by Leonhard Euler (pronounced "oiler" like oil) on the seven bridges of Königsberg, which we will talk about later. The main use of graphs today is representing data in computers. It is this aspect that we will focus on because, whether you know it or not, you use graphs every day of your life.

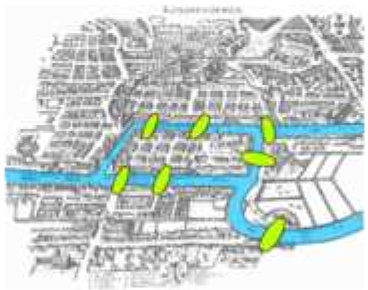
Four Color Theorem

Among the first uses of graphs was map coloring. While trying to color a map of the counties of England in the 1830's, Francis Guthrie found that he needed no more than 4 colors to color this, and any map. It was passed on to Alfred Kempe who proved the four color theory. Ten years later, the proof was shown to be false, but a five color theory was proved. After many years of failed attempts, it was finally proved in 1976.

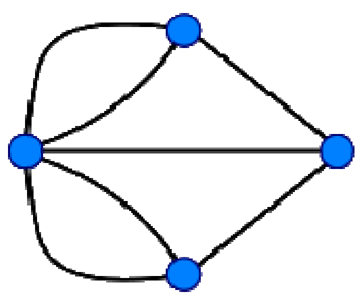
The first widely accepted proof of the four-color theorem was also the first major theorem to be proved using a computer. There are many algorithms which make graph coloring on the computer much easier, but the problem remains NP-complete to determine whether a graph can be colored with N colors, unless N=1 or N=2. The NP in NP-complete stands for Nondeterministic Polynomial time. To understand what that means, take the Fibonacci sequence for example. Using a recursive algorithm to compute the Fibonacci sequence is aproximately the equivalent of graph coloring. The fastest supercomputer in use today, the Cray XT5 Jaguar, just set a new record at 1.7 petaflops. A "flops" is one Floating point Operation Per Second. Peta- is the prefix which comes after tera-, which comes after giga-. That means this computer runs 1.75 QUADRILLION operations per second (an average computer can run at around 200 to 500 megaflops). Even with the Jaguar, computing the 200th Fibonacci number (not a real big deal, right?) would take 6.4995587438266819 * 10^18 years to compute(see final page for calculation). That is 6.5 quintillion years, which is more than a billion times longer than the history of the earth. Even when you take into account the fact that transistor density (for our purposes, computer speed) doubles every 18 months, it will be 99 years until there is a computer which can complete this in less than a month. As you can see, computers are not the answer to everything, and there are still problems computers cannot solve.

The Seven Bridges of Königsberg

One of the first problems explored by graph theory was the seven bridges of Königsberg, which Leonhard Euler explored in his paper. The city of Königsberg was split by the Pregel river, with 7 bridges crossing the river.



The story goes that the people of Königsberg liked to walk around in the afternoon when they had finished lunch (may or may not be true). They liked to walk all the way around the city, visiting both sides of the city, and both of the islands as well. The problem was to find a route through the city in which you could cross every bridge once, and only once. Euler found that the route through the city was unimportant to the route, since there were many ways to walk between 2 bridges. From this, he replaced each land mass with an abstract node. Each bridge then became an edge, completing the mathematical structure of a graph, and set the foundation for graph theory.



Traveling Salesman Problem

The Traveling Salesman Problem (TSP) is a famous graph problem first studied in the 1930's. Imagine you are a traveling salesman. You have several houses you have to cover today, going door to door for each one. What is the shortest path between all the houses, and your house where you are staying? Time is money and finding the shortest route is very important. When we turn to our trusty computer however, we find that the TSP falls into the NP-complete category of problems. The best algorithm for solving the TSP has a time complexity O(2^n), which means that it takes 2^n instructions to complete the problem for n nodes. If you had access to the Department of Energy's Jaguar computer at Oak Ridge National Labs, you could try to solve your door to door salesman problem. If you needed to visit 56 houses on Friday (plus your own), it would take 2^57 instructions to solve. After 23.5 hours on the Jaguar, you would have your answer, and a full half hour to complete your route!

So how do delivery companies such as UPS and the USPS get all of our stuff to us? The drivers need to know how to get to all of our houses and back to the headquarters in the shortest distance possible. However, there are ways to efficiently approximate to within 2% to 3% of the optimal TSP solution using hueristic algorithms.

Applications of Graphs

Google Maps, Mapquest, GPS units (Tom Tom) etc... use proprietary versions of Dijkstra's algorithm to compute driving directions, since all maps have weighted edges (the distance between 2 intersections) and none of these are negative (Google Answers). Dijkstra's algorithm is a complex algorithm that is nonetheless very efficient, finding the shortest path between two places. It has a time complexity in the worst case of $O(n^2)$.

Plumbing, electricity, phone, cable, and internet wiring can all use variations on Network Flow Modeling and Minimum Spanning Trees. Piping and wire is expensive, and you want to be able to provide service to everyone with the least amount of materials. With the Network Flow Modeling algorithm, the goal is to optimize the quantity flowing from the source to the sink (end). Due to the natures of how electricity and data (cable and internet) are transferred, they can be modeled very much like plumbing. While water flows through the pipes, so does electricity and data flow through wires.

The minimum spanning tree is a sub-graph which connects all of the nodes in a graph with the least amount of edge weight. Not every house in your neighborhood is directly connected to every other house with phone lines, and yet you can call anyone in your neighborhood. Kruskal's algorithm for finding the minimum spanning tree is very simple. You find the edge with the smallest weight, check to make sure that it does not make a cycle (a loop in the graph), and add it to the tree. Kruskal's algorithm is very fast, it has a time complexity $O(n \log n)$.

Graphs and the Internet

Many people are still studying the internet. They use graphs of the internet to study many critical computer science problems, such as security and communication. People use random generated graphs to represent the internet, because the real internet is too big to represent as a graph. In July of 2008, the search engine Google reached a milestone, as it indexed its trillionth page. Numbers like that make studying the actual internet rather useless.

While graph theory is one of the newest concepts in mathematics, having only been around less than 300 years, it is still extremely valuable.

Notes

Number of years to find the 200th Fibonacci number on the Cray XT5 Jaguar was computed using python command >>>2**138 / (1.7 * 10**15)/ (3600 * 24 *365)

2^{138} operations to compute the 200th Fibonacci number

1.7×10^{15} is 1.7 petaflops

$3600 \times 24 \times 365$ is one year of seconds

Works Cited

Google answers http://answers.google.com/answers/threadview/id/435679.html

Graph Theory http://en.wikipedia.org/wiki/Graph_theory