Study on Approximate Shortest Path & Implementation of Queries in Networks Distance

Introduction

Imagine you wanted to get in touch with Nelson Mandela using a sequence of personal introductions through friends and friends of friends. What is the shortest such sequence? Imagine you wanted to access a webpage on the Internet. Which routers should be used such that the necessary information is downloaded to your computer fastest? These questions have something in common, in that their (optimal) solution is the shortest path between two points of a network: a transportation network, a social network, and a router network.

The aim is to provide means to efficiently compute shortest paths in networks. With the development of biotechnology, more and more biological data is collected and available for analysis. One example is the GenBank and Proteins data from NCBI (National Center for Biotechnology Information). There is huge amount of data available, including DNA sequences, RNA sequences and protein sequences of all different species. And not much is known about this data. How can one extract the most interesting and knowledgeable patterns from that data which may guide us to more discoveries is an interesting task.

In the proposed work, efforts are proposed to be made to study Graphs are, of course, one of the prime objects of study in Discrete Mathematics. However, graphs are among the most ubiquitous models of both natural and human-made structures. In the natural and social sciences they model relations among species, societies, companies, etc. In computer science, they represent networks of communication, data organization, computational devices as well as the flow of computation, and more. In mathematics, Cayley graphs are useful in Group Theory. Graphs carry a natural metric and are therefore useful in Geometry, and though they are “just” one-dimensional complexes, they are useful in certain parts of Topology, e.g. Knot Theory. In statistical physics, graphs can represent local connections between interacting parts of a system, as well as the dynamics of a physical process on such systems.

For a computer network, routers and the connecting network cables are mapped to nodes and edges, respectively. In a social network, the connections are not physical. Individuals can be modeled by nodes; two nodes are connected by an edge whenever the corresponding individuals are friends. In other social networks, an edge may also indicate a private or professional relationship other than friendship. We first review various example structures, for which a graph serves as a suitable model. We then consider one central problem that can be solved using graphs: we investigate applications of the shortest path problem, in particular con-temporary applications of the point-to-point shortest path problem. We conclude the chapter by stating the
Networks & Graphs

Transportation Networks

Transportation networks are an integral part of the infrastructure in many countries. For this reason, the study of transportation networks is an important field of research. We give three examples of transportation networks: road, railway, and airline networks. A more realistic model of transportation would ideally integrate networks of all three types.

Road Networks

Since road networks often contain many bridges and tunnels, the corresponding graphs are in general not planar. However, road networks and planar graphs share some important properties, which render road networks tractable for many optimization problems. Navigation, for example, is not as difficult as it could be on a general graph, since road networks have some geometric and geographical orientation. Another characteristic of road networks is that, at every intersection, the number of streets (and thus the number of choices when navigating) is quite low. Still, planning efficient routes through a road network is very challenging since these networks may be huge (millions of nodes) and dynamic (travel times depend on various factors such as the current traffic situation and road maintenance).

Railway Networks

For Railway networks, Taking a train is represented by traversing an edge. Again, edges can be weighted with distances or travel times. This model arguably does not represent a real public transportation system very well.

Airline Networks

Air traffic networks can also be represented by graphs. Airports can be modeled by the nodes of a graph. Two nodes are connected by an edge if there is an aircraft that can start and land on and cover the distance between the corresponding airports. In this graph, almost all nodes are connected. While in road networks, each intersection had up to half a dozen of connections, in air traffic networks, airports can have hundreds of connections. Consequently, the corresponding graphs are significantly more dense.

Complex Networks

Various complex systems are highly interconnected: phenomena that were assumed to be local only are sometimes unexpectedly shown to influence the other end of the system. Researchers from fields such as physics, mathematics, computer science, biology, and social science analyze...
these systems to explain how (and sometimes why) everything is connected. The objectives of network scientists are (1) to make connections in real-world systems explicit, (2) to analyze and understand the network structures formed by these connections, and, possibly, (3) to exploit the structural properties of these systems.

**Networks in Biology**

Traditionally, proteins are identified based on their actions as building blocks, catalysts, or signaling molecules. The network view identifies a protein by its physical interactions with other proteins. This yields its contextual role in a protein–protein interaction network. Proteins and interactions are represented by nodes and edges of a graph, respectively.

Recently, many networks have been extracted from biological data. Examples other than protein interaction networks include metabolic networks, which encode biochemical reactions between metabolic substrates, and transcriptional regulatory networks, which describe the regulatory interactions between different genes.

**Social Networks**

Around 1929, the Hungarian writer Frigyes Karinthy formulated the following challenge: find a person who you can not be connected to by a friendship chain of at most five people. In other words, Karinthy conjectured that everybody knows everybody else through a short chain of personal connections. Milgram, in his famous small-world experiment, tried to verify this conjecture with an empirical study. Participants in the Milgram study were supposed to deliver a letter addressed to a specific person, not by mailing it directly, but by forwarding it to somebody they know on a first-name basis — somebody who is more likely to know the addressee. Milgram was interested in the number of forwarding steps, which corresponds directly to the length of a friendship chain. Although most letters never arrived, many letters reached the addressee after a very small number of forwarding steps only. "It’s a small world!!" Milgram’s result exhibits a stronger statement than Karinthy’s conjecture. For two individuals, the friendship chains connecting the two are not only short, it is also possible to “find” such a chain.

**Shortest Paths**

Once we have performed the abstraction by making connections explicit and modeling them by a graph, we can work with this model and compute properties of the graph without considering the exact details of the underlying network. Interesting properties could be graph distances between nodes. Tobler invoked... the first law of geography: everything is related to everything else, but near things are more related than distant things.

This law ought to be generalized to a law holding for many networks. An edge of a graph
essentially indicates a relationship between the two corresponding entities: between two interacting proteins, between two friends, between two intersections in a road network, or between two routers connected by a cable. If this relationship is somewhat transitive, which means that, for example, if protein A interacts with protein B, and B influences C, then A has an indirect influence on C. The shorter the chain of interactions, the stronger the influence. This is the first law of networks.

For example, two neighboring wireless devices cannot send at the same time due to interferences. These conflict graphs often occur in scheduling problems, which can be solved for example by coloring the graph. In this thesis, however, if Paul is friends with Linton and Stanley, then Linton and Stanley are probably closer to each other than if they were not friends with Paul. Or, to consider another example, if protein A interacts with and influences protein B, and B influences C, then A has an indirect influence on C. The shorter the chain of interactions, the stronger the influence. This is the first law of networks.

## Classical Results

Methods to find a shortest path were discovered and analyzed already in the late 50’s and early 60’s by Bellman, Bock and Cameron, Caldwell, Dantzig, Dijkstra, Floyd, Ford, Fulkerson, Hu, Klee, Leyzorek, Gray, Johnson, Ladew, Meaker, Petry, and Seitz, Minty, Moore, Mori and Nishimur, Parikh, Rapaport and Abramson, Shimbel, Warshall, Whiting and Hillier, and probably others. Despite the we restrict ourselves to graphs of the type encountered in the previous examples, assuming that an edge represents a somewhat positive relationship and not a conflict. If the user is interested in the distance to one target node only, it is possible to stop the search as soon as this specific target has been found. Some of the other classical algorithms solve the All Pairs Shortest Path (APSP) problem: the distance and the shortest path between all pairs of nodes is computed. The shortest path tree starting at the node representing the city of Los Angeles. Original by George B. Dantzig. “Hence the optimal path is from Los Angeles to Salt Lake City, then to Chicago, and finally to Boston.”

The importance of the problem and the numerous applications have stimulated research efforts for more than fifty years. A large fraction of these efforts targets the shortest path problem in transportation networks, since route planning is arguably one of the most important applications of shortest path algorithms.

## Traffic Solutions

Two objectives of traffic simulations are to forecast future traffic patterns and to predict the consequences of certain changes to the road network. While sometimes the consequences are easily predictable and intuitively clear, there are paradoxical situations where closing a road actually improves the overall traffic situation. Simulations can help to identify these counterproductive roads. Simulation results may also help in urban planning to reason about the economic and social impact of building new roads. With realistic estimates of population mobility and parameterized models for simulating the progress and transmission of a disease,
simulations may improve predictions in public health and epidemiology.

Image Segmentation

Image segmentation is an integral part of image processing applications such as accident disposal, medical images analysis, and photo editing. The image segmentation problem is to group together neighboring pixels whose properties are coherent. The grouping process often relies on shortest path computations.

Drug Target Identification

Shortest paths in interaction graphs are important in systems biology. Signaling paths for example are routes along which one molecule can affect another one. The average path length already reveals valuable information about a cell or a body.

Community Detection

Complex networks are often huge and thus difficult to analyze. One way to obtain an understanding of complex networks is to decompose the network at hand into related components, communities, and clusters. Several algorithms for clustering and community detection have been proposed. The algorithm by Girvan and Newman has been applied successfully to a variety of networks, including several social and collaboration networks, metabolic networks, and gene networks. Their algorithm iteratively removes edges with high betweenness centrality. If a network contains highly-connected communities that are only loosely connected by a few edges between clusters, then all shortest paths between different communities must go along one of these few edges, which will therefore have high edge betweenness. Removing these edges separates the communities from each other. The method works very well for small graphs but it does not scale due to its high computational demand.

Social Search

The sheer size of the web (currently, the web consists of billions of pages), renders the search for relevant information very challenging. Search engines are expected to find the “needle in the haystack.” The search interface is supposed to be kept simple and the average user is not entering much information; the search engine must find relevant information without knowing what the user actually wants (sometimes he may not even know it himself or he may not be able to express it appropriately). Imagine that a search engine would know basically everything about the user (this scenario may actually already be reality).

Involving context and social connections in ranking is a hot topic for search engines and computing social distances may soon be an important primitive, both for search engines handling keyword queries and for online stores recommending items for purchase.

Social Networking
Shortest paths in social networks seem to be of interest for end users. On the website oracleofbacon.org, for example, users can enter two names of actors and “the database server uses a breadth-first search (BFS) to find the shortest path between pairs of actors.”11 Such a webpage also exists for Mathematicians.12 In professional networking sites such as LINKEDIN or XING, users can add their business contacts in order to get in touch with potential clients or employers through a short chain of personal introductions. The corresponding web servers compute point-to-point shortest paths in an online setting.

Distance approximations for networks may be of interest for end users as well. If a sender can choose among different destinations, for example before downloading a large file from one of several replicated servers holding the same data; it is beneficial to predict the round trip time for each of the servers prior to actually communicating. This proximity estimate helps choosing the optimal server and connection.