

# CS 225 FINAL PROJECT PROPOSAL

QC15, PEILINH2, GZHAI5, YIKAIT2

## 1. LEADING QUESTION

For our main goal, we want to solve the page rank problem. PageRank can be thought of as a model of user behavior. We assume there is a random surfer who is given a web page at random and keeps clicking on links, never hitting “back”, and he will keep doing this until he achieves a steady state. Thus given a graph, its nodes can be seen as websites, and its (directed) edges can be seen as links within the source webpage, and the weight on the edges determines probability of him clicking that link. The main problem for us is to determine the probability of the surfer browsing a given website after a period of time long enough to reach the steady state.

In this case, we can modify the use of Page Rank to airlines and airports. With the nodes being airports, edges being airlines, and weight being the distance of taking that flight. With page rank, we can figure out the importance of airports in the world.

As a side project, a graph can be seen as a CW complex as well as a simplex. Thus we can compute the homology, cohomology, as well as the fundamental group of the topological space. This goal can be achieved by a spanning tree algorithm plus a certain degree of math knowledge. We will treat this part of the project as the least important and might skip this completely depending on the amount of time left after we finish everything else.

## 2. DATASET ACQUISITION AND PROCESSING

For our final project, we are going to use the dataset of “OpenFlights”. Open flights is an open-source data set of flight routes and airports suggested in the given documents. We will download the .dat files which includes the data of airlines, airport names, cities, and countries from the OpenFlight website. For processing those data, We plan to use airports as nodes, airlines as edges, and use distance as the weight. More specifically, we will use the airport data to compute the distance between each airport and use the reciprocal of those distances as the weight.

The website of the above mentioned dataset is “<https://openflights.org/data.html>”. The download links for the two dat files are “<https://raw.githubusercontent.com/jpatokal/openflights/master/data/airports-extended.dat>” and “<https://raw.githubusercontent.com/jpatokal/openflights/master/data/airlines.dat>”.

## 3. GRAPH ALGORITHMS

The followings are the algorithms and goals we will try to implement throughout this month:

- Graph Traversal (including BFS, and possibly DFS).
  - input: a graph (might have multiple path components).
  - output: a traversal of the graph given.
  - target big-O:  $O(m + n)$ , with  $n$  is the number of Nodes and  $m$  is the number of edges.
- Minimal Spanning Tree (including Kruskal’s Algorithm, and possibly Prim’s).
  - input: a graph (might have multiple path components).
  - output: a minimal spanning tree of the graph given.
  - target big-O:  $O(n + m \cdot \lg(n))$ .

- (possibly) Use spanning tree to calculate the homology groups of the graph (and possibly cohomology and fundamental group).
  - \* input: a graph (might have multiple path components).
  - \* output: the fundamental group, homology and cohomology groups of the graph.
  - \* target big-O:  $O(n + m \cdot \lg(n))$ .
- Page Rank algorithm.
  - input: a graph (might have multiple path components).
  - output: the steady state of Page rank.
  - target big-O:  $O(n^2) * p$ , where  $p$  is the number of iterations to reach steady state.

#### 4. TIMELINE

Since we meet the group discussion every Monday evening, we are going to set the coming five Mondays (4/12, 4/19, 4/26, 5/3, 5/10) as the deadline for different segments of our final project. We plan to finish the graph traversal and minimal spanning tree at most in 2 weeks. (Mid-Project Checkin). For the third week, we are going to write the code for page rank. We hope we could end our project one week earlier than the final week so that we can still have another week to optimize our code and consider adding more extra features.