

CS225 Fall 2021 -- Final Project Proposal

xx19, ruozhen2, kangyuf2, yuxuan19

Leading Question

- **What is the shortest path from one destination to another destination?**
- **What is the shortest distance from one destination to another destination?**
- **What are some of the popular airports/ destinations?**

Our group selected OpenFlights as our dataset, which covers Airport IDs, Cities, Countries, airlines, routes, longitude, latitude, and so on. Therefore, we believe that by knowing the starting point and the destination, we can solve potential routes that consist of the least transferring frequencies, least distances and create a world map that has the demonstration of different routes.

With these in mind, we decide to implement the shortest path algorithm for determining the least transfer routes and least distance routes between destinations, a Pagerank algorithm for determining popular airports and destinations, a GUI for users to interact and obtain shortest path information, and a world graph portraying airline routes.

Dataset Acquisition and Processing

- **Dataset**
OpenFlights (<https://openflights.org/data.html>). In specific, we will use airports, airlines, routes, planes, and countries dataset as supports for our project.
- **Download**
We will download all of the useful datasets from the OpenFlights website and store them in a CSV document type.
- **Process**
When processing our data, the primary programming language we will use is Python, specifically its libraries NumPy, dataprep.eda, Pandas, and SciPy. Since the datasets we obtained are all raw datasets, given different numeric features, we will need to clean all dirty data, remove duplicates, fix structural errors, and handle outdated, incomplete, as well as inconsistent data.

Graph Algorithms

For our project, we plan to implement two or three graph algorithms depending on the progress of our project. These graph algorithms will be implemented based on our dataset, trying to solve the proposed problems to discuss in the leading question.

- **Goal 1: DFS traversal**
We will implement a DFS traversal for our graph data structure. We expect the running time of DFS to be $O(N)$.
- **Goal 2: Find the shortest path/distance between two destinations**
We will try to use Dijkstra's Algorithm to find the shortest path between two airports. Here the input includes two destinations, and the outputs are the paths between two specific nodes (in vector or array form). At the same time, we will also consider the transit route questions, most minor distance questions (directed weighted graph), and maybe least time questions. We estimate the running time of our implementation of Dijkstra's Algorithm will be $O(N^2)$, where N is the number of vertices.
- **Goal 3: Find the most important airports/destinations**
We plan to use PageRank to find the most important airports/destinations. If it is not allowed, we will try to implement betweenness centrality to solve this problem. The input will be information on flights, while the output will be a list of airports sorted based on their importance. We expect the running time will be $O(k^N)$, where N is the number of airports and k is the number of iterations when implementing PageRank.

These goals are only preliminary ideas. The plan may be adjusted according to the difficulties and problems when programming, but the framework and direction of the entire project will not change.

Timeline

- **Nov 15th**
- focus on the OpenFlights dataset cleaning and filtering
- **Nov 22nd**
- complete the OpenFlights dataset cleaning and filtering
- construct a general code structure (graph class, necessary functions, DFS traversal)
- implement the shortest path algorithm for finding preferred routes
- **Nov 29th (Mid-Project Check-in)**
- complete the general code structure, including the DFS traversal function
- complete the implementation of the shortest path algorithm
- create some test cases for current algorithms
- **Dec 6th (Final Project Draft)**
- complete the Pagerank algorithm of finding the most important airports
- implement a GUI for checking shortest path, distance, and essential airports
- create some test cases for current algorithms
- create a world graph portraying airline routes
- **Dec 13th (Final Project Submission)**
- complete a written report of our project, including goals, development, and results
- complete the presentation slides, speech, and video
- any remaining implementation needed when inspecting code