

# Project Proposal

Liwen Sun, Jishu Yang, Zhekai Ge, Yihan Zhang

For this project, we are going to use Open flights as our dataset. This database contains over 10,000 airports, train stations and ferry terminals spanning the globe. We choose this dataset because it is closely related to our real life and we hope we can use our project to find the shortest path between airports all over the world which can be a convenient tool to solve relevant practical problems.

## Dataset Acquisition and Processing

We are going to download our dataset from this link: <https://openflights.org/data.html>. Since this dataset file is .dat, we will use ifstream and stringstream in cpp iostream to load and process these data. We mainly focus on the specific information of these airports like map coordinate, the name of source and destination etc. Every entry of this data in our project is like the name of airport, the source, the destination, distance, latitude and longitude. Since the real-world data may be messy, we will choose to contain the most accurate and detailed data entry in our project.

## Graph Algorithms

Firstly, we are going to use BFS algorithm to search each airport from source to destination since BFS is the most efficient and direct method for our project.

Secondly, we are going to use Dijkstra's algorithm to help us determine the shortest path between two airports. Applying this algorithm

to our project, we can make useful suggestion on how to make flight plans. Since we are going to use airport dataset, this algorithm takes it as a directed graph with edge weight. We anticipate that every edge weight is going to be the distance between every airport and our graph nodes will be the name and map coordinate of every airport.

Thirdly, we apply PageRank algorithm to our project. PR is an algorithm used by Google Search to rank webpage in their search engines to determine the importance of every webpage. In our project, we will use PR to determine the importance of every airport. The higher the probability of visiting airport is, the more important the airport is.

## **Timeline**

April 9-13 Data acquisition and processing

April 14-17 Complete first algorithm BFS

April 19-24 Complete second algorithm Dijkstra's algorithm

April 23-28 Complete third algorithm PageRank algorithm

May 1-4 Add test cases to our project

May 6-8 Check mistakes and improve the readability of codes

May 9-11 Produce the final deliverables