City University of Hong Kong
SDSC2004 Data Visualization
Homework 3

Zhumakhan Arman

SID: 56027756

Introduction

People use planes to go to other cities or countries and sometimes take several transitions in airports because of no direct flights. The report aims to research flight routes between airports over the world and analyze the distribution of routes using publicly available OpenFlights data. All visualizations are made using Gephi.

Description of data

All data were taken from the OpenFlights website. This website contains historical information about flight routes, airports, airlines around the world. Unfortunately, data hasn't been updated since June 2014.

The OpenFlights Flight routes data is used as edge data in my network. It contains 67663 routes between 3321 airports on 548 airlines spanning the globe. It has the following information: Airline, Airline ID, Source airport, Source airport ID, Destination airport, Destination airport ID, Codeshare, Stops, Equipment.

The OpenFlights Airports data contains information of 7698 airports. Each entry includes the following information: Airport ID, Airport Name, City, Country, IATA code, ICAO code, Latitude, Longitude, Altitude, Time zone, Daylight saving time.

In Gephi application, Source airport ID and Destination airport ID are used for defining edge. Airport Name is used as a label for node. Also, latitude and longitude of airport location are employed for visualizing flight routes. It gives an accurate representation of airports and enables straightforward interpretation of network visualization.

Figure 1. A network of airports and routes between them

Analysis and interpretation

In the network of flight routes, there are actually 7698 nodes and 36907 edges. Only 3214 of nodes have at least one edge. It means that the remaining 3484 airports are mentioned in Airports data but not used Flight Routes. The reason for ignoring these airports might be technical issues during data collection, age of these airports. Therefore, all analyses include only nodes that have at least one edge. Then the network turns to have 3214 nodes, directed 36907 edges. Also, when there are several airlines between airports, a number of airlines become a weight of edge.

Figure 1 shows the visualization of the network of flight routes and airports. A number of airports and flights in North America, Europe and China is more than anywhere in the world. The network density is 0.004. It shows that overall, an airport is connected to 0.4% of the other airports in the world. The average clustering coefficient of the network is 0.476. It means that most of the nodes are connected to each other in groups.

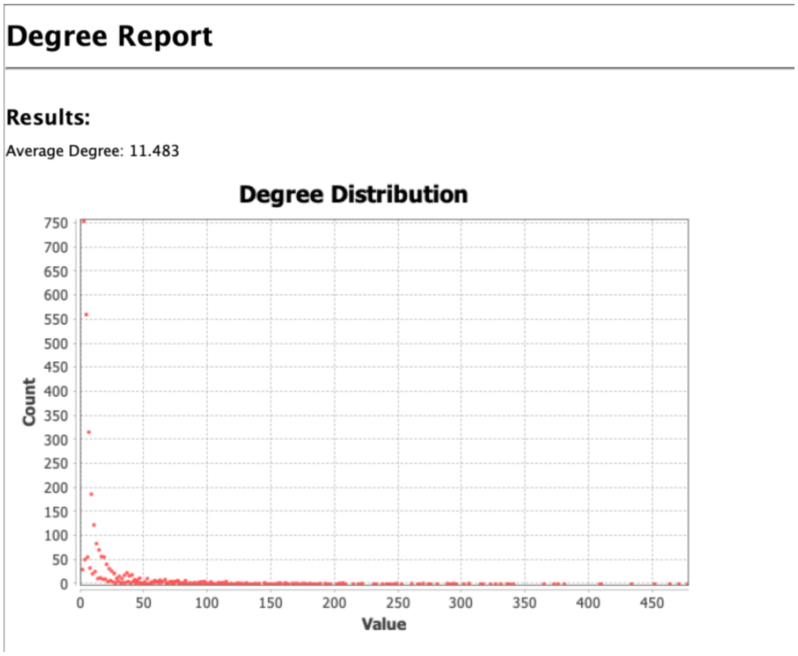

Figure 2. Degree distribution (using Gephi)

Average weighted degree of nodes is 11.483. It means that an average airport has 11 routes through it. On Figure 2, it's shown that the degree of nodes decreases exponentially. There are a lot of small airports with few routes in the world.

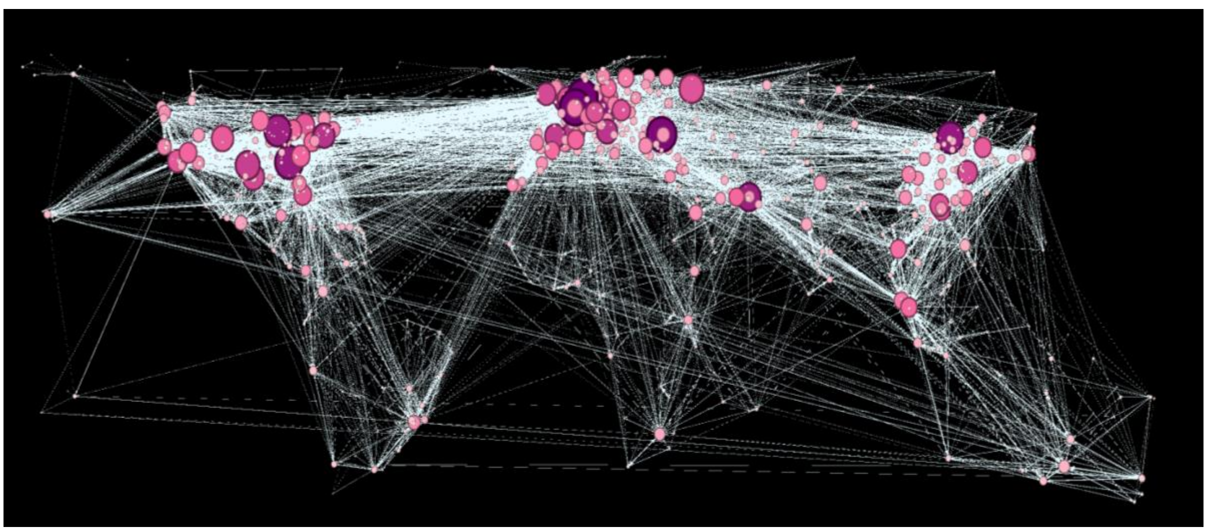

Figure 3. A network that sizes of nodes depends on degree

On Figure 3, it's clearly shown that airports in Europe have more connections than other airports. The main reason might be a vast number of cities and countries in Europe and the population. Though there are a moderate number of airports in Africa, they have few flights to other airports.

Figure 4. The airports with the greatest number of routes

Airports on Figure 4 are top 8 airports with the greatest number of routes. Five of them are located in Europe. It again shows the power of Europe in the airline industry.

The diameter of the graph is 13. It is the maximum eccentricity of any vertex in the graph (Wikipedia). In other words, the maximum number transition to go from one airport to any airport in the world is 13. The average path length of the network is 3.986. It means that on average to go somewhere in the world, people get four flights.

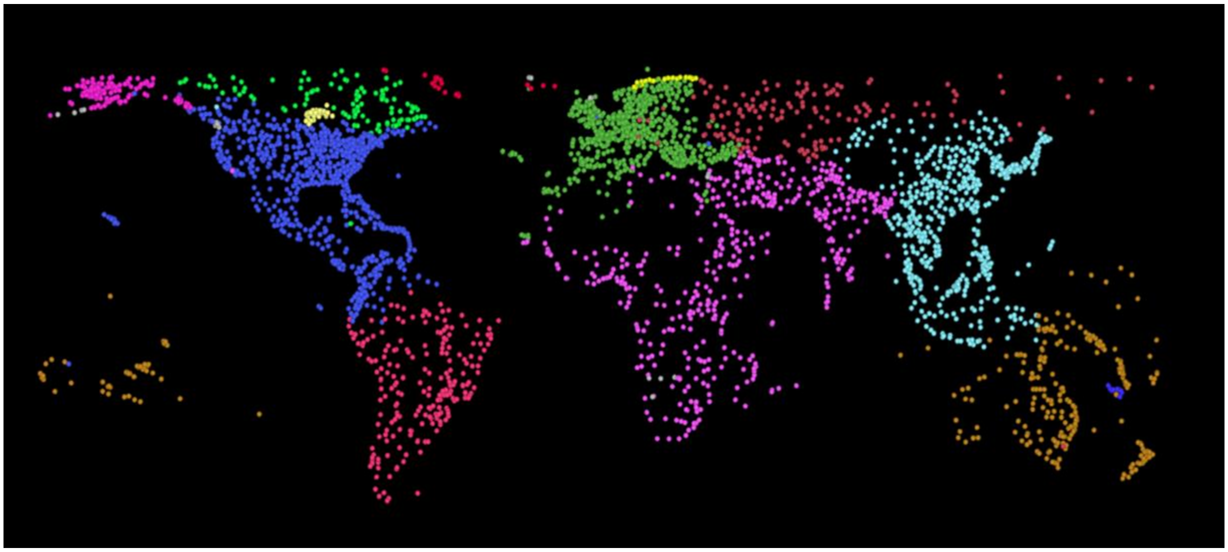

Figure 5. Nodes colored by modularity class with resolution 1

The modularity of the network with resolution 1 is 0.226, and a number of communities is 23. It is clearly shown on Figure 5 that most nodes are divided into modularity class by the continent and countries. It is clear evidence that airlines prefer to fly to airports inside the same country or continent. Moreover, this map illustrates that airports in countries that are politically close constitutes the same modularity group. For example, almost all airports in Europe are in the same community (they are in the European Union). Airports in Central Asia countries are in the same group of Russian airports due to the historical and political reasons (they were part of the USSR), though they are close to the South Asian or Chinese airports too. Also, airports in Southeast Asia countries are grouped together with Chinese airports, though some of them are closer to Australian airports.

References:

1. Gephi website. Available at: https://gephi.org/
2. OpenFlights website. Available at: https://openflights.org/data.html
3. Sebastien H. (2010). Map Geocoded data with Gephi. Gephi Blog. Available at: https://gephi.wordpress.com/2010/05/17/map-geocoded-data-with-gephi/
4. Wikipedia. Distance (graph theory). Available at: https://en.wikipedia.org/wiki/Distance_(graph_theory)