CS 458-558 Project 2

Most Popular Airports for Domestic Flights in the US from Chicago, Texas in 1997

Milan Thakkar, Conner Rhea, Grace Todd

The data described in this project represents the flights between international and domestic flights available from the United States. Figure 1 describes a directed graph of the serviceable domestic flights from the continental United States, with a service minimum of 30 airports for each node represented. The nodes are arranged based on their geographic location using their real longitude and latitudes. From this visualization we found that the airport with the highest number of flights to other airports is Chicago O'Hare International Airport, with Dallas-Fort Worth International Airport being the US airport with the second highest number of incoming and outgoing flights.

Similarly, figure 2 describes a directed graph of the relationships between all airports in the United States, including Hawaii, Alaska, and the Puerto Rico region. The nodes are also arranged based on their geographic location using their real longitude and latitudes. This figure visualizes all of the airports in the United States with the highest number of incoming and outgoing flights, with a more loose constraint on the minimum required number of incoming and outgoing flights.

Figure 3 visualizes the top 40 airports in the United States, with service to the most number of other airports across the world. The most popular airports were found to be Dallas-Fort Worth Airport and Chicago O'Hare International Airport.

Figure 4 describes a subset of the Airport Dataset described through the use of a glyph. This data set specifically looks at the Chicago O'Hare International Airport, and all the significant edges (degree > 30) that connect to it. The direction is determined through the vector of the difference of the longitude and latitude of Chicago O'Hare International and the other airport, and the color and size of the green ball corresponds to the number of flights between the two airports, the sphere being bigger for the larger number of flights between them.

To measure uncertainty, we implemented blurring in the vectors originating from the Chicago airport; the higher the uncertainty, the more blurry the vector appears. The size of the green sphere indicating the origin of the vector will also be smaller with higher levels of uncertainty.

Visualization with Gephi

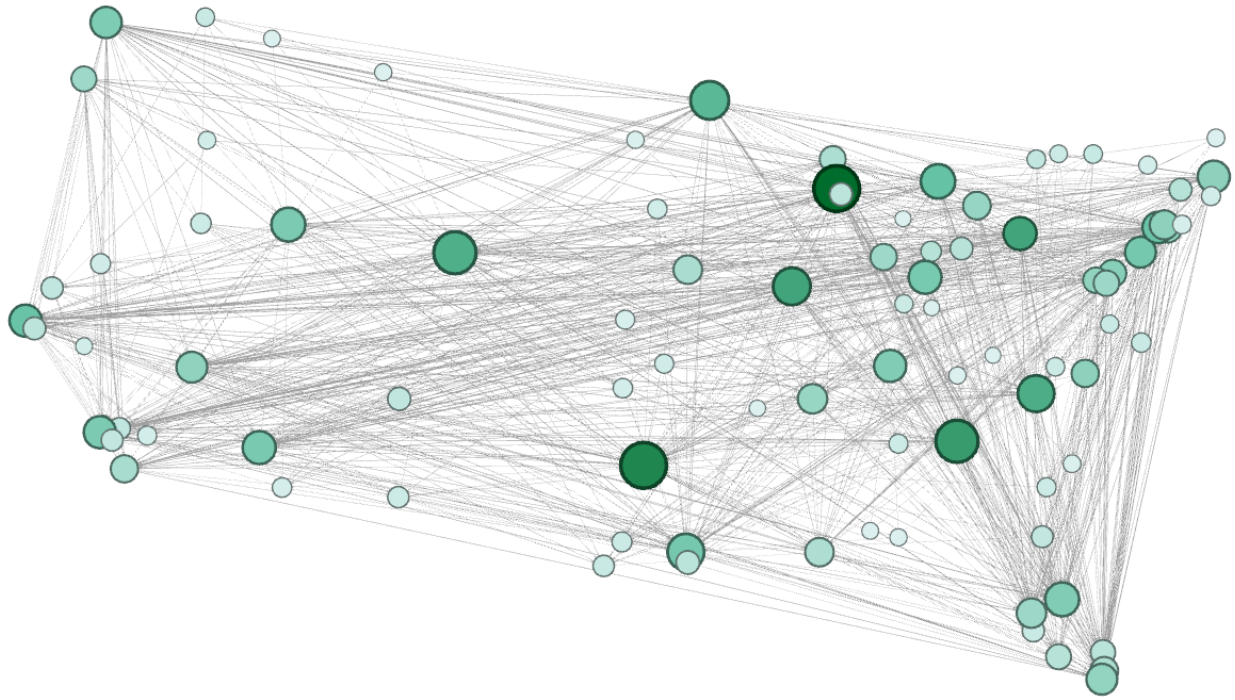

Figure 1: Map of airports with most domestic flights constrained to the continental United States. Airports with service to the most number of airports are larger and have a darker hue. Filters: Longitude, Edge weight

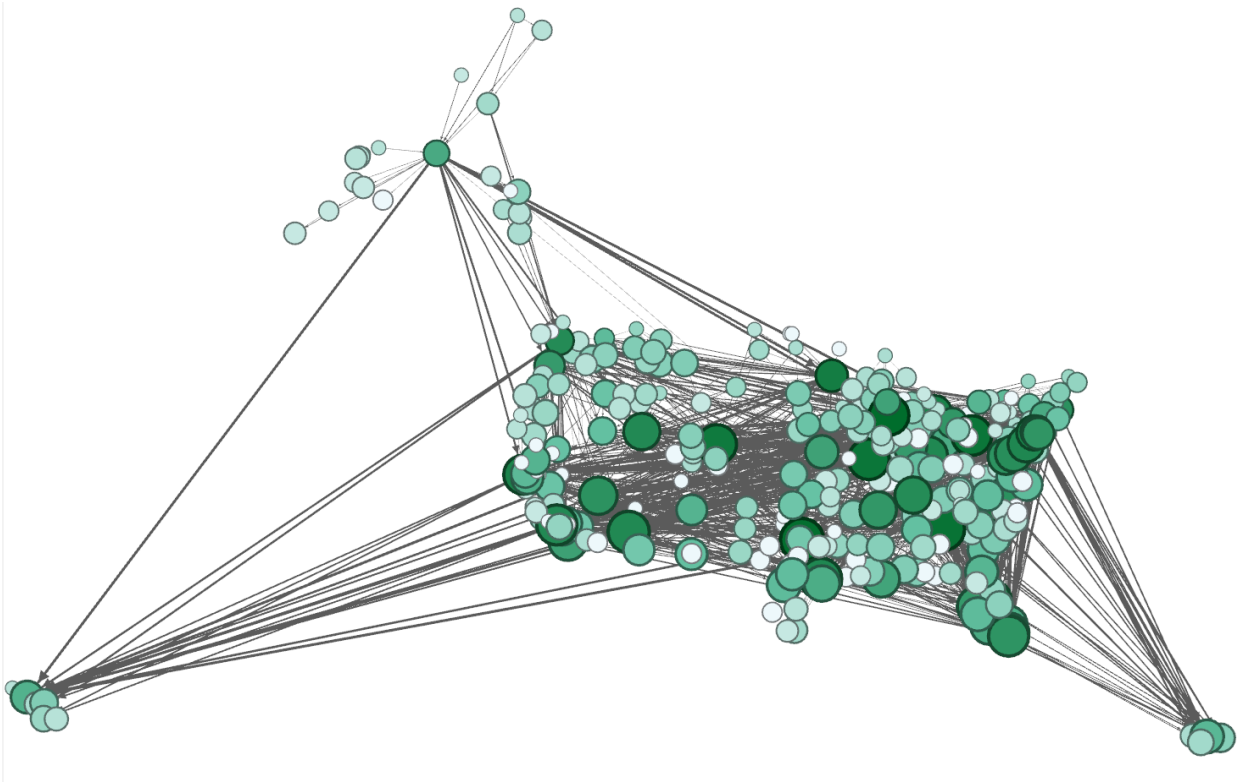

Figure 2: Map of airports with domestic flights to or from the United States with a constrained longitude, from Hawaii to Puerto Rico. Airports with higher frequency of service to other airports have a darker hue and a larger node size, with a heavier edge thickness. Airport nodes have not been labeled to increase readability. Filters: Edge weight, range.

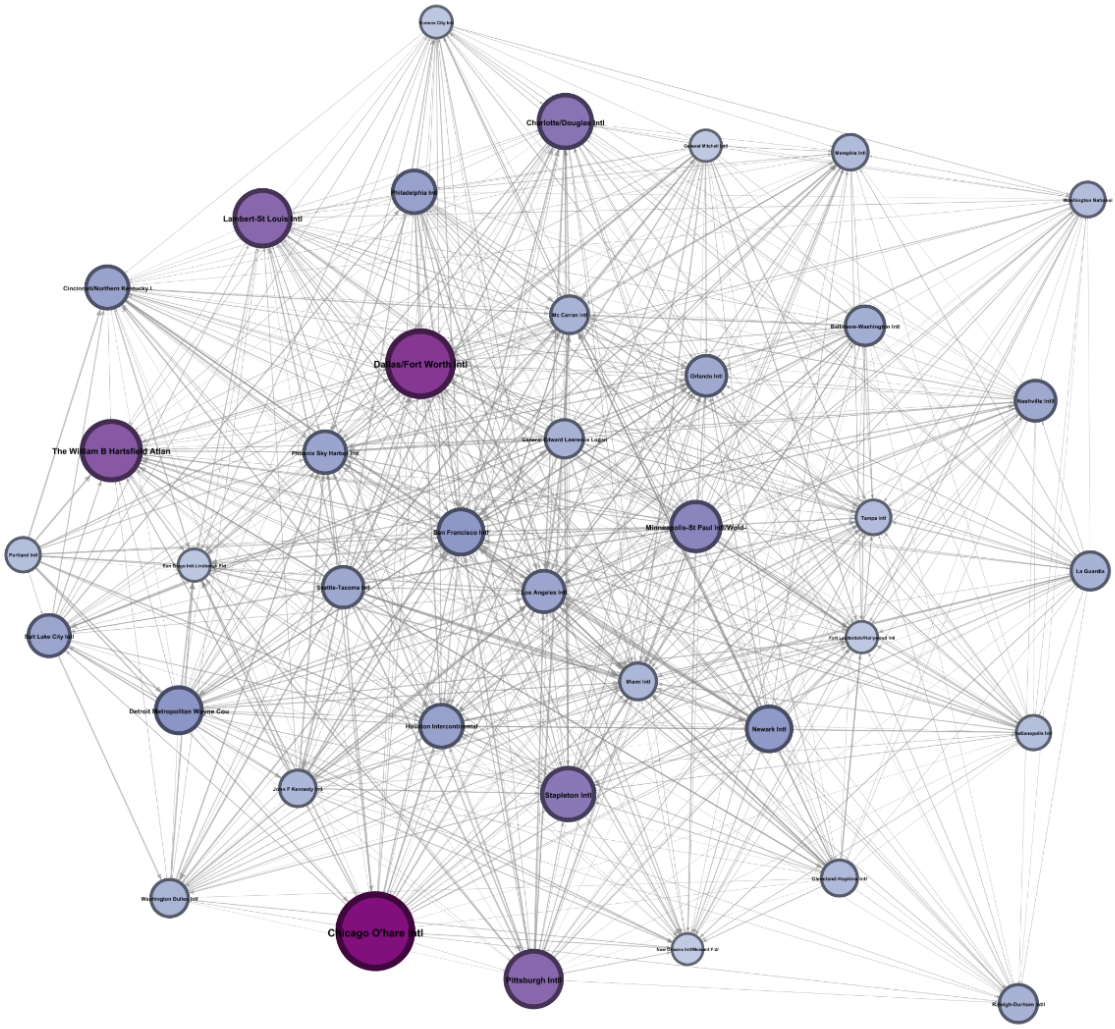

Figure 3: All airports in the United States with service to more than 30 other airports. Airports with service to the highest number of airports in bright purple. Edges connect each airport to its serviceable airports. Filter: Edge weight, degree range.

Glyph Designs

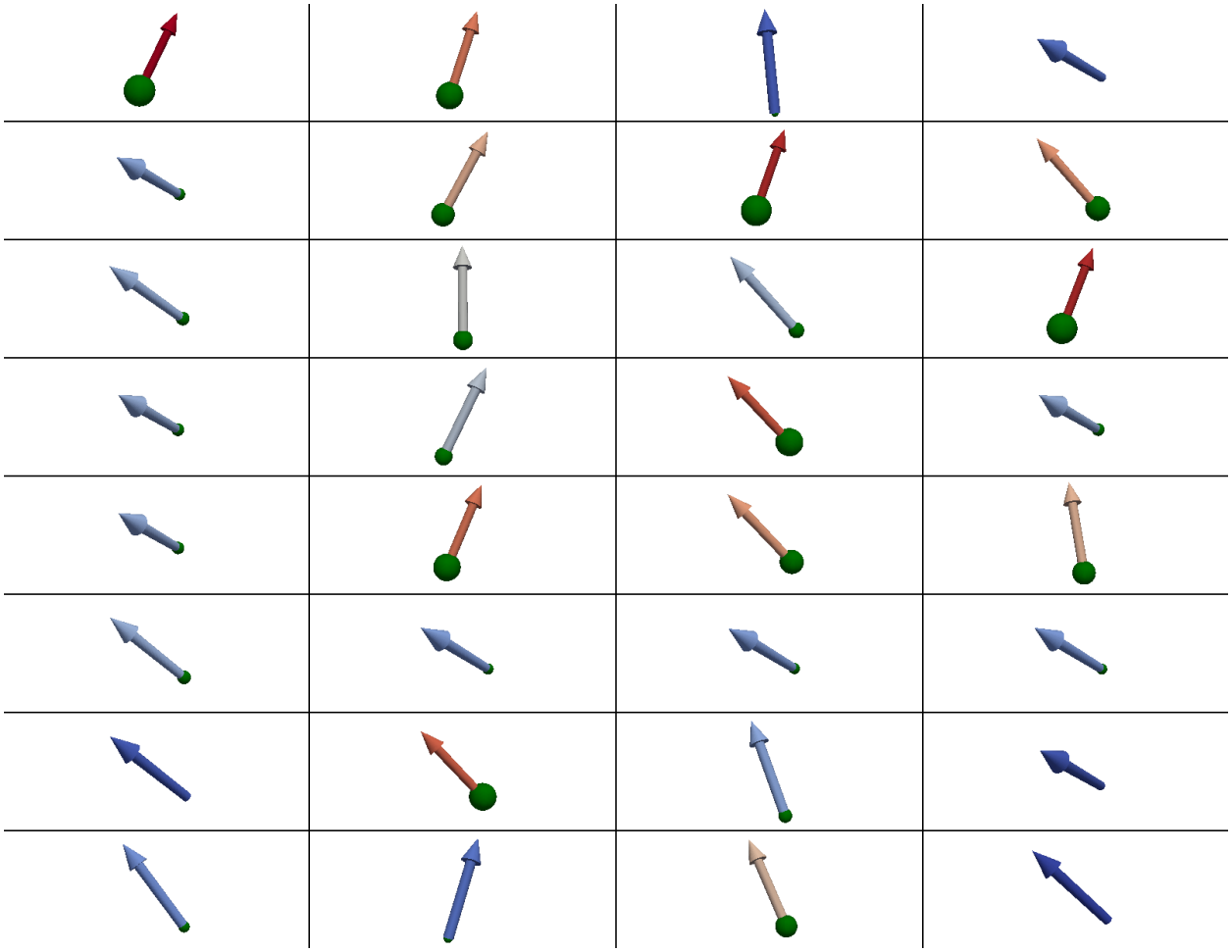

Figure 4: A subset of the Airport Dataset described through the use of a glyph. This Data Set specifically looks at the Chicago O'Hare International Airport, and all the significant edges (degree > 30) that connect to it. The Direction is determined through the vector of the longitude and latitude of the two airports, and the color and size of the green ball corresponds to the number of flights between the two airports, the ball being bigger for the larger number of flights between them.

Uncertainty

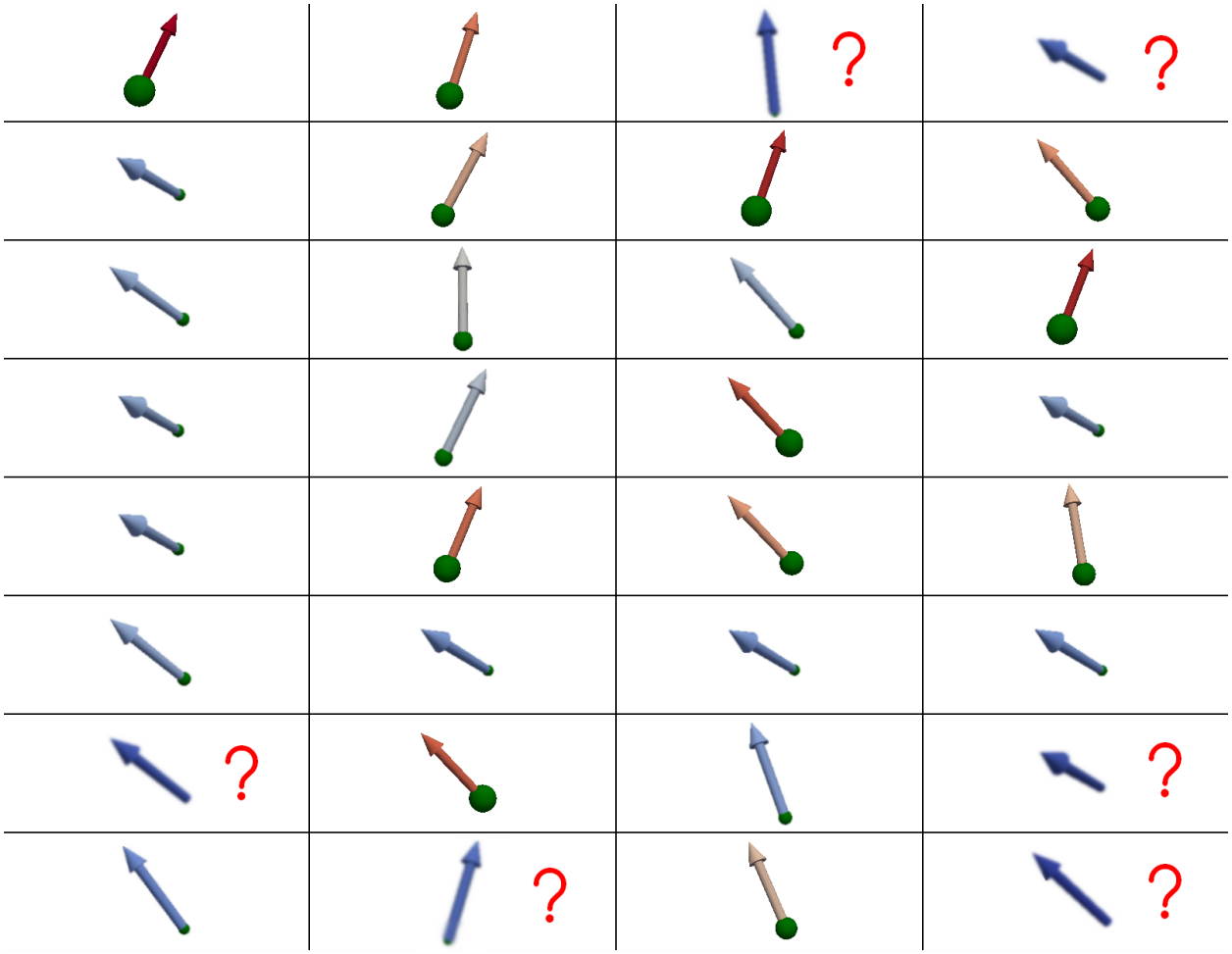

Figure 5: Glyphs with uncertainty added. Airports with smaller numbers of weights have increased blurriness to indicate the uncertainty of accurate representation of incoming and outgoing flights. Question marks denote ultimate blurriness and uncertainty in specific vectors.

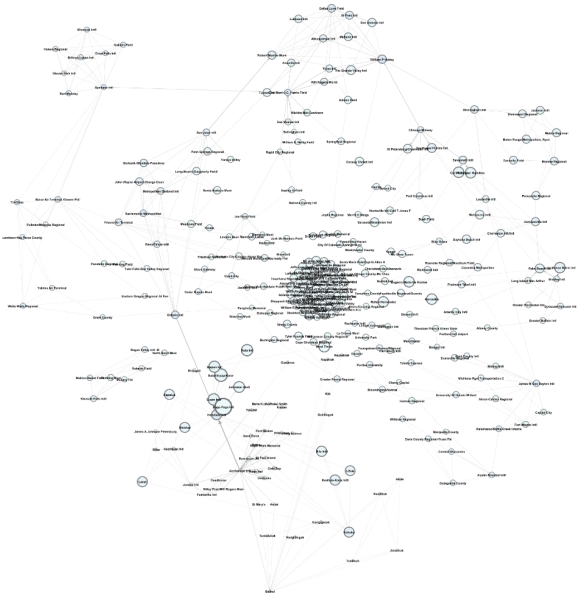

Figure 6: An inverse degree graph showing airports with degree < 30, these airports only have very loose connections and few data points recorded, including a cluster within the middle of the diagram showing airports that are only connected to larger airports free floating with no edges in a cluster. These airports are relatively unimportant in the grand scheme of American flights, however it's possible they may see more international flights that have not been recorded, meaning there is uncertainty in their importance overall.

There are a few variables that could contribute to uncertainty in the data visualization, the most prominent being:

- **Uncertainty with edge weights:** Smaller airports that get traffic from few other airports increase levels of uncertainty because we are not receiving as much data. 330 airports are represented in the data set, but only 40 of them are significant, as shown in figure 4. This could be due to the location of some of the less significant airports (i.e. airports in the Alaskan wilderness, where there is not much need for air traffic), but this could also mean that the incoming and outgoing flights are not being documented as thoroughly. This data set also might not consider the destinations of private planes, and might only take commercial airlines into account.

- **Global uncertainty:** This data set does not include international flights, therefore there is a level of uncertainty regarding how many international flights are incoming and outgoing from the United States.
- **Local uncertainty:** In order to represent the data based on the geographic location of the airports, we researched and added the latitude and longitude of each airport more or less by hand. This has the potential to cause uncertainties in the geographic location of each of the airports, which could skew the visualization results.
- **Global uncertainty:** Considering the age of this dataset, there have undoubtedly been more airports developed in the United States since 1997. This data set is incomplete, and therefore could raise uncertainty regarding the popularity of each airport. The recorded weights of the provided airports will be inaccurate with regards to incoming and outgoing flight numbers of 2023.

To measure uncertainty, we implemented blurry shapes to our glyphs to indicate which outgoing flights have the lowest weights. The airports with the lowest number of serviceable airports have the highest risk of inaccuracy. The principles for designing uncertainty visualizations in our data set are mostly global and local uncertainties, primarily revolving around the less popular airports and the age of the data set.