3/29/2019

By: Deane Buckingham

DS745-Sp19

Networking Project

US Airports

**Basic Network Details**

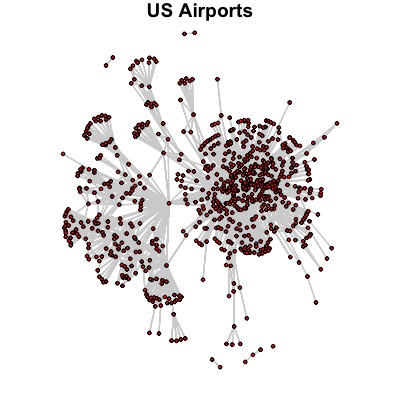
The network data I chose to model is called, “USairports” from the igraphdata package in ‘R’. The network is made of flights from December of 2010 (a busy travel period). There are 755 nodes representing airports and over23,000 directed edges representing a flight of a given type of aircraft with a given carrier.

The edges are directed since each edge is a flight going to/from the airport nodes in the graph. Since I’m only really concerned about connections between airports, I converted the network to be undirected where each flight simply represents a tie.

The node attributes for each airport are its 3-letter IATA code, city/state, and geographic location. There is not much complexity here, but the edge attributes include Aircraft Type, Carrier, Distance Traveled, Number of Departures, Number of Seats, and Number of Passengers.

The density was quite low at 0.08. While I tend to think of flights zipping to and from airports from all over, it definitely appears there is a limited number of potential routes that most flights take.

The diameter of this network is 8. This may seem small, but it actually was pretty close (if not larger) than what I expected. Although there are hundreds of airports all over the US, flights travel long distances and most folks would rather take a direct flight (or path) rather than go a roundabout way. I actually expected the diameter to be around 6, but I suppose some smaller flights could possibly take 8 hops to get from coast to coast.



Here we see a basic visualization of the network. We might be able to see a few airports with an abundance of connections (hubs). It even appears that certain groups of airports only funnel into these hubs as their main connection to the rest of the airport network. Like most scale-free networks, US airports have hubs which have an abundance of connections (number of flights) with other airports.

**Modularity and Community Detection**

I was very interested in finding out if the airline network exhibited clustering in ways other than hubs. Unfortunately, the edge attributes do not offer much information in this area. Each of the vertex attributes are basically different ways of labeling their location (by code, city/state or actual location). However, the edge attributes have much more aspects that flights could cluster by.

I hypothesized that Carrier or Aircraft Type would be the most likely edge attribute to show modularity. I explored the modularity of the network. I could not get my code to work for the Carrier attribute for some reason, but after running for all other edge attributes, Aircraft Type showed slightly higher modularity. Not very high, but certainly different than the other edge attributes.

Knowing this, I tried several community detection algorithms. I was only able to get the Walktrap, Label Propagation, Leading Eigenvector, and Louvain algorithms to complete. All algorithms increased the modularity of the network (i.e. explained more of the clustering). The best performing algorithm was Louvain as far as modularity. Leading Eigenvector returned the smallest number of clusters, but the clusters it drew are hard to interpret; if not nonsensical. It plotted many smaller clusters inside one supercluster. I’ve plotted all four results below.

**A picture containing text

Description automatically generated**

A close up of a logo

Description automatically generated

**A close up of text on a white background

Description automatically generated**

**A close up of text on a white surface

Description automatically generated**

**ERGM**

I also wanted to attempt ERGM on my airport network. After creating the null model, I tried to create different probability distributions using each of the options presented in class, but unfortunately, only gwesp worked.

Gwesp seemed to be yield a significant and meaningful predictor (low p-value with a coefficient of 5.4). The AIC value for this model went down, but still remained very high overall. The bright side is that the goodness-of-fit chart showed that gwesp definitely explained various node degrees, edges and triangles much better than the null model.

**Data Source**

“USAirports” from the igraphdata package in ‘R’

Most of this information was downloaded from The Research and Innovative Technology Administration (RITA). See http://www.rita.dot.gov/about\_rita/ for details. The airport position information was collected from Wikipedia and other public online sources.