Introduction

This dataset from Taxpayers for Common Sense captures Congressional appropriations earmarks for the 2010 fiscal year. For each appropriations bill in 2010, it captures which congressmen were sponsors of the bill, and who received the earmark. It also provides the state and party of each congressman, among other data.

I pulled out the sponsoring congressmen and recipient institutions for each appropriation bill, and analyzed the relationships between them. About half of the appropriations bills were filtered out from my analysis because they did not have any identified recipient in the data. I also had to normalize many of the recipient names because identical institutions were identified in different ways (e.g. Lockheed Martin, LMCO, Lockheed, Lockheed Martin Missles & Fire).

Overview -- Looking at Congressmen regions and earmark recipients

I first began to look at the entire earmark network, which is about 9,000 nodes total. I graphed Congressmen and recipients by first aggregating congressmen vertices into whole states, and then coloring these 50 states by each of the 4 major regions (Northeast, Midwest, South, West). I then filtered each graph by a single region to produce each region's network. Congressmen from the South and North do the most earmarking overall, as shown below. From these 4 filtered comparisons, you can see that the states from the South region and the Northeast regions are the most numerous and centrally connected. Vertices are sized by betweenness centrality, and filtered slightly by eigenvector and those with an edge weight greater than 1. The Harel-Koren Fast Multiscale layout algorithm is used.
Because of the size of the earmark network, I then decided to filter the network to only recipients with the name "university". Below are some interesting headlines from it.

1. Pelosi (CA) and Abercrombie (HI) buck the trend and belong to another group entirely apart from their state peers

When funding universities, almost all senators within a state are grouped together, with the exception of Pelosi from CA and Abercrombie from HI. They instead work more closely with congressmen from other earmark groups. Below, all labeled vertices are congressmen, which are sized by out-degree and grouped by color using the Clauset-Newman-Moore cluster algorithm. They are also filtered slightly by betweenness centrality to filter out the “small” bridges in the network. You also notice some interesting groupings of states.
Pelosi (CA) and Abercrombie (HI), circled in red, are the only two congressmen who belong to groups outside of the rest of their state peers. This network, which is laid out using the Harel-Koren Fast Multiscale algorithm, captures earmarks for university recipients. Vertices are sized by out-degree and they were filtered slightly by betweenness centrality.

Yale Medical, University of Hawaii, Arkansas State University, and Jackson State University are the top four most connected vertices in the earmark network. These three universities have the highest betweenness centrality, higher than many congressmen, which shows how influential they are in the earmark network. They clearly know how to get money from the US government. In this graph, we color the congressmen by chamber (blue=house, maroon=senate), and do very slight filtering (0.0001) by eigenvector centrality, to highlight the
more important connections.

Yale Medical, University of Hawaii, Arkansas State University, and Jackson State University are all influential academic institutions in the earmark network. They clearly know how to get money from the US government. Vertices are sized by betweenness centrality, which helps to highlight our headline. Congressmen are colored by chamber (blue=house, maroon=senate), and we do very slight filtering (0.0001) by eigenvector centrality, to highlight the more important connections.

3. Unexpected state partners such as ND and WA, HI and TN, CO and GA, and MI and NV help keep the earmark network cohesive

There are interesting groups of states that pop out which are unrelated to geography, such as ND and WA, HI and TN, CO and GA, and MI and NV. Who knew the congressmen from these states would work so closely
together. Interestingly, as the betweenness centrality is raised, the overall network keeps its cohesion and does not split into smaller networks. One could make the argument that enough senators maintain enough cross-state connections to keep the overall network intact. This might be necessary in order for senators to gain support for their own earmarks, in a kind of “I pat your back, you pat mine” mentality. Very few senators go it alone with earmarks.

This graph captures the earmark groupings of university recipients. You'll notice some really interesting congressmen partnerships unrelated to geography, such as ND and WA (forest green circles), HI and TN (lime green circles), and CO and GA (pink circles). These cross-state connections seem to play a role in the extraordinary cohesion of the network.

Critique of NodeXL
Pros

- I like how the dynamic filtering pane auto updates the graph. This makes it easier to discover interesting trends more quickly.

Cons

- I was unable to find the "group-in-a-box" feature. I had to ask someone where it is.
  - **Suggestion**: Rework the menu items on the ribbon to make it more obvious
- In Autofill Columns, under Vertex Color, Options, if you select the source column values as a category type, you are unable to choose the colors you want the different categories to be.
- Import from Open Workbook is a pain to work with because the size of the window can't be expanded, and there usually are a lot of columns and rows to display.
- It took me a while to understand how to multi-select vertices. The convention is to use a CTL-click to multi-select rather than a separate arrow.
- There are too many places to choose your vertex color, and it's unclear which place takes precedence. For example, if you have Vertex color filled out in the spreadsheet, but choose to group, group colors take precedence.