Congressional Influence on Student Debt

*Katherine Dix, Juan Padilla, William Archer*

**Introduction**

Note: The information that would be included in the summary.txt file has been put into this introduction

For our final project, we decided to analyze the relationships between people in Congress in order to find the most influential congresspersons. In Congress, the person who presents a new bill is called the sponsor. The people who support the bill are called cosponsors. Using the GovTrack.us database, we created a weighted directed graph using information from 600 congressional bills on student debt from 2000 to the present day, where nodes are congresspersons. For example, if congressman X sponsors a bill and congressman Y cosponsors it, there will be a directed edge from Y to X with a weight of 1. If X sponsors four bills and Y cosponsors each of them, the edge would have a weight of 4. With our graph, we used a variety of calculations to determine spheres of influence, similarities to different congresspersons, and absolute importance.

Type of Project: This is an empirical analysis project. We incorporated graph algorithms and social network algorithms.

Work Breakdown

William Archer: scraped the govtrack.us website for information about bills, sponsors, and cosponsors

Katherine Dix: created the weighted directed graph and used graph algorithms to find the most influential people

Juan Padilla: designed/used social network algorithms and tests to analyze the cluster coefficient/neighborhoods of the most influential people

Everyone worked on this paper

Dataset

All information came from [www.govtrack.us](http://www.govtrack.us). We do not have a link to the entire dataset that we used, since we got the information from scraping the website.

Note: We have also submitted some JUnit test cases that were used for this project.

**Methods**

Indegree Centrality

Indegree Centrality was used to calculate the sponsors with the largest set of co-sponsors, which would correspond to a larger neighborhood. Usually, degree centrality is calculated by dividing the degree of a node by the largest degree possible in a graph. Since this is a directed graph, we only took the indegree into account to calculate the number of cosponsors a congressperson has divided by the number they could possibly have.

PageRank

PageRank was used to calculate absolute importance of congresspersons by ranking them relatively among their peers. We tweaked the PageRank algorithm slightly to account for the fact that this graph is weighted. In the PageRank algorithm we learned in class, every node disperses its PageRank equally through each of its outgoing edges. We distributed the PageRank proportionally to the weight of the edge. For example, if a node A has two outgoing edges, X and Y, and the edge from A to Y has a weight of 2 and the edge from A to X has a weight of 1, two-thirds of A’s PageRank would be distributed to Y and one third would be distributed to X.

Node Strength

Node Strength was used to calculate the nodes with the highest weights. While indegree centrality will only measure the number of distinct cosponsors, node strength will take into account whether the cosponsors are sponsoring a person multiple times.

Neighborhood Overlap (outdegree and indegree)

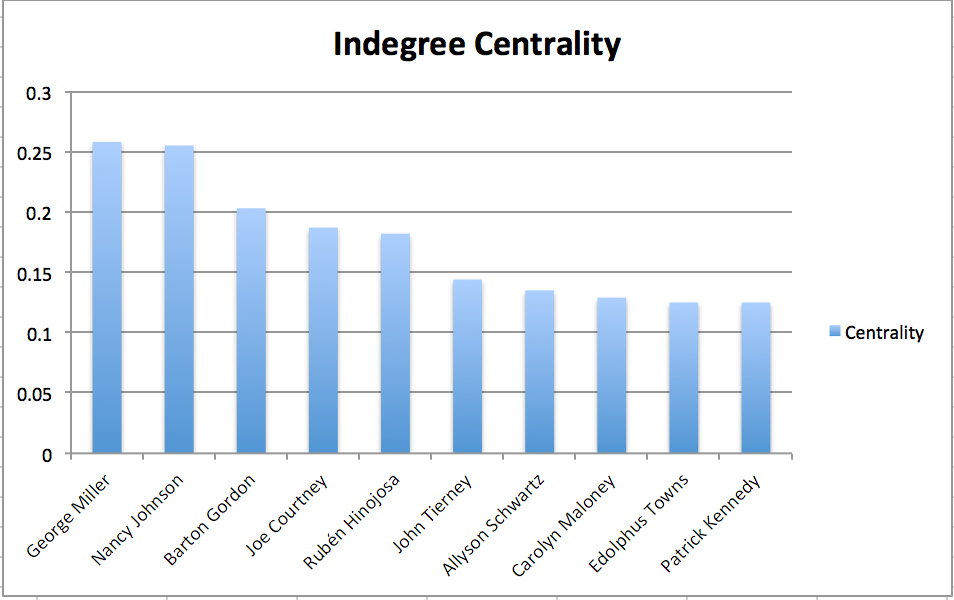
Neighborhood Overlap was used to calculate the congresspeople who had the most similar ideologies under the assumption that nodes with high neighborhood overlap would influence similar people because of a common push.

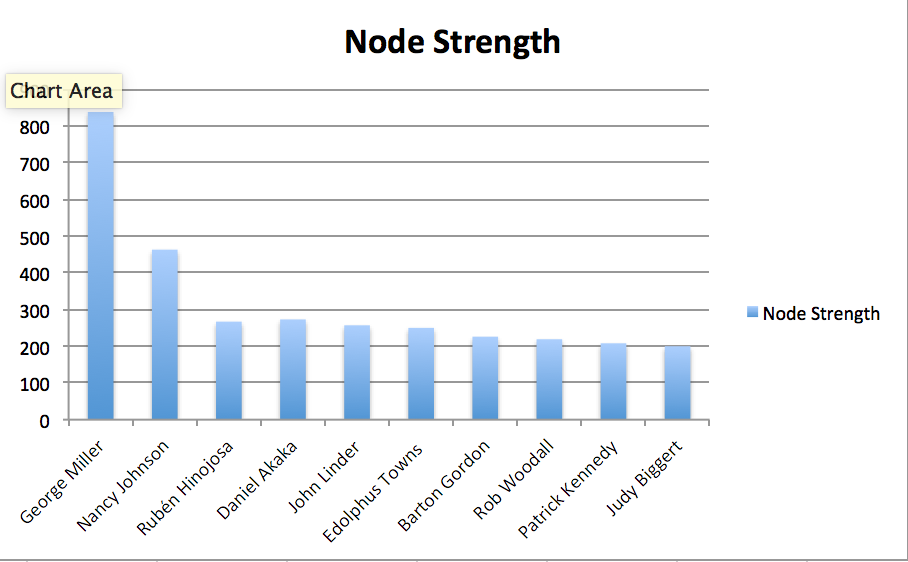
Clustering Coefficient

Clustering coefficient was used to calculate how dense spheres of influence on congresspeople were, which could show pockets where similar people only ‘preach to the choir.’

**Results**

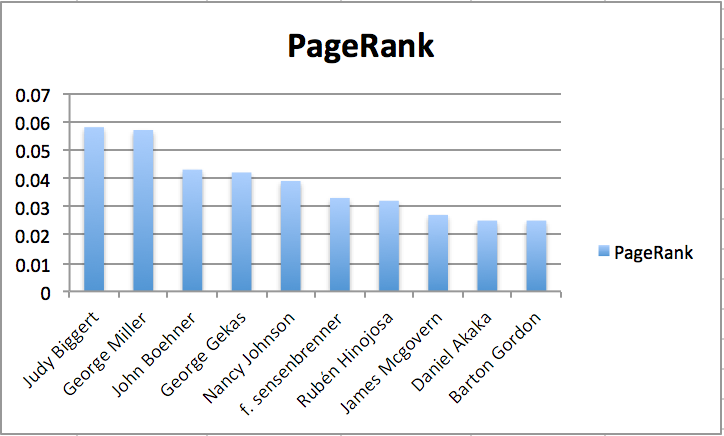
We used a combination of indegree centrality, node strength, and PageRank to find the most influential people in Congress with regards to student debt. First, we compared the results of the indegree centrality and the node strength. We wanted to find people who had both high indegree centrality scores and high node strength scores. These would be the people who had a large numbers of cosponsors that supported multiple bills each.





George Miller, Nancy Johnson, Barton Gordon, Rubén Hinojosa, Edolphus Towns, and Patrick Kennedy are in the top ten for both node strength and indegree centrality, suggesting that these are the most prominent people. George Miller appears to be the most powerful coming in first place both times. Nancy Johnson comes in a close second for indegree centrality but trails far behind in node strength. This must be because, even though George Miller and Nancy Johnson have similar numbers of cosponsors, George Miller’s cosponsors are supporting him multiple times, which creates larger edge weights and results in a higher node strength score.

Next, we analyzed the graph using PageRank to see if the results from the PageRank algorithm would confirm the results from the indegree centrality and node strength algorithms. The results from the PageRank algorithm made sense given the previous results.



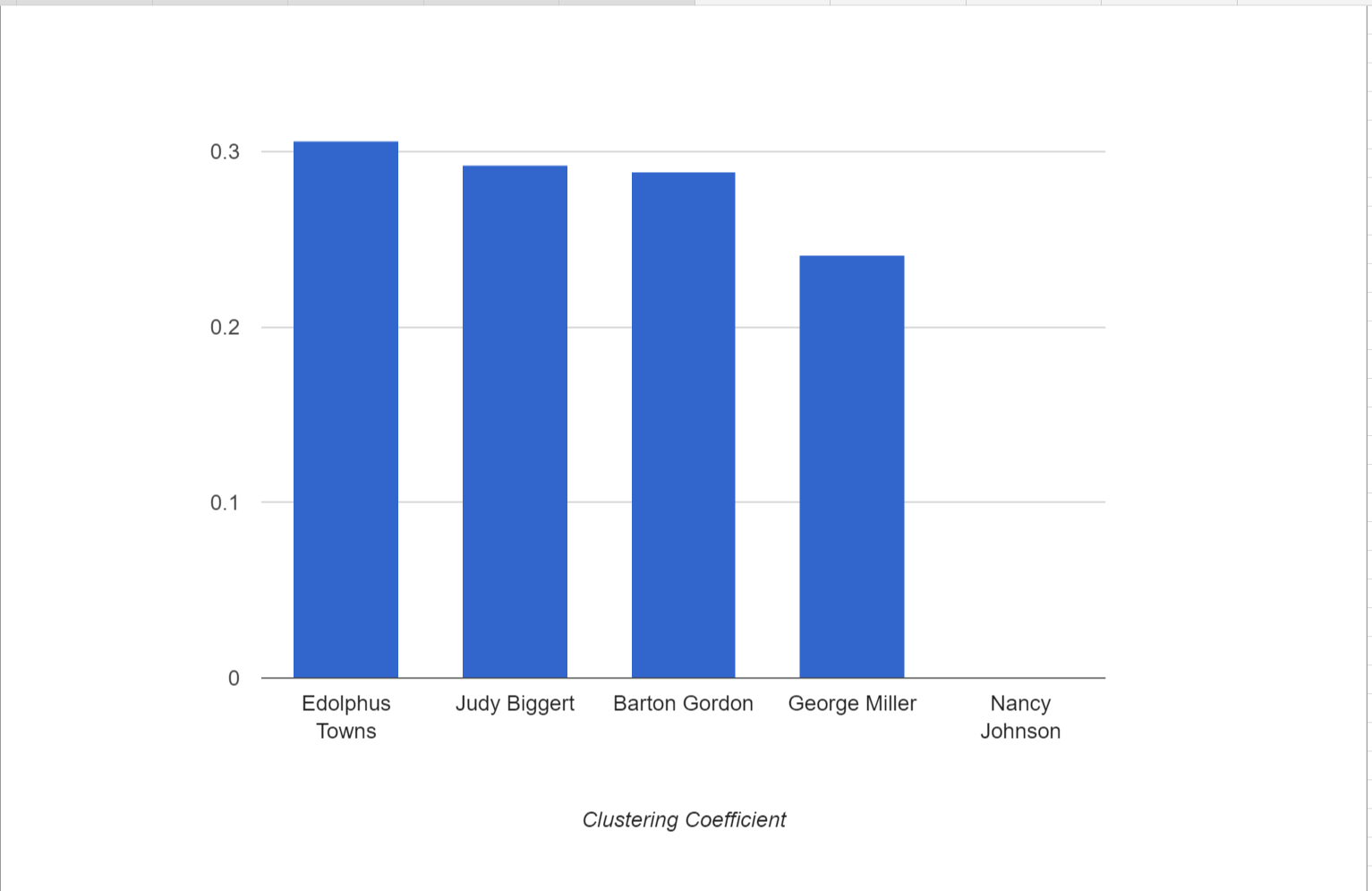
George Miller came in a close second place to Judy Biggert, however, his high scores across the board show that he is probably the most powerful with regards to student debt. Rubén Hinojosa, Barton Gordon, and Nancy Johnson also ranked in the top ten for node strength, indegree centrality, and pagerank. However, it is also important to take into account the length of time each person has spent in office. Our data is based on bills starting in 2000. George Miller was in office from 2000-2015. Other people in the graph who were not part of Congress for as long would have a hard time achieving Miller’s node strength and indegree centrality scores.

Looking at the most influential people in terms of PageRank rankings, such as George Miller, Nancy Johnson, among others, we decided to use the concept of neighborhood overlap to calculate the congresspersons who had the most similar ideologies, and whether influential congressmen had similar sponsors

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Neighborhood  Overlap (x and y) | **George**  **Miller** | **Nancy Johnson** | **Barton**  **Gordon** | **Edolphus**  **Towns** | **Judy**  **Biggert** |
| **George**  **Miller** | X | 0.0 | 0.172 | 0.205 | 0.0 |
| **Nancy Johnson** | 0.0 | X | 0.0 | 0.0 | 0.0 |
| **Barton**  **Gordon** | 0.172 | 0.0 | X | 0.25 | 0.105 |
| **Edolphus**  **Towns** | 0.205 | 0.0 | 0.25 | X | 0.0 |
| **Judy**  **Biggert** | 0.0 | 0.0 | 0.105 | 0.0 | X |

We noticed that the congressmen with the highest overlap were Edolphus Towns and Barton Gordon, and Towns and George Miller. This means that congressmen like Edolphus Towns shares a lot of the co-sponsors that influential people like Miller and Gordon also possess, even though Towns himself is not as influential in terms of PageRank compared to the others.

As far as clustering coefficient among the most influential, most of them did have dense spheres of influence as shown in this graph, with Edolphus Towns having the highest coefficient and Nancy Johnson surprisingly with zero coefficient.



**Conclusion**

Through this project, we were able to analyze the relationships between members in Congress to determine which have the most influence. The graph algorithms and social network algorithms provided pertinent information about the members in this group. To check our work, we returned to the govtrack.us website and looked at the profiles of the most influential members.  Govtrack’s information validated ours. For example, 31% of the bills George Miller sponsored are about education and 43% of the bills Rubén Hinojosa sponsors are about education. Using formulas like clustering coefficient further reinforced the fact that politicians like Miller had great influence in Congress. For example, Miller had around 0.25 clustering coefficient, which is a pretty high number and right up there with other influential congressmen like Barton Gordon. Neighborhood overlap also depicted the alliances between politicians like Miller and Gordon, both sharing a 0.172 overlap. Through this project, we learned about how we can use graphs and social networks to gain interesting information about groups of people.