Network of Political Books (June 2019)

Jason Huggy, Student, Lewis University

Abstract— This study looks at a network data set focused on a variety of political books before the 2004 presidential election. The books fall under three categories: conservative, liberal, or neutral. Using network analysis technology, each book is used as a node and edges are created to show the connections between each book. The strength of each node is measured using its degree, measuring the number of links to each node. The resulting graph shows a clear separation between conservative and liberal books and displays the books that are most heavily read in each political party.

I. INTRODUCTION

Right now, everyone is getting ready for the 2020 presidential elections, and the Democratic party will soon select its candidate to square off against the current President of the United States, Donald Trump. The political parties, including third parties, are ramping up their campaigns and the citizens of the US are trying to figure out what candidate best fits their ideals. Many books on both sides of the aisle have come out over the last few years discussing their side of the argument, but this is nothing new. In 2003, the US was getting ready to vote whether to keep President George W. Bush in office or to go with his Democratic opponent, Senator John Kerry. Like today, many books were released in the years leading up to the 2004 Presidential Elections.

This study looks at the books that were co-purchased on Amazon.com leading up to the 2004 elections. The data set was put together by Mark Newman as a way to align each book in the study to a political ideal, and then look at what books were purchased together [1]. This study serves as a way to understand the books that most heavily influenced the voters of the 2004 Presidential Elections.

A. Degree Versus PageRank

In this study each node, or book, is analyzed to determine the number of links attached to it. The more books that a book is linked to, the stronger its rank. While the PageRank algorithm is typically a very useful method to rank each node in a network, this study was too simple in a sense for it to be used. PageRank usually looks at how many links each node has, but also looks at the strength of the nodes linking to it [2]. For example, if a node has three links but only one is an in-link, then it is not as strong as a node with three links with three in-links. In-links specify a page, like a webpage, that links to the target node. Out-links specify that the target node includes a link to another page. In general, the more in-links, the better the node's rank. To even further rank a node one can also measure the strength of the nodes that are linked to it. If a node linked to the target node also has a good number of in-links, then it will improve the PageRank of the target node.

While PageRank is the preferred method to rank each node in a network study, in this study it is not needed. The only information presented is whether each book was bought with another book. Therefore, there is no way of telling whether a link is an in-link or an out-link. Each link is un-directed. For this reason, this study only uses the degree to rank each node [3]. The degree is simply the count of the links that each node has. The more links, the greater the degree.

II. METHODOLOGY

This study focuses on 105 books leading up to the 2004 Presidential Elections. Each book is considered to have some sort of political tie to politics, and is categorized as a conservative, liberal, or neutral book.

A. Software

The graphing of the network of books used in this study was graphed using Gephi, version 0.9.2. The data for this study was uploaded into the software and nodes and edges were assigned.

B. Nodes

Each book in the data set is labeled as a node. Each node is assigned a value categorizing it as either conservative, liberal, or neutral.

C. Edges

Each co-purchase of books on Amazon.com is labeled as an edge. This study does not focus on how often a pair of books was bought together, rather how many links each book has with other books. Edges are not directed, as books were bought as pairs. Meaning there is no say of what book led the buyer to buy the other book. All that is known is that if there is a link between each book that they were bought together at some point.

D. Creating the Network

Two layouts were used in the graphing of the network. The first layout used to graph the network was the Fruchterman Reingold layout. This layout essentially puts the network into a circle, which allows for the network to be easily visualized; but the layout also tries to keep linked nodes centralized to the nodes they are linked to. The second layout used is the Yifan Hu layout, which clusters each node based on its links. Nodes that are linked are closer together, while nodes that are not linked are further away from each other. Each node was then assigned a color based on its value: red for conservative, blue for liberal, and green for neutral. Each node was then sized using Gephi's degree feature. The larger the node, the higher the rank. The label of each node coincides with the book's title.

III. RESULTS & DISCUSSION

A. Results

After graphing the system of nodes there is a very clear separation between political ideals. Generally, conservative books are bought with conservative books and liberal books are bought with liberal books. There are exceptions and neutral books are bought with books from both political ideals. In Figure 1, the circle is cut in half. The red nodes take up one half and the blue nodes take up the other half. The green nodes are inter-mixed in the middle. In Figure 2, there is a conservative cluster and a liberal cluster. In between the two clusters are the green nodes, but there are also blue and red nodes which shows that they had links with neutral or opposite political ideals. Overall, the average book had 8.4 links. The highest number of links books had was 25, *A National Party No More* and *Off With Their Heads*, and the lowest, *The Future of Freedom*, only had 2 links.

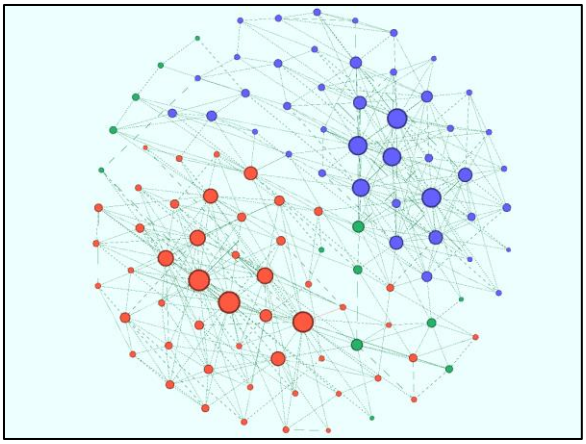

Fig. 1. Network using the Fruchterman Reingold layout. Red nodes are conservative, blue are liberal, and green are neutral.

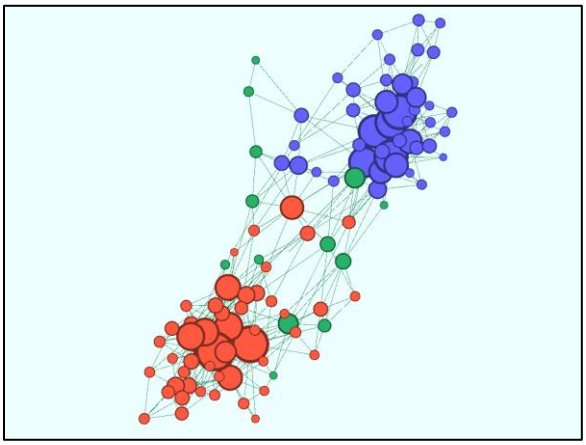

Fig. 2. Network using the Yifan Hu layout.

B. Discussion

In general, the results of this study turned out in a manner that is expected. Customers are more likely to buy two books that coincide with the same political ideology. However, there are liberal and conservative books in this study that have links with nodes of the opposite ideology. As can be seen in Figure 1, the nodes on complete opposite sides of each other are positioned there because they have no links with the opposite party. These are, most likely, more extreme books in the liberal or conservative view. Books that are in the middle of the split of the circle tend to have links with the opposite political ideology. The same phenomenon can be seen in Figure 2, where there are several red and blue nodes that wander from the tight clusters of each political ideology.

This may be because of two reasons. The first reason could be because these books may not focus too heavily on political ideals but be listed as conservative

or liberal due to the author of the book. For example, if former President Bill Clinton, a democrat, wrote a book on his life as president, that may not deter an idealistically conservative person from wanting to read his book. After all, he was still their president. The second reason for this occurrence could be due to "middle of the road voters". This includes citizens that are new to politics or are not sure which political ideals best fit them. For instance, a customer might buy a liberal and conservative book to see which principles best fit their beliefs. This means the book would identify as a respectable introductory book for each political ideology. After a person is committed to an idea, they are likely to extend their reading to more in-depth or extreme books that agree with their party.

The larger nodes in the graph represent the books that had the most links with other books. Essentially, the larger the node, the more popular the book. If a book is more popular, then it is more likely to be rated well by others. This would allow for Amazon's recommendation system to recommend these popular books to people that have similar books in their cart. As can be seen in Figure 1 and 2, these larger nodes are very centralized within their side of the political ideology. Therefore, it is likely that extreme, middle of the road, or average conservatives/liberals are all most likely to purchase one of these books with any other book they may get.

The green, neutral nodes are also worth examining. While some of the green nodes are directly in the middle in both graph representations, some of the green nodes seem to favor one side. This could show that even though the books are considered neutral, they are more favored by liberals or conservatives. This could also hint that the original author of this data was not exactly sure how to classify some of the books, so some could be mistakenly classified.

IV. CONCLUSION

In closing, this study looked at 105 political books from before the 2004 Presidential Elections and tried to determine which books were most commonly purchased with other books of the same political ideology. The books were graphed and represented as a network of links. Each link represented a couple that was bought together on Amazon.com. A clear separation was realized between the two political ideologies. It was also obvious which books are most often recommended and purchased within each ideology. The study was successful in classifying which books were most commonly bought, and which books most likely had the biggest impact on voters ahead of the 2004 elections. However, more research should be conducted to see what made these books so popular, and what characteristics lead to a more influential book.

V. References

[1] V.Krebs. UMICH [Online] Available at: http://www.orgnet.com/. [Accessed 22 Jun. 2019].

[2] *Pagerank Explained Correctly with Examples*. Princeton University. [Online]. Available: https://www.cs.princeton.edu/~chazelle/courses/BIB/pagerank.htm. [Accessed: 22-Jun-2019].

[3] *Quick Start*. Gephi. [Online]. Available: https://gephi.org/users/quick-start/ [Accessed: 22 Jun. 2019].