United States Congress Relations According to Liberal and Conservative Newspapers

Basim Mahmood  
BioComplex Laboratory, Computer Sciences  
Florida Tech, Melbourne, USA  
bmahmood2011@my.fit.edu

Ronaldo Menezes  
BioComplex Laboratory, Computer Sciences  
Florida Tech, Melbourne, USA  
rmenezes@cs.fit.edu

Abstract—The media have always played an important role in society: it acts as catalyst of information for most people. With the advent of the 24-hour news channels, people became accustomed to having access to the information the media outlets provide anytime, anywhere. In this paper, we analyse the social network of politicians in the United States from the point of view of two newspapers believed to have different political views: conservative and liberal. The idea is to reveal the relationships between US Congress members and Senators according to social network extracted from the newspapers. The newspapers we have chosen are from the same state (New York) to minimize differences in the network due to location. We selected the New York Times (NYT) which is believed to a liberal newspaper and the New York Post (NYP) which is believed to be a conservative newspaper. Our results indicate a significant difference between the social networks formed which means that the public is exposed to a different view of the US politicians depending the media they choose to read. The social network also confirms that the NYT focuses more on people from the Democratic party while the NYP focuses on Republicans. We believe this is the first kind of work that uses Network Science to establish bias in news media.

I. INTRODUCTION

The United States of America is a divided country when it comes to politics. The recent (November 2012) elections clearly demonstrate this divide showing that the population of the country is split 50-50 between republicans and democrats—the two main political parties. At the same time, there are many studies that demonstrates media bias to one of the other side of the spectrum [1], [2], [3], [4]. The relations between media bias and election as well as votes in congress is well known but one issue that has not received much attention relates to classifying media according to their bias.

The USA legislative (popularly known as the Congress) is the bicameral legislature of the federal government consisting of the upper house or the Senate (with 100 members), and the lower house or the House of Representatives (with 435 members). Two main parties control United States Congress: the Democratic and the Republican Parties. At the time of this study the current congress was referred to as the 113th Congress of the United States which had a Democratic majority in the Senate and a republican majority in the House of Representatives (as shown in Figure 1).

Fig. 1. The 113th Congress of the United States. US Senate with 53 democrats, 45 republicans and 2 independent senators; and the US House of Representatives with 200 democrats, 233 republicans and 2 current vacant seats. (images from Wikimedia Commons).

Newspapers in the United States are an integral part of the culture of the country and have been published since the 18th century with the introduction of the New England Courant by James Franklin [5]. The New York Times (NYT) and the New York Post (NYP) are one of the most important newspapers in the United States. According to the alliance for Audited Media, the daily circulation of NYT as of March 2012 was 1,586,757 issues (3rd in the USA) while daily circulation of NYP in 2012 was 555,327 (6th in the USA) [6]. This paper looks to see if the social network formed from the co-occurrence of names of US members of congress in newspaper articles differs from one newspaper to another.

Recent advances have allowed us to discern social relations from text [7]. There are many instances in which unstructured text is the only source of information between individuals. Newspapers form a great source of information about the relations between politicians because most newspapers have sections related to politics. Nowadays, most newspapers are represented on the WWW. There have been a few works looking at newspaper data as source of social studies (e.g. [8]).
Even relations between companies have been studied where the relations are created from text analysis of newspapers. [9].

Social Network Analysis (SNA) is a useful tool for studying relations between individuals. In simple terms, SNA consists of a collection of graph-analysis methods that many researchers developed to analyze networks of individuals who are connected based on their communication pattern, personal interaction, etc. [10]. Social networks represent individuals, groups, organizations, and related systems that are tied in one or more types of interdependencies, social contact, financial exchanges, and group events, among numerous other aspects of human relationships [11]. Politics influence many aspects of our life. Not surprising then, we have seen investigations of several aspects of politics and in a closer context of this work, American politics. The literature contains many works looking at public opinion [12], people’s voting behavior [13], the diffusion of policy ideas [14], bill sponsorships in the legislature [15], party factions [16], and other empirical phenomena such as causality [17].

One issue that has not yet been investigated in depth is the issue of media bias in politics. Obviously most people believe media outlets are biased but can we demonstrate this using SNA? In this work, we extracted the political news from two different newspapers: one believed to support Democrats (Liberal), the NYT; and another believed to support Republicans (Conservative), the NYP. If the newspapers report politics in a similar fashion, then their networks should be similar.

II. DATASET PREPARATION

The two networks used in this paper were extracted from the electronic version of two newspapers: the NYT and the NYP. The choice for these publications is justified in many ways:

- We wanted two newspapers that represent the same geographical area because they should have similar content being reported; in this case the newspapers are from the same city.
- The two publications are popularly believed to be liberal (the NYT) and conservative (the NYP).
- In order to build better relationships we wanted publications that are published daily.

Note that one of the goals of this work is to demonstrate/confirm media bias using the social network analysis of the politicians they report about. The second characteristic above is fundamental for us because we would like to see if the network confirmed the popular believe between the bias of the NYT and the NYP.

By using text analysis we extracted a dataset for each of the newspapers. The data extraction process has two stages. The first stage consists of getting the political news URLs for the NYT and the NYP and then using a Data Extraction Software to get two separated text files that represent the content of the newspaper. The second stage was performed by a Java program designed to process the text files and extract the relations between the members of congress and the weight of each relation. This program has three inputs (two coming from the first stage and the third is a text file containing the list of names of all members of congress). The Java program outputs two datasets representing two networks, one for each newspaper. The datasets were extracted for the period of September 18, 2012 to October 18, 2012 and the number of articles used for both newspapers added 674 articles. These articles had some noise because of the advertising on the pages; we removed the noise and used only the text related to the articles.

Networks represent nodes and the relationships between the nodes. In our case, the nodes are known and given by the list of US congressmen (from the House and the Senate). The relationship however has to be extracted from the newspapers’ text. In the extraction process we describe the output as containing the relations and their weight. This information comes from the articles themselves. Two nodes (congressmen) are related if they are mentioned in the same article. The underlying assumption is that if they are mentioned the same article, the newspaper believes they belong to the same story, or that they are part of the same context in which the article is set. This process of finding relations is repeated for each article. Hence if an article mentions the name of X, Y and Z, the extraction tool creates links between X–Y, X–Z and Y–Z, in other words, a clique is created connecting X, Y, and Z. This approach is again not new, other authors have used a similar approach; it is believed that people interacting with each other are likely to talk about the same topics or be mentioned in the same context [18].

Given that we are analyzing many files, it is possible to have repeated pairs. For instance, the relation between X and Y could be also found in a different article. When this is the case we increase the weight of the relation to reflect the multiple occurrences. After all is done the dataset consists of a list of undirected relations in the form \(<Member1>,<Member2>,<Weight>\). The final output is in the GDF format which has been proposed as the standard for GUESS (The Graph Exploration System).

III. NETWORK ANALYSIS AND RESULTS

Given the two networks, we set to identify differences between them. The first step was to try to visualize the networks. Figure 2 depicts the giant component of both networks. Republican members are shown in red while democrats are shown in blue.

One can note that visually these networks look very similar which might lead us to believe that both newspapers have similar social network of congressmen, however this argument could not be farther from the truth. First, the members represented in each figure are not the same and the weight of their connections are also different. Visualization does not help much in this case. So, we make a another comparison by providing a table with some important characteristics of each network; we used three of the many social network measures available [19] in order to find the differences between the networks (shown in Table I). These measures as follows:

ATASET II. D

REPARATION
Fig. 2. Connected component of the social network of members of congress according to an analysis of the text of the NYT (left) and NYP (right).

- **Average Degree**: The average degree describes the typical connectivity of a member of congress. The networks for both the NYT and the NYP are quite sparse. The average degree does not show significant difference, with the NYT being slightly more connected.
- **Average Clustering Coefficient**: This measure reflects the transitivity of relations between nodes, or the occurrences of cliques of size 3 in relation to the total number of cliques of size 3 that could exist in the network. The coefficient for the NYP network is higher than that for the NYT.
- **Average Path Length**: This shows how “short” the network is. The network for the NYP is “shorter” than the network for the NYT, which means that, on average, pairs of nodes are generally closer to each other in the NYP network than NYT.

There are also differences in the distribution of members with at least one connection. The NYT tends to have more democrats while the NYP shows more republicans (as shown in Table II). Although slight we can see already a bias in both publications.

<table>
<thead>
<tr>
<th>Networks</th>
<th>Avg. Degree</th>
<th>Avg. Clustering Coef.</th>
<th>Avg. Path Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>NYT Network</td>
<td>0.852</td>
<td>0.064</td>
<td>4.847</td>
</tr>
<tr>
<td>NYP Network</td>
<td>0.863</td>
<td>0.102</td>
<td>4.296</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table II</th>
<th>Statistics about the number of connected members due to the party.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network</td>
<td>No. of Democrats</td>
</tr>
<tr>
<td>NYT Network</td>
<td>84</td>
</tr>
<tr>
<td>NYP Network</td>
<td>72</td>
</tr>
</tbody>
</table>

Given the network analysis did not show many differences, we have decided to look at node measures for each network. We looked at the following measures for the 10 best connected individuals in each network:

- **Betweenness**: Total number of shortest paths that pass through a particular node; this metric is one possible indicator of who the most influential members in the network [10], [19].
- **Degree**: Members who have more ties to other members may be advantaged positions. Each network shows the connections between the members from the same party and with the other party [19].
- **Clustering Coefficient**: Clustering coefficient of a node is a measure of the degree to which that particular member is clustered (part of cliques of size 3) [20].
- **Closeness**: Defined as the sum of distances to all members. A member is considered important if he/she is relatively close to all other members [19].

It is easy to see that the analysis at the node level shows a huge bias between the newspapers. The most important individuals in the NYT network (most connected) are from the Democratic party, while the most important in the NYP network are from the Republican party (see Figure 3).

Table III shows the name of the top 10 best connected individuals for the NYT network ranked by betweenness, while Table IV shows the top 10 best connected for NYP network. The tables make it clear that people in the House of Representatives are more mentioned in newspapers than Senators, this may be due to the fact that they tend to sponsor more bills.

**IV. Conclusion and Future Work**

In this paper, we worked on relationships between members of the USA congress extracted from two well-known newspapers. The network revealed significant differences in the view of these newspapers given that the most connected
members found in each network have no overlap. This work shows that media bias may be measured and maybe even quantified according to the social network they reveal. One may notice the extraordinary fact that two newspapers present two completely different views of the political scene in the USA for exactly the same period of time.

In this work, we have focused on two well-known newspapers (the NYT and the NYP) because they are from the same city and should, in theory, discuss similar issues particularly when the data was extracted from the same time period. As we said above, the outcome of our study could not be further from the expected. The two newspapers present different views of the political scene in the USA.

There are many open problems that we would like to investigate in the future. Our work in this paper immediately brings to mind the question of the true nature of the networks extracted from the newspaper text. It is a known fact that congressmen collaborate on the sponsorship of bills and one could use this relation to create a social network. We will in the near future perform an analysis of the newspapers’ networks and compare them with the one generated from collaborations in congress. This in fact may lead to the possibility of ranking newspapers according to the true nature of the news they publish. If we assume the social network from co-sponsorship of bills is the ground truth, could the proximity of a newspaper network to the ground truth be an indication of the quality of the news being reported?

Last, we intend to reverse the question of this paper and look at many newspapers around the country and classify them according to their political leanings given that we now know that the network reflects the leaning. What are the conservative newspapers? What are the liberal ones? Are there newspapers that are more impartial than others? We expect the characteristic of the social network built from them could tell us more precisely their political views.

REFERENCES