2018 IEEE International Conference on Data Mining Workshops (ICDMW)

# Congressional Vote Analysis using Signed Networks

Extended Abstract

Tyler Derr
Data Science and Engineering Lab
Michigan State University
derrtyle@msu.edu

Jiliang Tang
Data Science and Engineering Lab
Michigan University
tangjili@msu.edu

## ABSTRACT

In today's era of big data, much can be represented as a network. However, most of the work in traditional network analysis is unable to handle many existing network types, which is due to certain networks having added complexities. For example, signed networks, which have both positive and negative links, have been shown to require dedicated efforts due to the methods designed for typical unsigned networks (those having only positive links) being no longer applicable. One specific type of signed network is that of voting records, such as the Senate and House of Representatives from the U.S. Congress, which form signed bipartite networks between the congresspeople and the bills voted upon. With the current tensions between the two prominent political parties in the U.S., it seems time to ask the question if signed network analysis methods are able to aid in our understanding of the underlying dynamics of the voting habits in the U.S. Congress, since they drive some of the most influential decision making processes in the country. To this end, in this paper, we conduct a thorough analysis on the behaviors of both current and past U.S. Congress voting datasets uncovering numerous patterns, extending and then investigating the applicability of balance theory in the signed bipartite setting, and then finally leverage our findings to accurately predict the sign of missing links.

## KEYWORDS

Signed Networks, Balance Theory, Signed Bipartite Networks, U.S. Congress

## 1 INTRODUCTION

Much of the data we produce today can be represented as networks. However, not all networks are the same, and many networks exist having complexities added to their structure that prevent traditional network analysis methods from being directly applied. Signed networks, which have become increasingly ubiquitous, are one such type that have benefited from dedicated efforts due to having differing principles and properties as compared to unsigned networks.

Simultaneously, the political science field has started to more widely adopt the use of network analysis in understanding their data. However, most efforts have only taken positive links into account (i.e., unsigned networks). One prominent political science area of study is towards the understanding of underlying dynamics driving how and why U.S. congresspeople vote the way they do. Naturally these voting networks of the U.S. congress form signed bipartite networks, where the bills being voted on form one node type, while the congresspeople form the other, where the voters can vote "yea" or "nay", which can be represented as positive and negative links, respectively. In fact, signed bipartite networks have been primarily overlooked. Previous works in signed networks have mostly focused on the unipartite setting. Therefore, if we seek to apply signed network analysis methods to the political science domain (specifically the U.S. Congress datasets), we need to ensure that the theories typically used, such as balance theory, are applicable in this signed bipartite network setting.

Balance theory [2, 3] is a fundamental signed network social theory that suggests a cycle in signed networks having an even number of negative links is balanced, while an odd number results in an unbalanced cycle. Many previous works have validated the theory in unipartite signed networks [4–6] and shown when harnessed the theory can aid in many real-word network analysis tasks [6]. However, in unipartite signed networks, balance theory is typically applied to the smallest cycles, i.e., signed triangles. In contrast, signed bipartite networks inherently do not contain any triangles and also contain two types of nodes. Therefore it is important to first validate and understand balance theory in signed bipartite networks before applying to the political science domain. Thus, dedicated efforts are desired for signed bipartite networks in addition to unipartite signed networks and unsigned bipartite networks.

In this paper, we perform an initial investigation of balance theory in undirected signed bipartite networks. As aforementioned, balance theory has been utilized to advance numerous tasks in unipartite signed networks and sign prediction is the one being benefited most. Hence, we then investigate how to utilize balance theory to boost sign prediction in the U.S. Congress datasets that are naturally signed bipartite networks. This paves the way for using balance theory for other network analysis tasks in signed bipartite networks and also for leveraging signed network analysis methods to obtain a better understanding of the voting behaviors of the U.S. congress.

## 2 DATA ANALYSIS

In this section we perform an analysis on extending balance theory to the signed bipartite setting.

In signed networks one of the most fundamentally studied social theories is that of balance theory [2, 3], which discusses the settings in signed networks that are socially "balanced" (i.e., expected), and those that are unexpected due to the social tensions involved


*WOODSTOCK'97, July 1997, El Paso, Texas USA*

ACM ISBN 123-4567-24-567/08/06.
https://doi.org/10.475/123_4

2375-9259/18/$31.00 ©2018 IEEE
DOI 10.1109/ICDMW.2018.00218

**Table 1: Signed Butterfly Statistics.**

| Signed Butterfly Isomorphism Classes | U.S. Sentate | | |
|---|---|---|---|
| | Count | % | $E\%$ |
| $(A)$ $(+,+,+,+)$ | 13404168 | 0.262 | 0.094 |
| $(B)$ $(+,-,-,+)$ | 5595440 | 0.110 | 0.122 |
| $(C)$ $(+,+,-,-)$ | 9404006 | 0.184 | 0.122 |
| $(D)$ $(+,-,+,-)$ | 5537080 | 0.108 | 0.122 |
| $(E)$ $(-,-,-,-)$ | 6815324 | 0.133 | 0.040 |
| **Balanced** | 40756018 | **0.797** | 0.500 |
| $(F)$ $(+,+,+,-)$ | 6225745 | 0.122 | 0.302 |
| $(G)$ $(+,-,-,-)$ | 4118075 | 0.081 | 0.197 |
| **Unbalanced** | 10343820 | **0.203** | 0.500 |

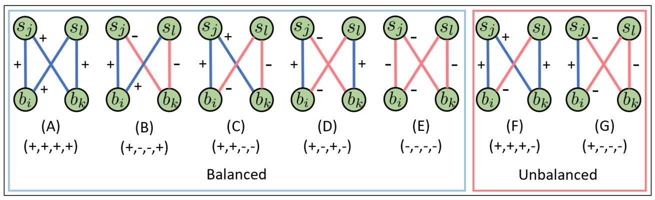


**Figure 1: Undirected Signed Butterfly Isomorphism Classes.**

in maintaining "unbalanced" and seemingly unnatural links. In recent signed network analysis works balance theory is usually investigated and then applied towards many tasks[6], but almost always in the form of triangles (or cycles of length 3) in a unipartite signed network. Furthermore, it is also unknown its applicability towards a bipartite setting where nodes can be of different types. Thus, we will introduce how we plan to extend the usage of balance theory to the smallest signed cycles (i.e., butterflies) in undirected signed bipartite networks.

## 2.1 Signed Butterfly Isomorphism Classes

In unsigned bipartite networks, one commonly investigated structure is that of a "butterfly" [1], which is a cycle of length 4. More formally, a butterfly is the simplest cohesive higher-order structure and also a complete biclique. Thus, this provides the most natural structure to investigate as a possible extension for balance theory in signed bipartite networks.

Just as there are different types of signed triangles, there are different types of signed butterflies. In Figure 1 we present the 7 non-isomorphic undirected signed butterflies; note that there are five that adhere to balance theory while only two are categorized as unbalanced. We use the notation $(*,*,*,*)$ to denote a signed butterfly isomorphism class that represents the links between the bills and senators $(b_i, s_j, b_k, s_l)$ (in that order with the last sign connecting $s_l$ and $b_i$). Note that we use "senators" here for simplicity, but the house representatives can be substituted in their place. The simplest of types are $(+,+,+,+)$ and $(-,-,-,-)$, which denote the classes having all positive or all negative links, respectively, and both are balanced due to having an even number of negative links (and can be seen in Figures 1(A) and 1(E), respectively). Next, we have $(+,+,+,-)$ and $(+,-,-,-)$ the two unbalanced classes of signed butterflies (since they have an odd number of negative links). In Figure 1(F) we have the signed butterfly isomorphism class that encompasses all the signed butterflies with a single negative link. We can observe that no matter where this single negative link is placed, we always have one buyer with two positive links, one bill with a positive and negative link, and similar structure for the two senators. The isomorphism class $(+,-,-,-)$ can be seen as the complement (if defined as swapping link signs in a signed network) of the class $(+,+,+,-)$ and defined in a similar way, but with swapping the positive and negative links in the definition. This leaves the signed butterflies having two positive and two negative links, of which we have three isomorphism classes. In Figure 1 we see the class $(+,-,+,-)$ is used to represent signed butterflies where all bills and senators have one positive and one negative link in their cycle. When one of the bills has two positive links, while the other bill has two positive links, we observe in Figure 1(B) that both senators have a single positive and single negative link, and define the isomorphism class of $(+,-,-,+)$. Finally, the last type of signed butterfly has both bills connected positively to one senator, and negatively to the other, which we represent as the class $(+,+,-,-)$ shown in Figure 1(C).

## 2.2 Signed Butterfly Analysis

In Table 1 we report our analysis after counting the number of signed butterflies for each isomorphism class as shown in Figure 1. We also calculated the percentage of each and the expected number of that type (i.e., "$E\%$") when randomly shuffling the link signs.

We first observation that the large majority of signed butterflies in our U.S. Senate dataset (consisting of the 1st to 10th senate votes) are indeed balanced. Furthermore, there are significantly more balanced than expected based on the link sign ratio in the given network (i.e., comparing columns "%" and $E\%$). The second observation is that all unbalanced signed butterflies. Similarly, across all datasets the $(+,+,+,+)$ and $(-,-,-,-)$ signed butterflies are significantly overrepresented, further strengthening the applicability of balance theory in signed bipartite networks. However, the isomorphism classes involving two positive and two negative links appear to not always be found overrepresented. For example, the class where all buyers and sellers have one positive and one negative link, i.e., $(+,-,+,-)$, is less commonly found than expected across all three datasets.

## 3 CONCLUSION

In this work we have focused our attention on applying signed network analysis to the political science domain. More specifically we have investigated the use of balance theory to provide a better understanding of the U.S. Congress datasets. Future work will consist of other interesting questions, e.g., can we predict when congresspeople will abstain, be a swing voter, or how aligned they are with others in congress?

## REFERENCES

[1] Sinan G Aksoy, Tamara G Kolda, and Ali Pinar. 2017. Measuring and modeling bipartite graphs with community structure. *Journal of Complex Networks* 5, 4 (2017), 581–603.
[2] Dorwin Cartwright and Frank Harary. 1956. Structural balance: a generalization of Heider's theory. *Psychological review* 63, 5 (1956), 277.
[3] Fritz Heider. 1946. Attitudes and cognitive organization. *The Journal of psychology* 21, 1 (1946), 107–112.
[4] Jérôme Kunegis. 2014. Applications of Structural Balance in Signed Social Networks. *arXiv preprint arXiv:1402.6865* (2014).
[5] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. 2010. Signed networks in social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1361–1370.
[6] Jiliang Tang, Yi Chang, Charu Aggarwal, and Huan Liu. 2016. A survey of signed network mining in social media. *ACM Computing Surveys (CSUR)* 49, 3 (2016), 42.