Cluster Boosting and Data Discovery in Social Networks

*

Maria Tomasso
Texas State University
San Marcos, Texas, USA
met48@txstate.edu

Lucas J. Rusnak
Texas State University
San Marcos, Texas, USA
lucas.rusnak@txstate.edu

Jelena Tešić
Texas State University
San Marcos, Texas, USA
jtesic@txstate.edu

ABSTRACT

We introduce a new ground-truth recovery boosting approach on antagonistic networks using the status-influence space obtained by the frustration cloud — a generalization of the frustration index which models nearest consensus-based states of a signed graph. A spectral clustering and k-means approach are both examined and are compared to existing clustering methodologies on two sentiment-based datasets. We demonstrate that our approach successfully recovers all community labels on a highly modular dataset and outperforms the leading clustering technique by a factor of 3.08 on a more complex network. Additionally, we demonstrate that status and influence, in combination with network data, can be used to detect and characterize anomalous outcomes in promotion networks.

ACM Reference Format:
Maria Tomasso, Lucas J. Rusnak, and Jelena Tešić. 2022. Cluster Boosting and Data Discovery in Social Networks: . In *Proceedings of ACM SAC Conference (SAC'22).* ACM, New York, NY, USA, Article 4, 7 pages. https://doi.org/10.1145/3477314.3507243

1 INTRODUCTION

1.1 Motivation

How do we analyze the abundance of publicly available survey data? We have witnessed the analysis of survey data from multiple statistical and visual perspectives. Very few trends were discovered beyond data summary: survey data does not reveal much information when looking at 1 dimensional data projections such as rise of cases vs. specific policy or rise of cases vs. political leaning of the region. In this paper, we go beyond statistical survey data analysis and look into the (i) considerations for the knowledge graph constructions from survey data within context, and (ii) knowledge graph analysis using consensus attributes. We outline the considerations, and the graph construction on multiple datasets, and demonstrate clear use of novel knowledge graph analysis by uncovering Wikipedia Elections data anomalies.

*The full version of the paper is available as `acmart.pdf` document

SAC'22, April 25 –April 29, 2022, Brno, Czech Republic

ACM ISBN 978-1-4503-8713-2/22/04…$15.00
https://doi.org/10.1145/3477314.3507243

In this paper we examine three data sets - Highland Tribes [11], Sampson's Monastery [13], and Wikipedia Elections [8]. We apply conventional clustering methods for signed graph and show that they fail to capture community structure in all but the highly modular Highland Tribes dataset. Then, we introduce a new method for boosting ground-truth recovery of conventional clustering algorithms that is robust to graph construction and demonstrate that it either meets the standard or out-performs state-of-the-art clustering methodologies for signed graphs on Highland Tribes and Sampson's Monastery. Finally, we retool the results of [12] to detect anomalies in the Wikipedia Elections dataset.

1.2 Paper Organization

Section 2 recalls the new techniques for sentiment analysis via the frustration cloud from [12] and the relevant metrics of *status* and *influence*. Section 3 presents the straightforward mapping of Highland Tribe Survey data [11] to signed graph, and compares community discovery to consensus analysis on small well defined network. Section 4 outlines the Sampson's Monastery incident [13], difficulties the researchers are facing when attributes do not fully capture the outcome in sociological experiment, and proposes a way to overcome it. Section 5 focus on the outcome analysis of Wikipedia Administrator Election voting data and the final outcomes [8] in spatio-temporal way, and characterizes the peculiarities in the final outcome. Section 6 concludes the paper with a summary.

2 SOCIAL NETWORKS, SIGNED GRAPHS, AND BALANCE

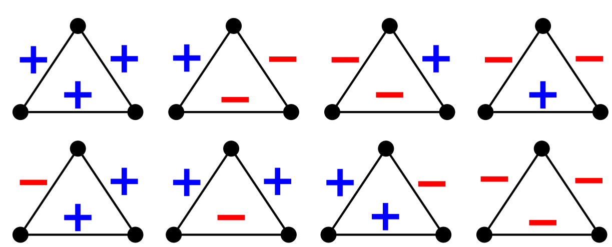

Figure 1: Heider's Triadic Relations: The top are balanced, the bottom are unbalanced.

In this section we describe prior related work, balance theory of signed graphs, a graph balancing algorithm based on spanning trees, and graph attributes derived from discovering minimal balancing states from [12]. We also describe contributions of this paper when graph balancing is used for survey and spatial data signed graph construction. Signed graphs are graphs where the edges store the attitude of a vertex (person, county, tribe) towards another. Two vertices connected by an edge can be agreeable or antagonistic: edge is labeled +1 if they are agreeable or −1 if they are antagonistic towards another.

2.1 Related Work

Some datasets naturally align with the mathematical definitions of a graph, such as a network of roads and intersections, airports and routes, or friendship/enmity in a social group. In these cases, the dataset can be converted to a graph structure with little to no preprocessing or ambiguity in the construction. Additionally, analysis of these 'intuitive' graphs is fairly straight-forward. In other cases; however, the construction of the similarity graph is a matter of debate and the methods used can have a profound impact on the final analysis.

Prior work has demonstrated that conventional clustering techniques for signed graphs fail on knowledge graphs that either encompass multiple sentiments or are large and sparse [14]. We test three spectral clustering approaches describes in [6] that are based on the signed Laplacian matrix (lap_none), symmetric Laplacian matrix (lap_sym), and symmetric separated Laplacian matrix (lap_sym_sep). We also include balance normalized cut [2] techniques that use the signed Laplacian (BNC_none) and normalized signed Laplacian (BNC_sym) as kernels and a novel spectral method solved via a generalized eigenproblem (SPONGE) [3] as well as its symmetric variant (SPONGE_sym). Finally, we include methods based the geometric means Laplacian (GM) [9] and matrix power means Laplacian (SPM) [10]. We assess clustering success with the adjusted Rand index (ARI) by comparing generated community labels to known ground-truth values.

Most of these methods rely on spectral methods rather than a structural approach. In [12], a method of characterizing vertices using repeated sampling of near-balanced states is presented. This technique provides the metrics status and influence, discussed further in section 2, which can be used to project the vertices of a graph into 2D space. In this paper, we examine using the status-influence projection for clustering.

2.2 Social Balance

Social balance theory [5] and the mathematical foundation of signed graphs [4] were the first to define and model balance in signed networks. A signed graph is *balanced*, i.e., in a global consensus state, if every cycle comprises an even number of antagonistic edges. This is the case because two antagonistic edges always cancel each other out i.e. a friend of my friend is also my friend, and an enemy of my friend is my enemy, as illustrated in Figure 1. Formally, a *signed graph* Σ is a pair (G, σ) that consists of a graph $G = (V, E)$ and an edge-signing function $\sigma : E \longrightarrow \{+1, -1\}$. A signed graph is *balanced* if the sign of every cycle is positive [1, 4]. If the graph Σ is not balanced then there exists a set of edges whose sign reversal produces a balanced signed graph called a *balancing set*.

Figure 2 illustrates a simple example of an unbalanced signed graph Σ (left), and all possible spanning trees T associated with that unbalanced graph (right) with darker edges. Given a connected graph G (ignoring the signs) and a spanning tree T, for each edge e outside of spanning tree T the subgraph $T \cup e$ contains a unique

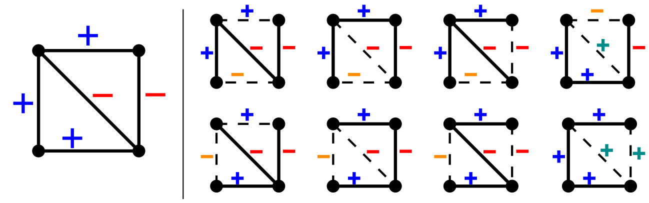

Figure 2: A small unbalanced signed graph example Σ (left) and its balanced states obtained from the frustration cloud (right). The spanning trees are represented with solid edges and inherit their signs from the original signed graph. Signs outside each tree that are changed from negative to positive appear as teal, while positive to negative appear as orange. Unchanged signs remain blue and red.

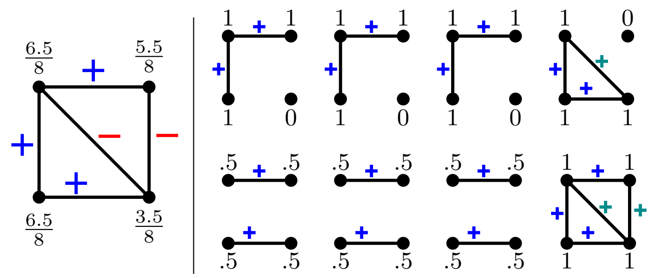

Figure 3: Vertex status measure (left) computed from consensus states (right). The deletion of negative edges results in the partition for all consensus states from Fig. 2. The nodes are labeled by majority value for sign graph Σ (right), and the normalized sum of all scores per vertex results in the status (left).

cycle, called the *fundamental cycle of T using e*. The set of fundamental cycles for each spanning tree forms a basis for the cycle space of the graph. Moreover, the number of edges outside a spanning tree is the dimension of the cycle space. In the event a graph is not connected we examine each connected component individually. In [12] the fundamental cycles were used to reconstruct all the cycles of the graph, but each edge outside of a spanning tree was re-introduced to produce a balanced version of the original signed graph. For Σ in Figure 2, there are 8 spanning trees of G, as depicted by the solid edges. For each edge e outside of spanning tree T the subgraph $T \cup e$ contains a unique fundamental cycle, and edge e is assigned whatever sign makes the fundamental positive. The edges outside of each spanning tree is emphasized by using a dashed line in Figure 2. Each spanning tree inherits the sign for the original signed graph, while the signs of the edges outside each spanning tree are possibly changed to produce a balanced sign graph.

Figure 3 (right) illustrates the resulting consensus components obtained by the deletion of the negative edges from each balanced state as guaranteed by Harary's Theorem [4]. For each part of these partitions we assign each vertex a value of 1 if its part contains more vertices, and a value of 0 if it contains fewer; in the event of a tie all vertices are assigned a value of 0.5. These values are then averaged over all spanning trees used to provide a percent value called *status* that represents the likelihood a vertex will belong to a majority consensus outcome. Given the signed graph from Figure 2, Figure 3 depicts the status values of each vertex along with the assigned vertex values for each balanced state. The dual concept of status for edges is called *agreement*, and the average agreement of the edges around a given node is called the *influence* of the node. From [12] it is known that the that the status of a vertex is always greater than or equal to its influence, and the resulting status-influence cone exists in the first half of the first quadrant.

Much of signed graph analysis focuses on the smallest size balancing set called the *frustration index* of a signed graph, which, in general is NP-hard to compute. In [12] these limitations were relaxed to a minimal (via containment) balancing set (not necessarily one of minimum size). This allowed for multiple different consensus outcomes to analyzed and aggregated to provide a more robust understanding of the original signed graph. We use the frustration cloud and the status and influence metrics to assess signed spectral clustering for identifying communities and anomalies.

3 COMMUNITY DETECTION ON HIGHLY MODULAR SIGNED GRAPHS

In this section, we demonstrate the effectiveness of the community discovery of state-of-the-art approaches on the highly modular benchmark dataset Highland Tribes [11] and compare them to clustering in the status-influence space.

The Highland Tribes dataset describes agreeable and antagonistic relations between 16 tribes of the Eastern Central Highlands of New Guinea. Each tribe is modeled as a node, with agreeable tribes connected by a positive edge and antagonistic tribes connected by a negative edge. There are three known communities in this dataset as shown by the colored nodes in Figure 4.

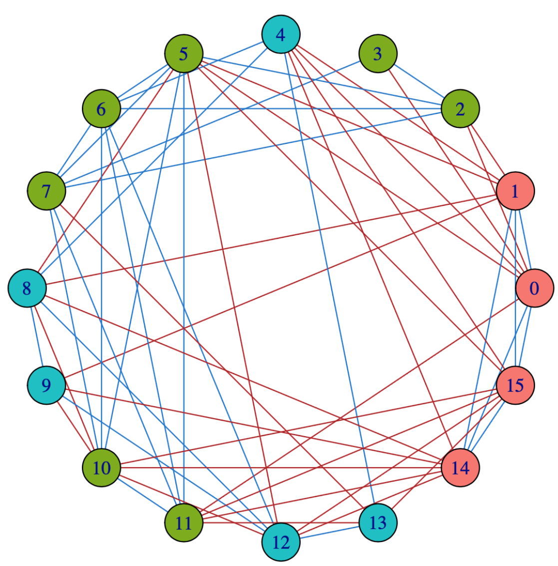

Figure 4: The Highland Tribes network with positive and negative edges.

Conventional community detection techniques are highly effective at recovering ground-truth labels on this dataset due to its inherent modularity: most positive edges occur within communities while most negative edges occur between communities. We applied nine clustering techniques in Figure 5 (blue): signed spectral clustering [7] using the signed Laplacian, normalized signed Laplacian, and normalized separated signed Laplacian; balance normalized cuts [2] using the signed Laplacian and normalized signed Laplacian; *SPONGE* and $SPONGE_{sym}$ [3]; clustering via geometric means of Laplacians [9]; and clustering via matrix power means [10]. As shown in Figure 5, we were often able to achieve 100% ground-truth recovery as measured by the adjusted Rand index (ARI), a similarity metric ranging from -1 to 1 in which 1 represents a perfect labeling, 0 represents random labeling, and -1 represents a maximally incorrect labeling.

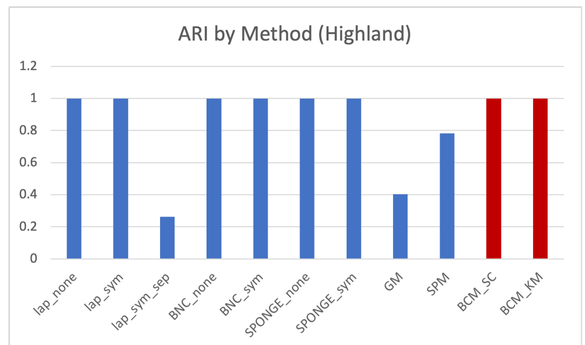

Figure 5: The Highland Tribes Adjusted Rand Index (ARI) clusterability measure of signed spectral clustering methods on Highland tribes.

Next we map the nodes from Highland Tribes into 2D space using two metrics from the method introduced in [12], status and influence. As shown in Figure 6, status and influence effectively separated the nodes based on ground-truth community membership. We exploit this separation by applying two known clustering methods, spectral clustering (BCM_SC) and k-means clustering (KM_SC), to the status-influence coordinates before assessing the generated community labels using the ARI. As shown in red in figure 5, both BCM_SC and BCM_KM both succeeded in fully recovering the ground-truth labels.

4 SIGNED GRAPH ANALYSIS FOR COMPLEX SURVEY OUTCOMES

In this section, we describe how we extended the balancing theory analysis of the signed graphs when applied to survey data. We show signed graph construction and analysis using Sampson data, and how status-influence characterize complex communities when spectral clustering methods fail.

Sampson's monastery describes social relationships between eighteen novice monks in St. Anthony's, a New England monastery, between 1966-1967 [13]. Ground truth is present in this dataset because four social groups were identified throughout the study: (i) the Loyal Opposition, the group of monks who began first; (ii) the Young Turks, the group who joined after the Loyal Opposition and during the period of change associated with the Vatican II; (iii) the

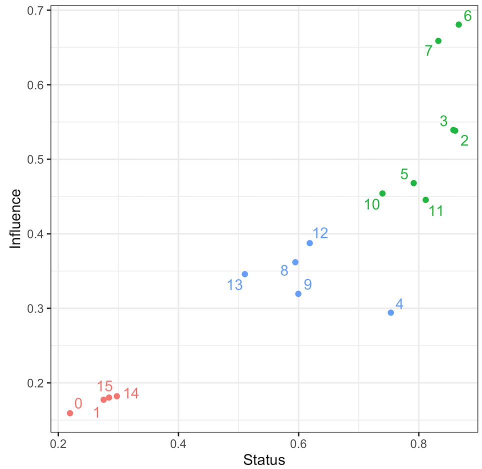

Figure 6: Ground truth communities for Highland Tribes in the status-influence space.

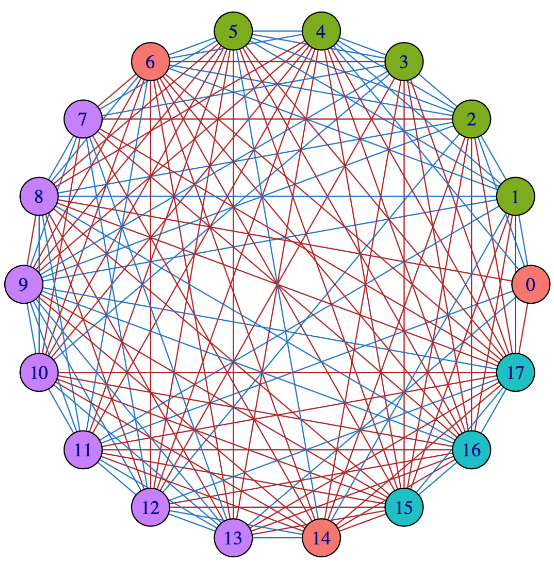

Figure 7: Sampson data analysis: graph labeling of edges and ground truth labels of Sampson nodes.

Outcasts, who were the monks not accepted by either the Loyal Opposition or the Young Turks; and (iv) the Interstitial Members, who did not take a clearly defined stance. Data was collected through observations, interviews, and questionnaires. Four categories of relations were recorded: (i) *affect*: "various forms of cathectic orientation relationships in which the object of ego's orientation is an alter"; (ii) *esteem*: "composed of forms of cognitive orientations of alters", (iii) *influence*: "composed also of forms of cognitive orientations of alters, but relating to alter's impact upon ego's experience or behavior", and (iv) *sanctioning*: "comprised of cognitive orientations of ego regarding his direction of positive and negative sanctions toward alter." During each survey, the novices were asked to rank

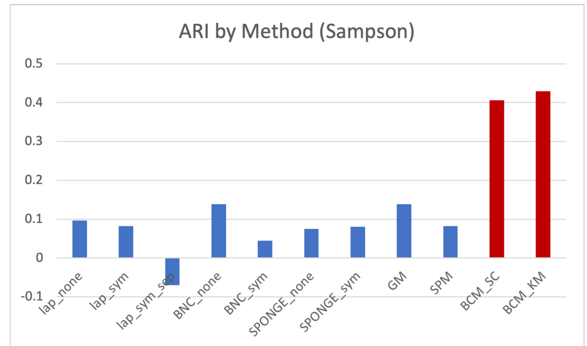

Figure 8: Sampson data analysis: ARI recovery score of ground-truth labels for Sampson using various signed clustering techniques.

their top three and bottom three Brothers in each category. For this experiment, we only use data from the fourth survey and combine the four categories into a single +1/-1 sentiment for each pair of monks. To accomplish this, we sum the adjacency matrices for each category of relations where the i, j entries represent the directed score between i and j, i.e., if i ranks j as their first choice for Affect, the i, j entry in the Affect matrix will be 3.

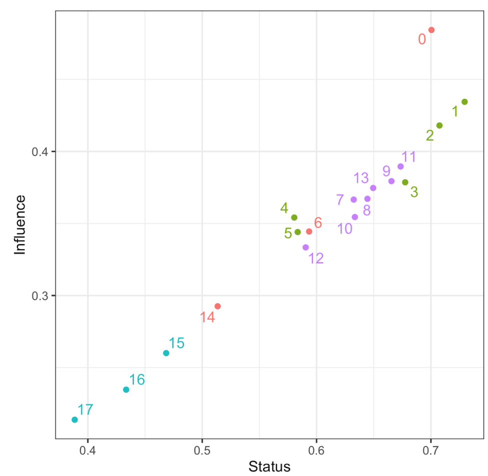

Figure 9: Sampson ground truth view in status influence space.

To construct a graph, we assign edge weights of +1 or -1 between monks based on the 8 matrices. We examine each category to assign an edge weight. If all categories are positive, we assign a +1 edge. If all edges are negative, we assign a -1 edge. To handle cases with mixed sentiments, we implement a threshold to determine the edge weight. Only 17 monk-monk relations had mixed sentiments. To calculate the threshold, we sum the scores across the four categories. The strongest mixed sentiment cases involve either 3 positive scores of 3 and 1 negative score of 1 for a total score of 8, or 3 negative scores of 3 and 1 positive score of 1 for a total score of -8. Thus, scores for mixed sentiments will range from -8 to 8. We divide mixed sentiment scores by 8 to normalize, and assigned any edge with a weight greater than 0.125 a sign of +1 and any edge with a weight less than -0.125 a sign of -1.

The edges in Sampson data are complex in that the weights capture multiple sentiments. We repeated the clustering experiment from Section 3 and found that none of the signed clustering methods were able to capture the nuance of this underlying grouping, as illustrated in Figure 8. Next, we mapped the nodes of the Sampson dataset into status-influence space and colored by ground-truth community label (Figure 9). We found that the Outcasts had both the lowest status and the lowest influence of the group, while the Interstitial Members were scattered throughout the space. The Loyal Opposition members, who entered the monastery first, consistently placed higher in both metrics than the Young Turks. Thus, we have recovered some useful information via balanced cut metrics that did not arise through the use of traditional signed graph clustering methodologies. We applied the same clustering routine used in 3 to the status-influence coordinates and found that, while ARI was still low, our methods recovered over three-times more information than the next-best algorithm. In conclusion, while conventional signed graph clustering techniques were generally unable to recapture the ground-truth community labels as shown in Figure 8 we were able to boost clustering success by first projecting the points into status-influence space.

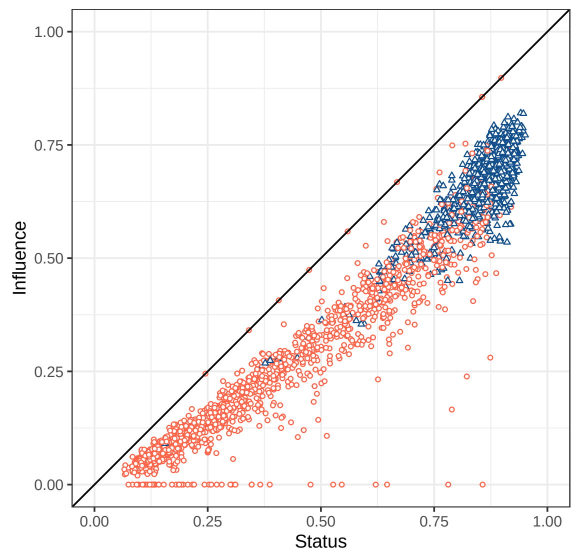

Figure 10: wiki elections projects in status/influence space.

5 WIKIPEDIA ELECTIONS ANALYSIS

The wiki-Vote data set was curated by Jure Leskovec, et al. and downloaded from the Stanford Network Analysis Project's (SNAP) repository of network data. It contains data on all Wikipedia adminship elections that took place before January 2008 and includes 2794 elections, 103,663 votes, and 7066 users. Roughly half of the votes were cast by current administrators while the rest were cast by registered editors. Of the 2794 elections, 1235 were successful for a promotion rate of 44.2 percent. The data set is presented as nodes and directed edges, where each node represents a user (either a voter, loser, or winner) and each edge represents a vote of Support, Oppose, of Neutral cast by one user with respect to another.

In [12] status and influence were used to distinguish between adminship and promotion to adminship, and the status-influence cone for promoted users appears in Figure 10. However, the community ground-truths are not as clearly defined. One possibility is to group based on existing admins, promoted to admins, or those not promoted; however the work in [12] indicates that there is a clear difference between influence and status in the detection in leadership and promotability, respectively, in this very sparse network. Another option is to bin by individual elections, but a user may belong to multiple elections in varying capacities and outcomes; this was addressed in [12] by cloning a user based on each election cycle, but has not been examined using our framework, and is planned for future work.

Wikipedia administrators are editors who have been granted the ability to perform special tasks. These tasks include blocking and unblocking users, IP addresses, and IP ranges; applying, removing, or modifying page protection; deleting pages meeting specific criteria; granting and revoking user permissions; viewing and restoring deleted pages; granting and revoking requested user permissions; restricting and restoring the visibility of information in individual logs and page revisions; editing protected pages; overriding the title blacklist; moving pages; and editing some pages in the WikiMedia namespace. Administrators are chosen through a *community review process* that seeks *consensus*, not a majority-rules decision. The following passage is taken from Wikipedia's policy documentation: *Decisions on Wikipedia are primarily made by consensus, which is accepted as the best method to achieve Wikipedia's goals, i.e., the five pillars. Consensus on Wikipedia does not mean unanimity (which is ideal but not always achievable), neither is it the result of a vote. Decision making and reaching consensus involve an effort to incorporate all editors' legitimate concerns, while respecting Wikipedia's policies and guidelines.*

To be eligible for adminship, a user must have an account that is at least 30 days old and have made at least 500 edits. Once this threshold is met, the user may file their own nomination or be nominated by another user which is known as request for adminship (RfA). Next, a discussion is opened where all users, including those without an account or not logged in, may comment and ask questions. Editors who are logged in can cast votes of Support, Oppose, or Neutral with a brief explanation of their rational. The discussion period generally lasts a minimum of seven days and ends with a *bureaucrat* (a user who has been granted privileges exceeding those of administrators) reviewing the discussion for evidence of consensus for promotion. In theory consensus does not need to be unanimous nor does it depend on a numerical threshold and is assessed by reviewing the comments attached to votes, but in practice we have found that RfA percentage is a strong predictor of promotion. An RfA's percentage is calculated by considering only Support and Oppose votes; Neutral votes are not included in the calculation but the attached comments are still considered to assess consensus. Since 2015, Wikipeda policy states that RfAs finishing

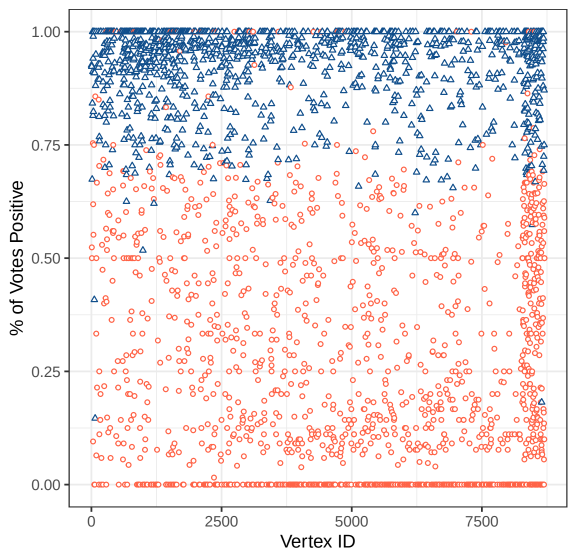

Figure 11: RfA versus vertex ID for Wikipedia Elections. Blue triangles were promoted, red circles were not promoted.

between 65 and 75 percent are subject to the reviewing bureaucrat's discretion and generally RfA's with a percentage below 65 percent do not result in promotion, but we have found that these guidelines hold for pre-2008 elections as well. Figure 11 depicts the RfA versus Wiki-ID colored by election outcome.

5.1 RfA Analysis and Outliers

Rather than predicting promotion, and due to the lack of a clear ground-truth, we are examining outliers through the lens of status and influence to identify anomalous elections. Non-promoted outliers were first examined by logistic model on outcomes with RfA as a predictor. Mis-classified points were separated based on outcome and projected into status-influence space, as depicted in Figure 12 and Figure 14.

The non-promoted logistic model outliers were traceable to three scenarios: mislabeled data in the original SNAP dataset when cross-checked via Wikipedia; the nomination was withdrawn by the candidate; consensus was not reached as determined by a bureaucrat. Their placement within the status-influence cone is shown in Figure 12. The accuracy of the logistic model was 0.981, the recall was 0.975, precision was 0.976.

Next, we created a linear regression model with RfA as the outcome and status and influence as predictors and high leverage points were labeled as outliers. A high leverage point indicates an unusual combination of independent variables. Since status and influence generally increase together, this could mean a high-status low-influence election or vice versa. The response variable, RfA in this case, will generally only play a minor role in determining leverage. Thus, we expected to see a different set of users and issues than those identified by the logistic model.

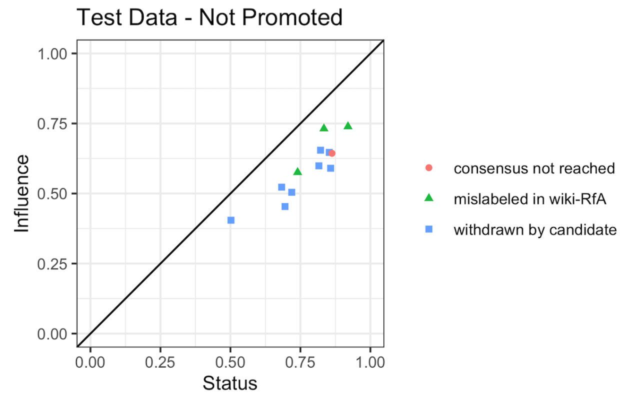

Figure 12: Non-promoted outliers from the logistic model, projected in status/influence space.

The non-promoted high leverage outliers in the linear model and their placement within the status-influence cone is shown in Figure 13. These failed elections were traced to a lack of consensus; withdrawn by either the candidate or bureaucrat; or were simply a very unsuccessful election. The most interesting finding is that the high leverage points are at the extreme end of the status-influence cone where the status-influence gaps are the largest.

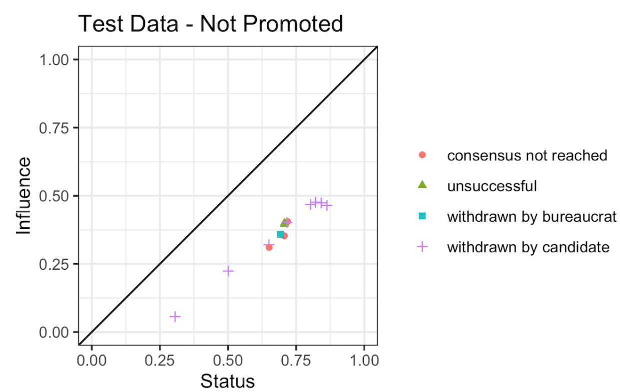

Figure 13: Non-promoted outliers from the linear model, projected in status/influence space.

The promoted logistic model outliers were traceable to five scenarios: inexperience; judgement concerns; lack of need for admin tools; removed by arbitration; or a spcific issue raised during the election cycle. Their placement within the status-influence cone is shown in Figure 14, the accuracy of the model was discussed previously for non-promoted users.

The promoted high leverage outliers in the linear model and their placement within the status-influence cone is shown in Figure 15. These failed elections were traced to inexperience or near or total unanimity. Again, the most interesting finding is that the high leverage points are at the extreme end of the status-influence cone where the status-influence gaps are the largest. This was present for both promoted and non-promoted users.

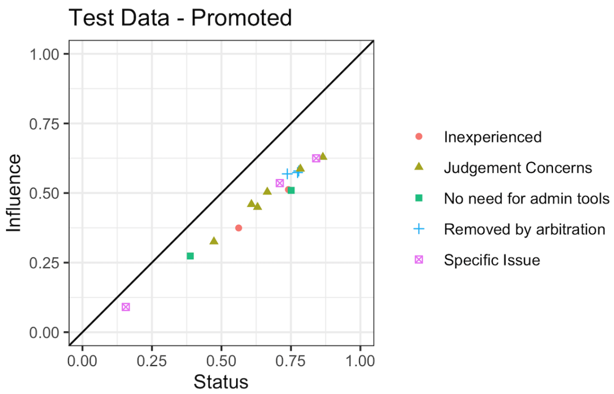

Figure 14: Promoted outliers from the logistic model, projected in status/influence space.

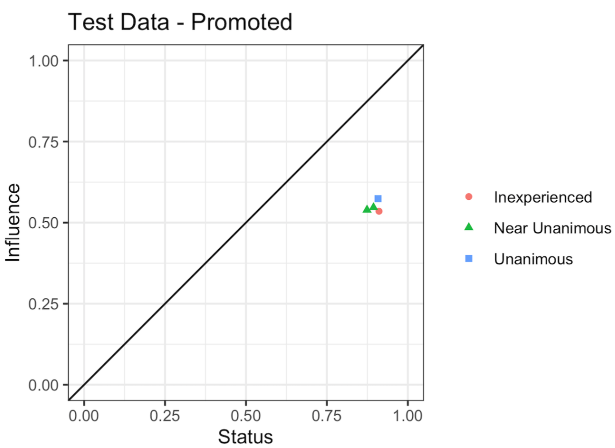

Figure 15: Promoted outliers from the linear model, projected in status/influence space.

6 CONCLUSION AND FUTURE WORK

In this paper, we expand on the notion of frustration cloud introduced in [12] and propose to model cognitive "correction" in a social network using generalized balance theory and a vector of measures of vertex and edge impact in a graph. The agreeable nature of the Highland Tribes dataset provided a proof of concept on using status and influence to supplement clustering approaches. The antagonistic nature of the Sampson dataset generally fails all clustering attempts, but status/influence clustering with basic spectral and k-means clustering were over 3.08 times more successful that prior attempts just by translating to the status/influence space. The lack of ground truth in the large sparse Wikipedia admin dataset was then examined for status and influence based outliers in both a logistic and linear model. Outliers for each model were analyzed, but the high leverage outliers in the linear model trended along the lower angular extreme of the status-influence space and needs further study. We hope to examine the possibility of discovering ground-truths on the Wikipedia dataset via this process to discern the nature of promotional distribution and their outliers. Additionally, we intend to continue exploring anomaly detection via consensus attributes by comparing our approach with state-of-the-art anomaly detection algorithms in future work.

REFERENCES

[1] D. Cartwright and F. Harary. Structural balance: a generalization of Heider's theory. *Psychological Rev.*, 63:277–293, 1956.
[2] Kai-Yang Chiang, Joyce Jiyoung Whang, and Inderjit S. Dhillon. Scalable clustering of signed networks using balance normalized cut. In *Proceedings of the 21st ACM international conference on Information and knowledge management - CIKM '12*, page 615, Maui, Hawaii, USA, 2012. ACM Press.
[3] M. Cucuringu, P. Davies, A. Glielmo, and H. Tyagi. SPONGE: A generalized eigenproblem for clustering signed networks. *arXiv:1904.08575 [cs, math, stat]*, 2019.
[4] F. Harary. On the notion of balance of a signed graph. *Michigan Math. J.*, 2:143–146, 1953.
[5] F. Heider. Attitudes and cognitive organization. *J. Psychology*, 21:107–112, 1946.
[6] Jérôme Kunegis, Andreas Lommatzsch, and Christian Bauckhage. The slashdot zoo: Mining a social network with negative edges. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09. ACM, 2009.
[7] Jérôme Kunegis, Stephan Schmidt, Andreas Lommatzsch, Jürgen Lerner, Ernesto W. De Luca, and Sahin Albayrak. Spectral analysis of signed graphs for clustering, prediction and visualization. In *Proceedings of the 2010 SIAM International Conference on Data Mining*, pages 559–570. Society for Industrial and Applied Mathematics, April 2010.
[8] Jure Leskovec and Andrej Krevl. SNAP Datasets: Stanford large network dataset collection. http://snap.stanford.edu/data, June 2014.
[9] Pedro Mercado, Francesco Tudisco, and Matthias Hein. Clustering signed networks with the geometric mean of laplacians. *Advances in Neural Information Processing Systems*, 12 2016.
[10] Pedro Mercado, Francesco Tudisco, and Matthias Hein. Spectral Clustering of Signed Graphs via Matrix Power Means. In *Proceedings of the 36th International Conference on Machine Learning*, pages 4526–4536, 2019.
[11] Kenneth Read. Cultures of the central highlands, new guinea. *Southwestern Journal of Anthropology*, 10(1):1–43, 1954.
[12] Lucas Rusnak and Jelena Tešić. Characterizing attitudinal network graphs through frustration cloud. *Data Mining and Knowledge Discovery*, 6, November 2021.
[13] S. Sampson. A novitiate in a period of change: An experimental and case study of relationships. *Ph.D. Thesis*, 1968.
[14] M. Tomasso, L. Rusnak, and J. Tešić. Advances in scaling community discovery methods for large signed graph networks. *CoRR (under review)*, 2021.