Comparison of Faction Detection Methods in Application to Ukrainian Parliamentary Data *

Thomas Magelinski[1], Iain Cruickshank[1][0000−0002−4205−5806], and Kathleen M. Carley[1][0000−0002−6356−0238]

Center for Analysis of Social and Organized Systems (CASOS),
Institute for Software Research, Carnegie Mellon University, Pittsburgh 15217
tmagelin@andrew.cmu.edu, icruicks@andrew.cmu.edu,
kathleen.carley@cs.cmu.edu

Abstract. In this study we compare two general methods of faction detection from Ukrainian Parliamentary roll call data, MacRae's method and Gower's method. Both methods were adapted to the special voting procedures and patterns of the Ukrainian Parliament, such as its non-binary voting scheme. Our analysis shows that each method is viable for faction detection individually, and that both can be used in tandem for results with higher confidence. Viability was demonstrated through the construction of the cooperation network between official parties, and by listing key parliamentarians based on their centrality in the factionized voting network. While the party-party network is intuitive, it was found that a pair of key actors in one faction were from opposing parties.

Keywords: Voting Analysis · Faction Detection · Clustering.

1 INTRODUCTION

Political factions impact political outcomes in many political systems. The formation and detection of factions can thus be very important in determining future responses of a body politic to different votes and issues. A "faction" is often defined as a recognized political group with a defined political agenda and sometimes with formal membership requirements [1], [2].

Recent qualitative work on faction dynamics has demonstrated the importance of shared objectives and work in faction formation [1]. Consequently, actions such as voting together on bills that satisfy a political objective is a strong theoretical construct for detecting faction formation in a political setting. Herein, we will detail two different methods of faction detection from roll call data:

* This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1745016 and Office of Naval Research Multidisciplinary University Research Initiative (MURI) under award number N00014-17-1-2675. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or Office of Naval Research. Additionally, Thomas Magelinski was supported by an ARCS foundation scholarship.

MacRae's method and Gower's Method for using roll call data to find factions within the Ukrainian Parliament.

The Ukrainian Parliament, along with the Ukrainian Nation as a whole, has seen incredible political tumult over the recent years. Some of the recent crises include the *Euromaidan* crisis in 2013-2014 and the Russian invasion of Crimea in 2014. Possibly as a result of these crises, the Parliament has seen several significant episodes of factions forming and disintegrating. Knowing this, we can expect to find complex faction structure when analyzing Ukraine's parliamentary data. One interesting difference between the Ukrainian Parliament and other legislative bodies, like the U.S. Congress, is that a parliamentarian in the Ukrainian Parliament has six options for any given vote. A parliamentarian could not only vote "for" and "against" the bill, but also "absence," "abstaining," "do not vote," or "no vote." Not only this, but bills are voted on multiple times before passing. Successful bills must receive at least 3 majority votes, which can take as many as 9 attempts. Ukraine's history makes it a great contemporary example of faction formation, while its complex voting scheme provides new challenges for roll call analysis.

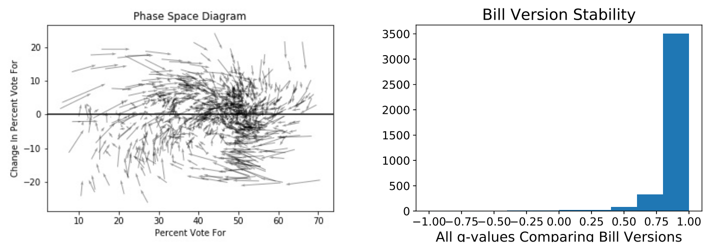
Analysis of roll call data is an important method for determining political factions. Prior research has featured spatial models, statistical models, and some bi-clustering [4]. Spatial models typically score legislators based on how they have voted either through a Guttman-type scale, a proximity-type scale, or euclidean distance in a shared voting space [5], [6], [7], [10]. In essence, legislators have two outcomes and a utility function, resulting in a probabilistic vote outcome to the roll call vote. This probabilistic outcome can then be used to determine the median ideal points for legislators in a space.

An older, less popular, approach to analyzing these votes is factor analysis. Factor analyses of roll call votes typically involve computing an association metric between each legislator and every other legislator using agreement or disagreement on bills [5], [6], [8], [9]. Association metrics typically used are Pearson's correlation coefficient, Yule's q, and ϕ/ϕ_{max} [8], [9].

2 IMPLEMENTATION: Adjustments, Assumptions and Decisions

2.1 MacRae Method of Faction Detection

Only Using Bill's Final Votes We explore only convocation 8, as it is the most stable in the multiple votes across any given bill. The impact of re-voting on bills was analyzed for stability. Each bill in convocation 8 containing two or more votes was plotted in the phase-space of the percentage of votes "for" the bill, shown in Figure 1. It can be seen that $\%change = 0$ is an attractor, indicating that bills become more stable each time they are brought to the floor. Given our focus on stability, we use only the last vote on each bill in our analysis. To assess the robustness of this choice, Yule's q-value from MacRae's method was used to test the information lost in removing earlier versions. The q-value



(a) (b)

Fig. 1. Bill phase diagram shown in (a), histogram of q-values comparing consecutive vote iterations in (b).

was calculated for each bill version and the previous version. As a measure of similarity, the higher the q-score, the less information lost by dropping previous bills. A histogram of the bill version comparison scores, shown in Figure 1, indicates that the overwhelming majority of bill versions show high similarity to their predecessor.

Further Data and Model Adjustments MacRae's initial implementation was for the U.S. Congress, which has a 2-option voting scheme: "for" and "against." Things like absences were counted as "error" votes. In the Verkhovna Rada there are six voting options. The additional 4 voting options are variations of "no vote." Unlike the U.S. system, these 4 options cannot be counted as error votes, since they are used 51.7% of the time, whereas votes "against" are only used 1.1% of the time. This analysis combined with expert opinion has lead to the conclusion that while votes "against" are a strong signal, the other 4 categories are used when one does not want to vote "for" a bill. As such, the voting options were categorized as "for" and "non-for" so that MacRae's method may still be used. With this modification, there are no longer "error" votes, and so no need for the error-correction rule.

Modern Adjustments After the creation of the Q-Matrix, MacRae proposed binarizing the fully connected, weighted network with a threshold of 0.7, to show connections between bills with high similarity. Visual analysis of data was preformed to find clusters. We, in contrast, use modern grouping algorithms, Louvain and Spectral clustering, to efficiently and objectively locate groups of bills.

While U.S. representatives are largely defined by party affiliation, Rada members show less party loyalty. Thus, MacRae's issue by party analysis is unlikely to be informative, so we calculated scale-scores per issue for each politician. The distance in scale-score between each pair of politicians was averaged across issues, resulting in a politician-politician distance network. The weights in this network ranged from 0 (identical actions) to 9 (opposite actions). The distance metric was converted into a closeness metric: closeness = $\exp(-\gamma * \text{distance})$. Closeness has weights between 0 and 1, with a higher value indicating a stronger tie. The γ value has been tested at values of 1, 1/2, and 1/4. Like bill grouping, the closeness matrix was thresholded (this time at the 50^{th} percentile) and the result was grouped into factions using either Louvain or spectral clustering.

To recap, there are 2 options for grouping bills into issues, 3 options in creating the closeness matrix, and 2 options for grouping the parliamentarians into factions, resulting in 12 potential MacRae faction detection algorithms.

2.2 Gower Method of Analyzing Groups

For Gower's Method, we first define the network of parliamentarians and voting using a bipartite network , $G(V, V', E)$. Since E characterizes the relationship of a given parliamentarian, V to a given bill V', it will take on values of "for," "against," "absence," "abstaining," "do not vote," or "no vote." These values are categorical and symbolic in nature. Thus, the weightings for each element of E will be *symbolic*. We now use Gower's Coefficient to compute pair-wise similarities between each pair of parliamentarians. The equation for Gower's Coefficient is given by: $S_{ij} = \frac{\sum_{k=1}^{N} w_{ijk}\delta(x_{ik}, x_{jk})}{\sum_{k=1}^{N} w_{ijk}}$. Where, in this case, S_{ij} is the similarity between parliamentarian i and parliamentarian j, k is the particular vote (there are N votes total), x is the response of a parliamentarian to a given vote (i.e. entries in the aforementioned incidence matrix), and $\delta(x_{ik}, x_{jk})$ is an indicator function between parliamentarian i and parliamentarian j, that outputs 1 if parliamentarian i and j had the same response to vote k, and 0 otherwise, and is akin to a Jaccard Index.

One of the important considerations with Gower's Coefficient is the weighting scheme w_{ijk} [3]. In the absence of subject matter knowledge about what factors should be weighted heavier, one scheme is to weight each sample by the entropy in the feature space [3]. So, more contentious votes of bills that split members of the parliament will have a higher entropy and should be better for ascertaining factions. The final pair-wise comparison equation for two Ukrainian Parliamentarians becomes: $S_{ij} := \frac{\sum_{k=1}^{N} [(-\sum_{m=1}^{\text{unique}(E_k)} p_m \ \log \ p_m)\mathbb{I}(x_{ik}, x_{jk})]}{\sum_{k=1}^{N}(-\sum_{m=1}^{\text{unique}(E_k)} p_m \ \log \ p_m)}$, where the entries are how similar a parliamentarian is to another parliamentarian. Note, that this affinity matrix will have values $\in (0, 1)$ and be symmetric with all ones on the main diagonal.

3 RESULTS

The methods are compared with the co-faction network, which is constructed by linking 2 MP's if they appear in the same faction under that method. We first look at the number of shared links in the co-faction network.

Figure 2 compares all the methods from a shared link perspective. Louvain grouping on this network reveals four intuitive groups of methods: Gower,

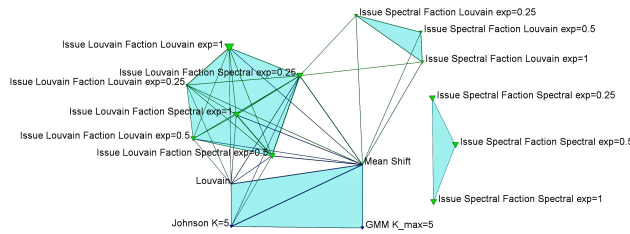
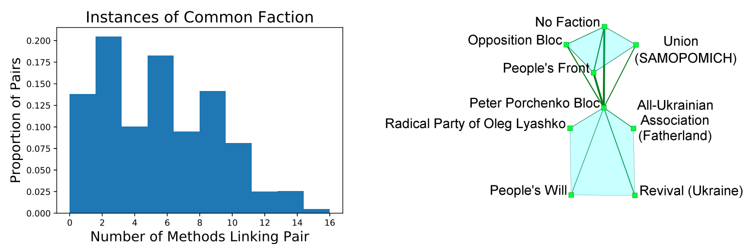


Fig. 2. The method-method network is shown, linked by the number of shared links in each method's co-faction network. Methods are colored by class and are sized by the number of factions they formed. The shaded polygons show the Louvain grouping of similar methods. Link weights less than the mean are not shown.



(a) (b)

Fig. 3. Histogram of number of methods linking 2 MP's in (a), party-party network using links with > 10 methods, with only links weighted greater than the mean in (b).

MacRae starting with Louvain, and the two types of MacRae starting with spectral clustering.

The weighted co-faction network can be created by summing the co-faction networks from each network. Analysis of this network shows that 13% of the links appear in 10 or more of the 16 possible methods. Using these links is one way to create a "high confidence" network of parliamentarians that takes advantage of information from each method. To study the relationship between official parties, the high-confidence parliamentarian network was used. Each link was converted to a party-party link, to create a weighted party-party network. That network is visualized in Figure 3, which also shows the degree distribution of the high-confidence network.

4 DISCUSSION

For MacRae's method, Louvain has the advantage of automatically determining the number of groups. Since the number of true factions is unknown, the user input needed for spectral clustering is a drawback. Perhaps the biggest assumption under MacRae's method comes from the use of the voting options. Information is lost from our assumption that all non-for votes are essentially "against." While this assumption better fits the original use of MacRae's method, Gower's method has clear benefits in this regard. Another underlying assumption in the current use of MacRae's method is that factions are formed based on all issues simultaneously. A more nuanced method may define factions for individual issues.

Finally, the internal structures of the MacRae Louvain-Spectral-0.25 faction network were analyzed to find the politicians with highest centrality. This method was chosen since it had the highest eigenvector centrality in the method-method network. The most important MP's are shown in Table 1. All of the important MP's come from faction 1 or 2. Probably because faction 1 is the dominant faction, with 359 members. Interestingly, many of the members have no formal party affiliation. Furthermore, two of Faction 1's key players are from opposing parties, the Peter Porchenko Bloc (the presidential and controlling party) and the Opposition Bloc.

Table 1. Most central politicians from MacRae-Louvain-Spectral-0.25.

MP	**Party**	**Faction**
Illenko Andriy Yuriyovych	No Faction	2
Golovko Mikhail Yosifovich	No Faction	1
Marchenko Alexandr Aleksandrovich	No Faction	1
Nikitas Maksim Viktorovich	No Faction	1
Bohdan Ruslan Dmitrievich	All-Ukrainian Association	1
Nasirov Roman Mikhailovich	Peter Porchenko Bloc	1
Voropayev Yuri Nikolaevich	Opposition Bloc	1

5 CONCLUSION

In this study we adapted two different methods for roll call voting data to the Ukrainian Parliament, where the vote responses and voting style are more complex than the U.S. Congress. Multiple methods were used simultaneously to link parliamentarians with higher confidence than individual methods. The result is a party-party network that groups the Oposition Bloc, People's front and those not affiliated with a party together, opposed to the other parties. The most central method was used to list the most central parliamentarians in the network, which listed members from opposing parties as important members of the same faction.

The methods presented both demonstrate that they can be used for faction detection. Gower's methods have the benefit of their ability to be used without collapsing non-binary voting responses. However, further work is needed to identify optimal weightings for Gower's and provide a more thorough comparison of roll call voting analysis methods.

References

1. Friedkin, N.E.: A structural theory of social influence, vol. 13. Cambridge University Press (2006)
2. Friedkin, N.E., Johnsen, E.C.: Social influence networks and opinion change. Advances in Group Processes **16**(1), 1–29 (1999)
3. Gower, J.C.: A general coefficient of similarity and some of its properties. Biometrics **27**(4), 857–871 (1971)
4. Hartigan, J.A.: Bloc voting in the united states senate. Journal of Classification **17**, 29–49 (2000)
5. Jr., D.M.: Dimensions of Congressional Voting: A Statistical Study of the House of Representatives in the Eighty-First Congress. University of California Press (1958)
6. Morrison, R.J.: A statistical model for legislative roll call analysis. Journal of Mathematical Sociology **2**, 235–247 (1972)
7. Poole, K.T., Rosenthal, H.: A spatial model for legislative roll call analysis. American Journal of Political Science **29**(2), 357–384 (1985)
8. Shubert, G.: The Judicial Mind: Attitudes and Ideologies of the Supreme Court. Northwestern University Press (1965)
9. Weisberg, H.F.: Dimensional Analysis of Legislative Roll Calls. Ph.D. thesis, University of Michigan (1968)
10. Weisberg, H.F.: Scaling models for legislative roll-call analysis. The American Political Science Review **66**(4), 1306–1315 (1972)