Multiple Partitioning of Multiplex Signed Networks

Nejat Arinik, Rosa Figueiredo, Vincent Labatut

Multiple Partitioning of Multiplex Signed Networks

Nejat Arınık[1], Rosa Figueiredo[1], Vincent Labatut[1]

Laboratoire Informatique d'Avignon, France
{firstname.lastname}@univ-avignon.fr

Mots-clés : *Signed Graph, European Parliament, Graph Partitioning, Correlation Clustering, Structural Balance.*

1 Introduction

In a *signed graph*, each link is associated to a sign, which can be either positive (+) or negative (−). This type of graph can be used to model any system containing two types of antithetical relationships (like/dislike, for/against, similar/different...). A signed graph is considered *structurally balanced* if it can be partitioned into two [6] or more [7] clusters, such that positive links are located inside the clusters, and negatives ones are in-between them.

However, it is very rare for a real-world network to have a perfectly balanced structure : the question is how to quantify its imbalance. Various measures have been defined for this purpose, the simplest consisting in counting the numbers of misplaced links, i.e. negative ones located inside the groups, and positive ones located between them [6]. Such measures are expressed relatively to a graph partition, so processing the graph balance amounts to identifying the partition corresponding to the lowest imbalance measure. In other words, calculating the graph balance can be formulated as an graph optimization problem.

Our goal is to use this paradigm to study the roll-call voting activity of the Members of the European Parliament (MEPs) by bringing together three different disciplines : Operations Research, Social Science and Data Science. In this work, we want not only to detect groups of MEPs which would be cohesive in terms of votes, but also to identify the different characteristic voting patterns of the European Parliament (EP), i.e. the characteristic ways in which the MEP set is partitioned by these votes. In literature the standard approach to study this type of system is to extract a vote similarity network, in which nodes represent MEPs and weighted (possibly signed) links represent the similarity between two MEPs, averaged over the series of roll-calls (e.g. [14, 13]). However, this averaging leads to some information loss due to the temporal integration performed on the raw data [3].

In the current work, we propose to adopt an approach based on a multiplex signed vote similarity networks, in which each layer models a single roll-call as a signed unweighted graph. All approaches proposed in the literature are based on the assumption that one is looking for a *single* partition in the end. However, this single-partition assumption is not compatible with all our objectives. Indeed, we look for the *different* characteristic polarizations of the EP, so we want our method to be able to identify *several* partitions. To this aim, we propose a new partitioning process for multiplex signed graphs. We apply our method to a dataset representing the voting activity during the 7$^{\text{th}}$ term of the EP. As desired, it allows identifying groups of cohesive voters, but also their different characteristic voting configurations, as well as the context (i.e. legislative propositions voted through roll-calls) to which they apply. In particular, we focus on certain case studies previously analyzed and discussed by Arinik *et al.* [3] using a standard approach. Not only does our method confirm certain of their assumptions, such as the fact that Socialists and Liberals alternatively align with the left- and right-wing blocks, but it also identifies the contexts that happen. In addition, it uncovers previously overlooked properties, for instance the emergence of a strong antagonism for a proportionally small number of roll-calls, that was invisible to the traditional approach. We invite reader to see the extended work of this summarized paper for further details [4].

The rest of the article is organized as follows. First, in Section 2, we give the formal definition of the Correlation clustering (CC) problem. We turn to the methods in Section 3, and describe the approach we propose for the analysis of multiplex signed networks. In Section 4, we present our results on a few specific cases selected from the dataset, and discuss them. Finally, in Section 5 we summarize our findings, comment the limitations of our work and describe how they can be overcome, and how our work can be extended.

2 Correlation clustering problem

In this section, we give the mathematical formulation of the CC problem. Let us introduce our notations before defining the problem itself. Let $G = (V, E)$ be an *undirected graph*, where V and E are the sets of vertices and edges, respectively. We note $n = |V|$ and $m = |E|$ the numbers of vertices and edges, respectively. Consider a function $s : E \to \{+, -\}$ that assigns a *sign* to each edge in E. An undirected graph G together with a function s is called a *signed graph*, denoted by $G = (V, E, s)$. An edge $e \in E$ is called negative if $s(e) = -$ and positive if $s(e) = +$. We note E^- and E^+ the sets of negative and positive edges, respectively.

Let $P = \{F_1, ..., F_\ell\}$ $(1 \leq \ell \leq n)$ be an ℓ-partition of V, i.e. a division of V into ℓ non-overlapping and non-empty subsets F_i $(1 \leq i \leq \ell)$ that we call *factions* in the context of this work. For $\sigma \in \{+, -\}$, the set of positive or negative edges (depending on σ) connecting two factions $F_i, F_j \in P$ $(1 \leq i, j \leq \ell)$ is $E^\sigma[F_i : F_j] = \{\{u, v\} \in E^\sigma \mid u \in F_i, v \in F_j \text{ or } v \in F_i, u \in F_j\}$, and its cardinality, i.e. its total number of links, is expressed as $\Omega^\sigma(F_i, F_j)$.

The *Imbalance* $I(P)$ of a partition P is defined as the total number of positive edges located between factions, and negative edges located inside them, i.e.

$$I(P) = \sum_{1 \leq i \leq \ell} \Omega^-(F_i, F_i) + \sum_{1 \leq i \neq j \leq \ell} \Omega^+(F_i, F_j). \tag{1}$$

The CC problem is formally described as follows.

Problem 1 (CC problem). *For a signed graph $G = (V, E, s)$, the Correlation Clustering problem consists in finding a partition P of V such that the imbalance $I(P)$ is minimized.*

This NP-hard minimization problem appears under this name for the first time in Bansal's paper [5], although it was addressed before in the literature, e.g. [8].

3 Methods

In this section, we describe the method that we propose to analyze our multiplex signed networks. We will handle various types of partitions, so we need to clarify our terminology first. We call *voting behavior pattern* of the EP (or *pattern*, for short) a partition of the set of all MEPs obtained for a given roll-call. The subsets of MEPs constituting this partition are called *factions*. A *pattern* represents the way the parliament is split at the occasion of a roll-call concerning a specific subject. We reserve the term *clustering* to refer to a partition of the set of all patterns (i.e. the patterns corresponding to all considered roll-calls). Since each pattern describes the EP behavior for a given roll-call, a clustering can also be interpreted as a partition of the set of all roll-calls. The subsets of roll-calls constituting a clustering are simply called *clusters*. For instance, if roll-calls #1, 2 and 3 form a cluster, this means the pattern was similar for all three corresponding roll-calls.

The goal of our method is to identify the main types of patterns occurring at the EP, and to characterize them in terms of political groups, individual MEPs and topics of the concerned legislative documents. To this aim, we propose a three-stepped method, summarized in Figure 1, and detailed in the rest of this section. The first step is to separately partition the p roll-calls,

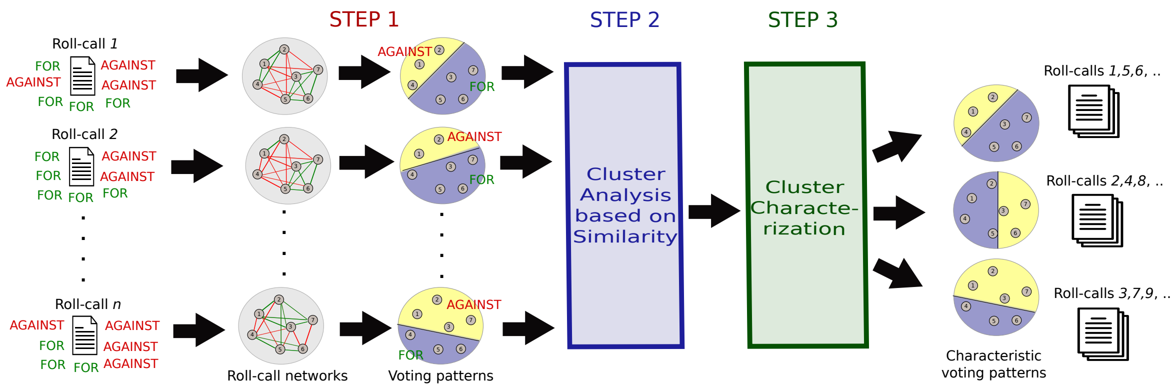


FIG. 1 – General workflow of the proposed analysis method.

in order to get as many patterns (Section 3.1). The second step consists in applying a clustering method onto those patterns (Section 3.2). This leads to a set of k clusters, each one gathering similar patterns. The third step is to process what we call the *characteristic pattern* of each cluster, which is supposed to consensually represent all the patterns belonging to the cluster (Section 3.3). This results in a set of characteristic patterns, each one associated with a cluster of roll-calls, which can be used as the basis of the interpretation work.

3.1 Processing the Patterns

Detecting the pattern associated with the i^{th} roll-call amounts to partitioning the i^{th} layer of our multiplex signed network. This layer is an unweighted signed graph G_i whose nodes are connected depending on how the MEPs they represent voted during this roll-call. MEPs who voted similarly are connected together by positive links, and are connected by negative links to MEPs that voted differently from them. MEPs who did not vote at all are isolated (nodes without any neighbor). We consider that absent MEPs should not affect the pattern, since they did not express any opinion regarding the matter at hand during this roll-call, so we simply ignore them at this stage. For those who abstained, we keep them in the graph, since abstaining is a way of expressing some middle ground position.

Next, we need to identify the factions of similarly voting MEPs remaining in the graph, which can be done by solving the *Correlation Clustering* problem (CC). In its original version [5], and consistently with the definition of structural balance given earlier in the introduction, it consists in finding a partition of the set of vertices which maximizes both the number of positive links located *inside* the subgroups, and that of negative links located *between* them. In order to identify this partition, we coded Ex-CC (Exact Correlation Clustering), which is a mixed integer programming (MIP) based method able to solve the CC problem exactly [1]. It is well known that exact approaches of any clustering problem do not scale much due to its NP-hard nature. Moreover, the process can be very time-consuming even for medium-sized networks (e.g. 100 nodes). One way to deal with this issue is to strengthen the underlying MIP model through the *cutting plane* approach with the following valid inequalities whose efficiency is empirically proved by Ales *et al.* [2] for another graph clustering problem : The 2-partition and the 2-chorded cycle inequalities.

3.2 Performing the Clustering

At this stage, we have identified the pattern associated with each roll-call. We now want to gather similar patterns together. For this purpose, we use a classic cluster analysis approach.

We first compute a similarity value based on the *Purity* [12] for each pair of patterns. We then build a dissimilarity matrix summarizing these comparisons. Next, we apply the k-medoids clustering method to the previously obtained similarity matrix [10]. It is similar to the well-known k-means algorithm in the sense that it tries to partition the dataset in k clusters, while minimizing the distance between the members of each cluster and some center of the cluster. The difference is that in k-means, this center is an average value, whereas in k-medoids it is one of the actual data points from the dataset. It is generally used in place of k-means when one cannot perform the required average operation, which is true in our case (we cannot straightforwardly process an average pattern).

3.3 Computing the Characteristic Patterns

We now have k clusters, each one containing a certain number of patterns. The patterns constituting a cluster may differ slightly, but overall they are supposed to be very similar. The next step is to compute a characteristic pattern representing the whole cluster, such that these small differences are smoothed.

For this purpose, we use a similarity network-based approach, inspired by the work of Lancichinetti & Fortunato [11]. Based on a collection of n partitions of the same set, we derive a consensual partition by first extracting a weighted signed similarity network, and then applying Ex-CC to this network, in order to identify a partition corresponding to the characteristic pattern of the cluster. The network is built as follows : each node represents a MEP, and the weight of the link connecting two MEPs is the difference between the proportion of patterns putting both MEPs in the same faction, and the proportion of patterns putting them in different ones.

Like any pattern, a characteristic pattern can take one of three forms : a single faction in case of unanimity, 2 antagonistic factions if some MEPs agreed on the concerned roll-calls whereas others opposed them, and 3 in case of an additional faction of abstentionists.

4 Results

As mentioned in the introduction, our goal is to compare our proposed method with the standard approach traditionally used when studying vote networks. For this purpose, we analyze the same data as Arinik *et al.* [3], and more particularly the votes of French and Italian MEPs related to agricultural questions during the 2012-13 legislative year. We expect our method to be able to answer some of the questions left open by their analysis. We take their discussion as a reference when commenting our own results, highlighting both differences and similarities between these approaches. Our source code is publicly available [1].

4.1 Conventions and General Remarks

To denote the clusters, we use a notation indicating first the concerned country (Fr for France), then the value of k used when detecting the considered clustering through k-medoids, and finally the number of the cluster in this clustering. For instance : Fr-$k3$-$clu2$ is the name of the second cluster of the clustering obtained with $k = 3$ for the French MEPs.

We represent all networks and voting patterns using a circular layout (Figures 2) generated through Circos [2]. We describe them generically here, for matters of convenience. They shall be read from the center to the periphery. The negative and positive links are drawn at the center, in red and green, respectively. Next, the inner colored ring represents the nodes (MEPs), and these colors correspond to the factions constituting the detected pattern (i.e. partition of the MEPs). If a MEP was often absent, he is ignored and appears in white. The names of the MEPs are not included due to lack of space. Finally, the outer ring shows the European political groups to which the MEPs belong. They are ordered according to the political spectrum, from left to right : GUE-NGL (red), G-EFA (green), S&D (pink), ALDE (orange), EPP (light blue), ECR (dark blue), EFD (purple) and NI (brown).

1. https://github.com/CompNet/MultiNetVotes
2. http://circos.ca

4.2 Baseline

Arinik *et al.* [3] have applied a classic method, and extracted a vote similarity network by integrating their data over the whole considered legislative year (2012-13). Moreover, they have filtered the weakest links to sparsify the network. Figure 2a and Figure 2d show the best partition that Arinik *et al.* obtained by solving the CC problem, and a variant called RCC problem, respectively. They state that the network is highly polarized, as it contains many negative links and results in good partitions in terms of structural balance. They identify two antagonistic factions respectively led by the environmentalists (G-EFA) and the right conservatives (EPP), joined by some other groups or individual MEPs. The position of S&D and ALDE is interesting, because they belong to the right-wing faction according to CC, whereas they hold an intermediate position with RCC. Arinik *et al.* were not able to give a solid explanation for this. But they assumed that these groups were sometimes voting like the left-wing faction, and sometimes like the right-wing one.

4.3 Characteristic Patterns

As mentioned in Section 3.3, k-medoids requires us to specify the number of clusters. In theory, there exist several internal criteria to do it automatically. However, in practice, one possibly has to consider other factors to make a choice. Therefore, we identify $k = 5$ as our best trade-off (which is also consistent with the internal criteria), and discuss each cluster obtained for $k = 5$ and its corresponding pattern in the following.

4.3.1 Unanimity

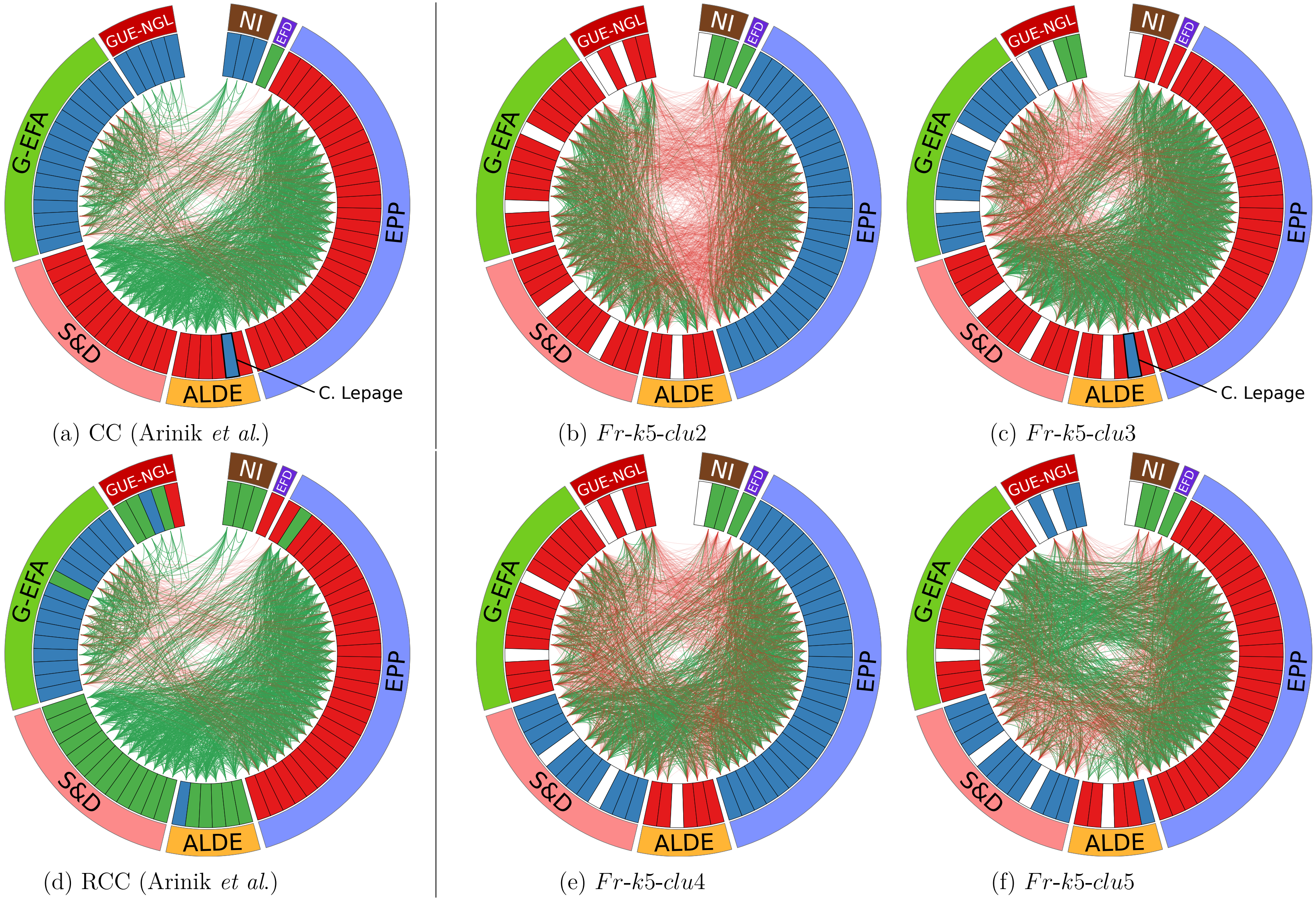
Cluster Fr-$k5$-$clu1$ corresponds to a unanimity situation, so for space considerations is not included in Figure 2. Indeed, it contains a single faction, and only one negative link, between P. Le Hyaric (GUE-NGL) and M. Le Pen (NI). The emergence of such a high level of agreement was completely hidden when considering only a temporal integrated network (i.e. averaged over the series of roll-calls), and therefore could not be detected by Arinik *et al.* It is the largest cluster with 100/232 roll-calls (43%), so we can assume it represents the regular voting behavior in the considered context. All the other clusters correspond to patterns containing varying antagonistic factions. This is consistent with the fact that our clusters are supposed to correspond, by construction, to distinct voting patterns.

4.3.2 Conservatives vs. All

The characteristic pattern associated with Fr-$k5$-$clu2$, shown in Figure 2b, finds the right-wing conservative group (EPP) opposing the rest of the MEPs, while both Euroskeptic groups (EFD and NI) abstain. This cluster contains 34/232 (15%) roll-calls. An examination of the content of the corresponding legislative documents voted through those roll-calls reveals that this voting behavior corresponds to EPP trying to block radical changes related to the CAP. These changes, as well as their blocking by the right-wing conservatives, are confirmed in a positioning paper published by S&D about the 2013 CAP reform [9].

4.3.3 Environmentalists vs. All

Cluster Fr-$k5$-$clu3$, shown in Figure 2c, contains 74/232 roll-calls (32%). Its characteristic pattern finds the environmentalist group (G-EFA) opposing a large faction constituted of the rest. The far-left group (GUE-NGL) is apart, as one MEP agrees with the environmentalists whereas the rest of his group abstains. This is very similar to the pattern obtained by Arinik *et al.* when solving CC on their integrated network, except for the NI group and a few MEPs. In particular, Corinne Lepage, which was described by Arinik *et al.* as an environmentalist member of ALDE, is placed in the G-EFA faction by our method. The relevance of this faction is confirmed by her activity at the EP, where she is very active on issues such as *food safety*.



(a) CC (Arinik *et al.*) (b) Fr-$k5$-$clu2$ (c) Fr-$k5$-$clu3$

(d) RCC (Arinik *et al.*) (e) Fr-$k5$-$clu4$ (f) Fr-$k5$-$clu5$

FIG. 2 – Voting behavior patterns of the French MEPs on AGRI questions in 2012-13. Red and green lines at the center represent negative and positive links, respectively. Around the links, each MEP is represented by a colored tile, whose color corresponds to the MEP's faction in the displayed pattern. The green factions in plots (b), (c), (e) and (f) correspond to abstentionists. The outer ring represents the political groups at the EP. The left plots show the patterns obtained by Arinik *et al.* [3] on the temporal integrated network when solving (a) CC and (d) RCC. The right plots show the second to fifth clusters obtained with our proposed method for $k = 5$: (b) Fr-$k5$-$clu2$ (%15 of roll-calls), (c) Fr-$k5$-$clu3$ (%32 of roll-calls), (e) Fr-$k5$-$clu4$ (%8 of roll-calls), and (f) Fr-$k5$-$clu2$ (%3 of roll-calls). The first cluster, Fr-$k5$-$clu1$, corresponding to a unanimity situation, is not included to lack of space.

4.3.4 S&D/EPP vs. the Rest

Cluster Fr-$k5$-$clu4$, shown in Figure 2e, represents 18/232 (8%) roll-calls. Its characteristic pattern contains a faction formed by the far-left, environmentalist and liberal groups (GUE-NGL, G-EFA, ALDE), vs. another faction containing the socialists and conservatives (S&D, EPP), while both Euroskeptical groups form an abstentionist faction. This constitutes a new type of pattern, different from all the others met until now, including in the baseline. In particular, it is is worth noticing that S&D and ALDE do not belong to the same faction. Thus, if these groups alternatively side with left- and right-wing groups, as already assumed by Arinik *et al.*, our method shows that they do not always do so simultaneously.

4.3.5 Unholy Alliance

For Fr-$k5$-$clu5$, as illustrated in Figure 2f, the characteristic pattern finds a faction gathering environmentalists and right-wing liberals and conservatives (G-EFA, ALDE, EPP), opposing a faction composed of the far-left and socialist groups (GUE-NGL and S&D), while the Euroskeptics abstain once again. These factions are surprising from a political standpoint, as they exhibit a somewhat unholy alliance between environmentalists and conservatives, whose views generally clash for AGRI matters. But the pattern is also surprising when considering Arinik *et al.*'s results, as they do not detect this alliance at all. The cluster contains only 6 roll-calls (2%), which shows that this situation does not happen often. A careful examination of the texts voted at these roll-calls reveals that the interests of these two groups happen to match, very punctually, and for completely different reasons.

4.3.6 Comparison with the Baseline

Our results confirm in a more objective way the assumption of Arinik *et al.*, based on the RCC pattern from Figure 2, and according to which S&D and ALDE sometimes vote like the left-wing groups (as in Fr-$k5$-$clu2$) and sometimes like the right-wing ones (Fr-$k5$-$clu3$). Our method additionally identifies the context (i.e. roll-calls) for which the EP adopts these two patterns. But our method also shows that these two groups vote differently on a number of occasions (Fr-$k5$-$clu4$ and Fr-$k5$-$clu5$), a fact overlooked when using the traditional approach.

In addition, our results uncover the fact that the Euroskeptics systematically abstain on most roll-calls, and only vote for a specific subset corresponding to the *Green vs. All* pattern. This specific behavior put them apart from the rest of the groups, and maybe this is why they had been categorized by Arinik *et al.* as an intermediate group, like S&D and ALDE. However, our results show that this is an artifact of the Euroskeptics' abstentionist behavior, and that they hold a completely different position than S&D and ALDE.

Finally, our method allows identifying the *Unholy Alliance* pattern, which had completely been overlooked by Arinik *et al.* It corresponds to a very surprising coalition, politically speaking, which emerged when voting for a very specific set of roll-calls.

5 Conclusion

In this work, we introduce a method to the partition of multiplex signed networks. We show the interest of our approach by applying it to a subset of the 7^{th} term European Parliament dataset presented by Arinik *et al* for France and Italy. By comparison to existing approaches, our method has the following advantages. First, it undergoes much less of the information loss appearing when integrating the raw voting data to extract the voting similarity networks, since it treats separately each roll-call vote in the partitioning process. Second, in addition to antagonistic groups of voters, it allows identifying sets of legislative propositions causing the same polarization among these groups. This additional information can be leveraged by the end-user to better explain the observed outcomes. Third, unlike other methods, it does not require to filter out (quasi-)unanimous propositions, or to discard week links appearing in the model for interpretation or computational purposes. Fourth, it explicitly represents abstention in each roll-call vote layer, which allows detecting relevant groups of abstentionists.

Our method is generic and can be applied to any system with similar properties. For example, in the context of document/artwork classification, the opinion expressed by a selection of specialists can produce a set of polarizations, each one representing a variety of opinions regarding an item [5]. Our method could allow identifying clusters of items leading to similar expert polarization. It can also be used to group specialists sharing the same point of view on certain issues. In the near future, we will apply our method more systematically to the whole EP dataset. Also, we will perform a textual content analysis of the voted documents, in order to provide the information required to properly interpret the identified characteristic voting patterns.

Références

[1] Z. Ales and A. Knippel. An extended edge-representative formulation for the k-partitioning problem. *Electronic Notes in Discrete Mathematics*, 52 :333–342, 2016.

[2] Z. Ales, A. Knippel, and A. Pauchet. Polyhedral combinatorics of the k-partitioning problem with representative variables. *Discrete Applied Mathematics*, 211 :1–14, 2016.

[3] N. Arinik, R. Figueiredo, and V. Labatut. Signed graph analysis for the interpretation of voting behavior. In *International Conference on Knowledge Technologies and Data-driven Business - International Workshop on Social Network Analysis and Digital Humanities*, Graz, AT, 2017.

[4] N. Arinik, R. Figueiredo, and V. Labatut. Multiple partit. of multiplex signed networks. *Soc Netw*, 2019.

[5] N. Bansal, A. Blum, and S. Chawla. Correlation clustering. In *43rd Annual IEEE Symposium on Foundations of Computer Science*, pages 238–247, 2002.

[6] D. Cartwright and F. Harary. Structural balance : A generalization of heider's theory. *Psychological Review*, 63 :277–293, 1956.

[7] J. A. Davis. Clustering and structural balance in graphs. *Human Relations*, 20(2) :181–187, 1967.

[8] P. Doreian and A. Mrvar. A partitioning approach to structural balance. *Social Networks*, 18(2) :149–168, 1996.

[9] Group of Progressive Alliance of Socialists & Democrats (S&D). CAP Reform - a good deal despite resistance from some Governments. Technical report, December 2013.

[10] L. Kaufman and P. J. Rousseeuw. Partitioning around medoids. In *Finding Groups in Data : An Introduction to Cluster Analysis*, page 10.1002/9780470316801.ch2. John Wiley & Sons, Hoboken, NJ, USA, 2009.

[11] A. Lancichinetti and S. Fortunato. Consensus clustering in complex networks. *Scientific Reports*, 2 :336, 2012.

[12] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA, 2008.

[13] V. A. Traag, G. Krings, and P. van Dooren. Significant scales in community structure. *Scientific Reports*, 3(1), 2013.

[14] A. S. Waugh, L. Pei, J. H. Fowler, P. J. Mucha, and M. A. Porter. Party polarization in congress : A network science approach. *arXiv*, physics.soc-ph :0907.3509, 2009.