

Analyzing Voting Behavior in Italian Parliament: Group Cohesion and Evolution

Alessia Amelio

National Research Council of Italy (CNR)
ICAR & DEIS, Università della Calabria
Via Pietro Bucci, 41C
87036 Rende (CS), Italy
Email: {aamelio}@deis.unical.it

Clara Pizzuti

National Research Council of Italy (CNR)
ICAR
Via Pietro Bucci, 41C
87036 Rende (CS), Italy
Email: {pizzuti}@icar.cnr.it

Abstract—The roll calls of the Italian Parliament in the current legislature is studied by employing multidimensional scaling, hierarchical clustering, and network analysis. In order to detect changes in voting behavior, the roll calls have been divided in seven periods of six months each. All the methods employed pointed out an increasing fragmentation of the political parties endorsing the previous government that culminated in its downfall. By using the concept of modularity at different resolution levels, we identify the community structure of Parliament and its evolution in each of the time periods considered. The analysis performed revealed as a valuable tool in detecting trends and drifts of Parliamentarians. It showed its effectiveness at identifying political parties and at providing insights on the temporal evolution of groups and their cohesiveness, without having at disposal any knowledge about political membership of Representatives.

Keywords-SVD; hierarchical clustering; network evolution;

I. INTRODUCTION

In the last years political parties in Italy have been affected by a steady fragmentation, with a high number of Parliamentarians leaving the group that allowed them to be elected to join another one, often changing party many times.

In this paper we investigate Italian Parliament by using different tools coming from Data Mining and Network Analysis fields with the aim of characterizing the modifications Parliament incurred, without any knowledge about the ideology or political membership of its Representatives, but relying only on the votes cast by each Parliamentarian. We consider the roll calls of the period of three years and an half from April 2008 until October 2011, after which there was the fall of the center-right coalition that won the elections. This period has been equally divided in seven semesters and the votes cast by each Parliamentarian have been stored. Note that in our analysis we do not consider the Italian Senate, but only the Parliament.

Voting records have been used in two different ways. In the first approach we directly use them to show party cohesion during the considered period, and apply a multidimensional scaling technique to reveal political affinity of Parliamentarians, independently of their true party membership. This kind of analysis is interesting because it is able to reproduce the effective political alliances, without

assuming parties as relevant clusters. In the second one, from voting records we compute similarity between each pairs of Representatives and try to detect structural organization and evolution of Parliament by applying hierarchical clustering and community detection based on the concept of modularity. All the approaches conduct to coherent results. However, by using the modularity concept, we identify communities of members that voted similarly, and investigate how the party cohesion evolves along the semesters. The analysis provides an explicit and clear view of the steady fragmentation of the coalition endorsing the center-right government, that caused the majority breakdown. Thus modularity allows a more deep analysis of the internal agreement of parties, and demonstrated a powerful means to give insights of changes in majority party.

The investigation of voting records with computational techniques is not new [1], [2], [3], [4], [5], [6], though this is the first study regarding an Italian institution.

The paper is organized as follows. In the next section we give a brief description of the Italian Parliament organization and the data set used for the analysis. In section III we describe the application of multidimensional scaling approach to the voting records. In section IV the similarity metric used is defined, and the groups obtained by applying hierarchical clustering and community detection are discussed. Section V argues about the results obtained for the last semester. Section VI, finally, concludes the paper and outlines future developments.

II. DATA DESCRIPTION

The current Italian Parliament has been elected in April 2008 and it is constituted by 630 representatives originally elected in 5 main political parties: People of Liberty (PDL), League of North (LN), Democratic Party (PD), Italy of Values (IDV), and Democratic Union of Center (UDC). The majority of center-right that governed Italy until November 2011 was composed by the first two parties. To better understand the analysis we performed, it is important to know that two main events characterized the political organization of Parliament: (1) in July 2010 a group of Representatives divided from PDL to form a new political party named

Future and Liberty (FL); (2) in December 2010 some Parliamentarians, mainly coming from the center-left coalition, separated from their party to constitute a new coalition, named People and Territory (PT), that endorsed the center-right government, allowing it to rule the country for other almost ten months. Furthermore, along all the three years and an half, several Representatives abandoned their party to move in a group called Mixed. The Italian Parliament maintains a database of the legislative activity by storing, for each bill voted, the list of votes cast by each Representative. From the web site <http://parlamento.openpolis.it> it is possible to download the voting record of each Parliamentarian, together with some personal information, such as territorial origin, and actual group membership. For every roll call, the Openpolis database stores the vote of each Parliamentarian in three ways: 'yes', 'no', and 'not voting'. This last kind of vote can be due to either absence or abstention, but they are treated in the same manner.

Table I
NUMBER OF VOTED MEASURES FOR EACH SEMESTER.

I	II	III	IV	V	VI	VII
386	422	328	343	373	332	89

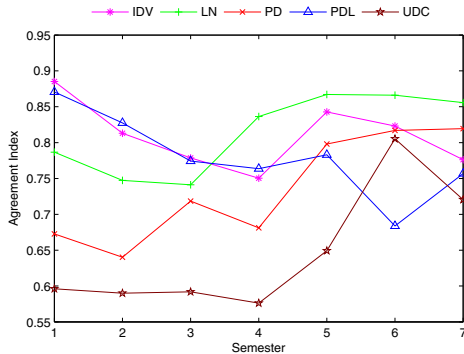


Figure 1. Agreement index of parties for all the semesters.

III. ANALYSIS OF VOTING PATTERNS

We collected the roll calls of the Italian Parliament in the period starting from April 2008 until October 2011, after which there was the fall of the center-right coalition that won the elections. This period of three years and an half has been equally divided in seven semesters and the votes cast by each Parliamentarian have been stored in matrices of size $n \times m$, where n is the number of Parliamentarians, and m is the number of bills voted in the reference period. Since some Parliamentarians, for several reasons, never voted, they have been eliminated. Thus the number n of Representatives reduced to 612. As regards m , it assumes a different value, depending on the semester. The

number of bills voted is reported in Table I. Seven voting matrices have been built in the following way: an element A_{ij} of a voting matrix A is +1 if the Representative i voted *yes* on measure j , -1 if he voted *no*, and 0 if he did not vote. The voting matrices are exploited to study the voting behavior of the Italian Parliament in two different ways. In the first approach we use them to compute party cohesion and to characterize the political affinity of Parliamentarians, independently of their true party membership. In the second one, we compute similarity for each pairs of Representatives and try to detect structural organization and evolution by applying hierarchical clustering and community detection based on the concept of modularity.

A. Party Cohesion

Given the voting matrices, the first investigation that can be done is to compute the cohesion of each political party along the considered period and compare the results obtained. To this end, the *agreement index* [7] measures the level of cohesion within a party by exploiting the number of equal votes for each roll call. The agreement index for each roll call is defined as follows:

$$AI_i = \frac{\max\{y_i, n_i, a_i\} - \frac{y_i + n_i + a_i - \max\{y_i, n_i, a_i\}}{2}}{y_i + n_i + a_i}$$

where y_i is the number of members who voted *Yes* in the voting i , n_i is the number of members who voted *No*, and a_i is the number of members who did not vote. Group cohesion is then computed as the average of agreement indices for all the roll calls:

$$AI = \frac{\sum_i^m AI_i}{m}$$

The agreement index ranges from 0 (complete disagreement) to 1 (complete agreement). Figure 1 displays the trend of agreement index of the 5 main political parties during the seven semesters. It is clear from the figure that the opposition parties show an increasing cohesion, while PDL, that started with a value near to 0.9, has a constant downtrend until the sixth semester, with a slight increment in the last semester. The variation of internal cohesion well reflects the actual political situation along the considered periods.

B. Singular Value Decomposition

We now analyze the voting behavior of Italian Parliament by applying the well known multidimensional scaling technique known as *Singular Value Decomposition (SVD)* [8], whose advantages with respect to other techniques have been discussed in [9]. Let A be an $n \times m$ voting matrix where rows correspond to Representatives and columns to the votes cast to approve a law. The *Singular Value Decomposition* of A is any factorization of the form

$$A = U \times \Lambda \times V^T$$

where U is an $n \times n$ orthogonal matrix, V is an $m \times m$ orthogonal matrix and Λ is an $n \times m$ diagonal matrix with

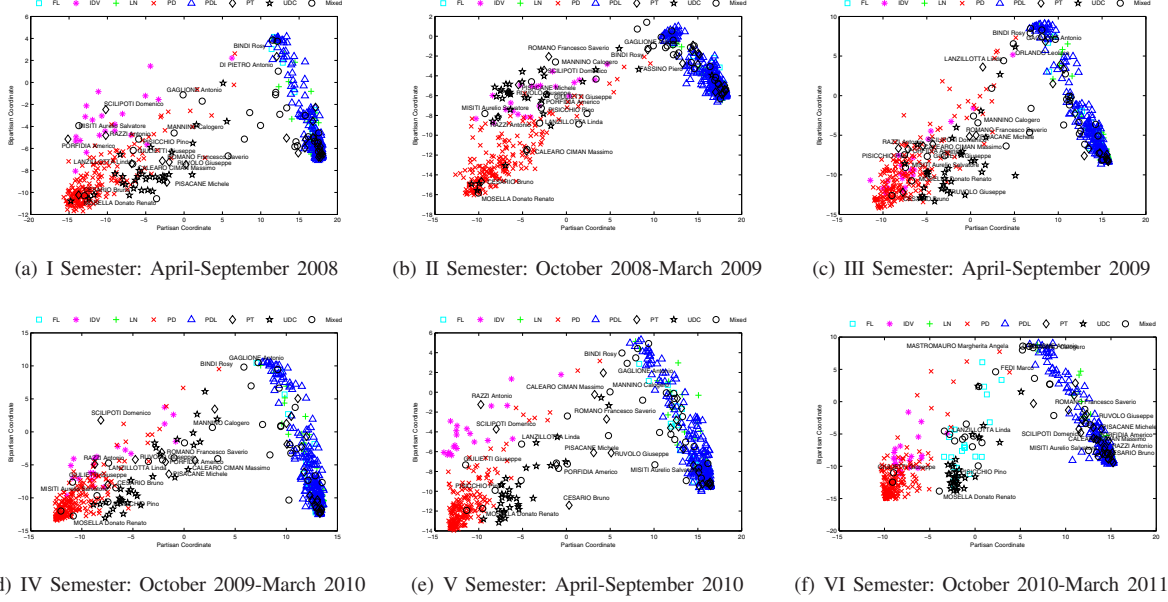


Figure 2. Singular value decomposition of the Italian Parliament voting behavior for each of the six semesters starting from April 2008 until March 2011.

$\lambda_{ij} = 0$ if $i \neq j$. The diagonal elements λ_i are called the *eigenvalues* of A . It has been shown that there exist matrices U and V such that the diagonal elements of Λ are the square roots of the nonzero eigenvalues of both AA^T and $A^T A$, and they can be sorted such that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ [8]. Geometrically this factorization defines a rotation of the axis of the vector space defined by A where V gives the directions, Λ the strengths of the dimensions, and $U \times \Lambda$ the position of the points along the new axis. Intuitively, the U matrix can be viewed as a similarity matrix among the rows of A , i.e. the Representatives, the V matrix as a similarity matrix among the columns of A , i.e. the votes cast for each law, the Λ matrix gives a measure of how much the data distribution is kept in the new space [10]. If the singular values λ_i present a fast decay, then $U \times \Lambda$ provides a good approximation of the original voting matrix A . In particular, by projecting on the first two coordinates, we obtain a compressed representation of the voting matrix that approximates it at the best. The visualization of the projected approximation matrix, allows to identify groups of Representatives that voted in a similar way on many bills. As observed in [3], the first coordinate correlates to party membership, thus it is called the *partisan* coordinate. The second coordinate correlates to how often a Representative voted with the majority, thus it is called the *bipartisan* coordinate.

Figure 2 shows the application of *SVD* on the voting records of the Italian Parliament for the first six semesters of the current legislature. Each point corresponds to the projection of votes cast by a single Parliamentarian onto the

leading two eigenvectors partisan and bipartisan. Each party has been assigned a different color and symbol. The main objective of this analysis was to study the changes in voting behavior of those Parliamentarians that moved from the opposition coalition to the majority one. Thus we selected some members of PT and Mixed groups, and visualized their names on all the figures. First of all we point out that the representation of the two coalitions center-right and center-left, and their evolution along the three years, summarized by the six figures, is very impressive.

Figure 2(a) clearly shows a compact center-right aggregation, a less cohesive, but clearly distinguishable, center-left alliance, and a strong connected PD sub-group (left bottom). It is worth to note that this sub-group maintains its connectedness for all the time periods, with a slight dispersion in the second semester. The same cohesiveness is shown by PDL and LN, as expected. Moreover FL, which was included in PDL until July 2010, demonstrated its political disagreement in the sixth semester by coming nearer to UDC, as effectively happened. As regards the chosen members of PT and Mixed groups, we can observe a steady movement from the center-left coalition to the center-right one since the fourth semester. This shift is much more evident in the 5th semester, when the voting behavior of these Representatives approached closer and closer to center-right majority. In fact, all the Parliamentarians located in the central part of Figure 2(e), appear at right in Figure 2(f), indistinguishable from the majority coalition. We also notice that there is a PD Parliamentarian positioned upper, near the right coalition, for five semesters. Because of the interpre-

tation of the bipartisan coordinate, her location means that she mostly voted with the majority. This dissimilarity from the own political party, perhaps can be explained by the fact that this Representative is vice-president of the chamber.

Analysis of voting behavior with Singular Value Decomposition is thus a powerful tool to characterize political ideology of Parliamentarians, and to trace the evolution of their position along consecutive time periods. SVD is able to find structural patterns and latent information in the voting records without any knowledge about the political orientation of Representatives.

IV. PARLIAMENTARIANS SIMILARITY

There can be different ways of defining similarity between two Parliamentarians from the voting matrix. For example, Jakulin and Buntine [1] used the mutual information concept. However, as observed by the authors, if two members always vote in the opposite way, they also are considered similar. We think that this kind of proximity measure misrepresents the Representative closeness, thus we employed a more suitable measure. Considering that when two Representatives cast a vote the values *yes* and *no* should be considered equally important in comparing their political affinity, we adopted the proximity measure known as *simple matching coefficient (SMC)* [11]. We ignored the cases when at least one of the two did not vote because, as already pointed out, this means either abstention or absence, and we cannot distinguish between them. Thus there can be four different outcomes: (1) *yy*, both voted *yes*, (2) *nn*, both voted *no*, (3) *yn*, the first Parliamentarian voted *yes* and the second one *no*, (4) *ny*, the first Parliamentarian voted *no* and the second one *yes*. Then the *SMC* of Parliamentarians p_1 and p_2 is defined as

$$SMC(p_1, p_2) = \frac{yy + nn}{yy + nn + yn + ny}$$

The simple matching coefficient thus computes the fraction of equal votes, both *yes* or *no*, with respect to the total votes they cast. The similarity metric defined allows us to measure the closeness of each pair of Parliamentarians on the base of their voting behavior. In such a way a symmetric similarity matrix M among all the Parliamentarians can be built, and their proximity with the members of the same or opposite parties studied. A summarized view of the affinity between each couple of Representatives can be done in different ways. In the following we first apply a hierarchical clustering algorithm, and then we give a graphical representation of the similarity matrix.

A. Hierarchical clustering

We apply the agglomerative hierarchical clustering method known as *single linkage clustering* [11]. The algorithm uses the smallest distance between two Parliamentarians and it generates a hierarchical cluster tree known as

dendrogram. The dendrogram shows the cluster-subcluster relationships and the hierarchical structure of the merged groups. Figure 3 represents very well the political alliances along all the semesters. The colors inside the dendrogram represent the clusters found by the algorithm. Attached to the leaves there are the names of the corresponding politicians, painted with the colors of the true associated party.

In Figure 3(a) we can observe as the two main political parties, PD in red and PDL in blue, correspond to the two main clusters of the dendrogram for all the semesters. The other parties (IDV in magenta, FL in cyan, LN in green, PT in orange, UDC in brown, and Mixed in violet) are clusters of smaller size, or they are merged inside the main clusters. For example, LN party is grouped together with PDL in all the semesters, reflecting the real political (center-right) alliance between PDL and LN. Another similar case is IDV: most of the members are grouped with the PD while some of them appear in different clusters for all the semesters.

Let us now consider the remaining parties. FL, as already described, was included into PDL until July 2010, when internal problems caused the movement of FL in the direction of center-left alliance. This phenomenon is captured from the clustering process. In fact FL is included into the majority for the semesters I-V (Figures 3(a-e)), while in the 6th semester all the members of FL are separated from PDL and grouped together with the opposite part (Figure 3(f)).

In order to analyze more clearly the trend of PT and Mixed parties, we looked not only at the dendrograms but also at the confusion matrices generated for all the semesters. They show what really happened along the semesters of the legislature: the gradual movement of PT and of some members of the Mixed group in the direction of the center-right alliance.

Furthermore, it is interesting to observe that UDC is recognized from the clustering process as a group (Figure 3(a)), while in the 6th semester (Figure 3 (f)) it appears together with FL and grouped with PD. This is due to the political alliance between the UDC and FL and to the movement of both parties in the direction of the center-left alliance.

It is worth to note as the main voting patterns revealed by hierarchical clustering totally agree with the results of the *SVD* analysis performed in the previous section.

B. Similarity matrix visualization

In order to visualize the similarity matrix M , a binarized matrix B has been built from M by assigning 1 to the element B_{ij} if $M(i, j) \geq 0.6$, and zero otherwise. B has been then reordered such that Parliamentarians of the same party are located as consecutive rows/columns.

Figure 4 shows how the two political parties PDL (rows 304:521) and LN (rows 45:98), that supported the center-right government, progressively reduce their intra-group similarity, while the opposition parties PD (rows 99:303), IDV (rows 24:44), and UDC (rows 546:578) present the

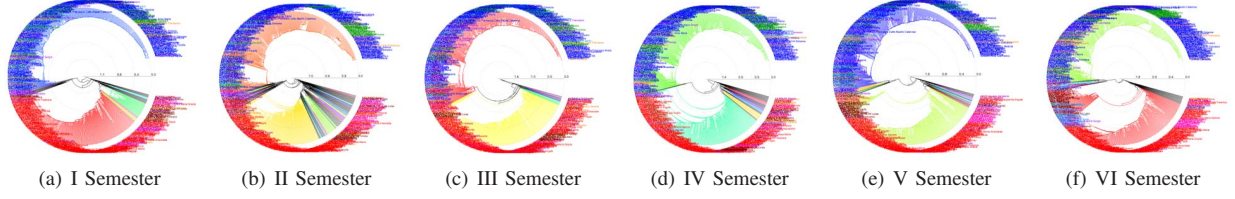


Figure 3. Dendrograms obtained by the single linkage clustering algorithm for each semester. Internal colors correspond to the clusters found by the algorithm, external colors to the true parties. The association color-party is the following: FL: cyan, IDV: magenta, LN: green, PD: red, PDL: blue, PT: orange, UDC: brown, Mixed: violet.

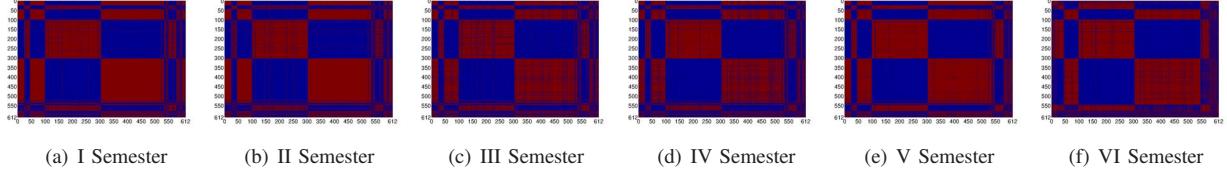


Figure 4. Visualization of the binary similarity matrices sorted by party membership, for each of the six semesters. The row intervals corresponding to each party are the following: FL [1:23]; IDV [24:44]; LN [45:98]; PD [99:303]; PDL [304:521]; PT [522:545]; UDC [546:578]; Mixed [579:612].

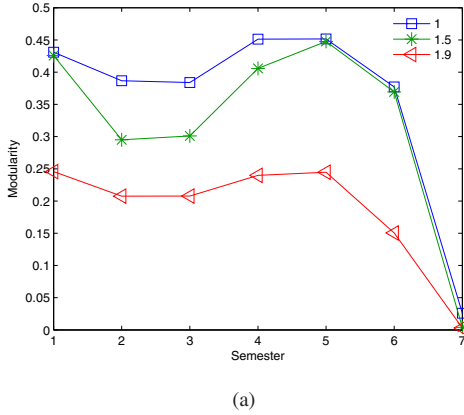


Figure 5. Modularity values.

opposite trend, i.e. in the first three semesters their intra-group similarity slightly diminishes, in the second three semesters, on the contrary, it increases. It is interesting to note that members of FL (rows 1:23) maintain their high similarity for all the periods, although they separated from PDL in 2010. Another important observation regards the new formed group PT, whose Representatives come from the center-left parties. Although this was constituted in the sixth semester to avoid the government fall, its members showed a good political affinity since the first semester (rows/columns 522:545). The figures clearly show the boosting of agreement from the first to the last semester.

C. Network representation of Italian Parliament voting records

In this section we apply network analysis to the voting records of Italian Parliament to verify if the results obtained

with the approaches employed in the previous sections are comparable when changing the analysis method. The binary matrix B , derived from the similarity matrix M , can be used to build an undirected and unweighted network $\mathcal{N}$, where nodes correspond to Parliamentarians and there exists an edge between two nodes p_i and p_j if the entry B_{ij} is 1. This means that two Representatives are connected if they voted in the same way in at least 60% of the overall roll calls. The community structure of $\mathcal{N}$ can then be investigated by optimizing the well known concept of *modularity* [12], based on the intuitive idea that a community should have more internal connections among its nodes than interconnections between its nodes and those in other communities. The modularity is defined as

$$Q = \frac{1}{2r} \sum_{ij} (A_{ij} - \gamma \frac{k_i k_j}{2r}) \delta(C_i, C_j)$$

where r is the number of edges in the network, k_i is the degree of node i , C_i is the community to which i belongs, and $\delta(C_i, C_j)$ is 1 if nodes i and j belong to the same community, 0 otherwise. γ is a resolution control parameter introduced by Granell et al. [13] to overcome the resolution problem stated in [14] and study community structure at multiple scales. In fact it has been proved that the optimization of modularity has a topological resolution limit that depends on both the total size of the network and the interconnections of groups. This implies that small, tightly connected clusters could not be found. Thus, searching for partitioning of maximum modularity, may lead to solutions in which important structures at small scales are not discovered. When $\gamma = 1$ the equation reduces to the standard formulation of modularity [12].

We used an algorithm optimizing modularity [15] extended with the resolution parameter, and executed the

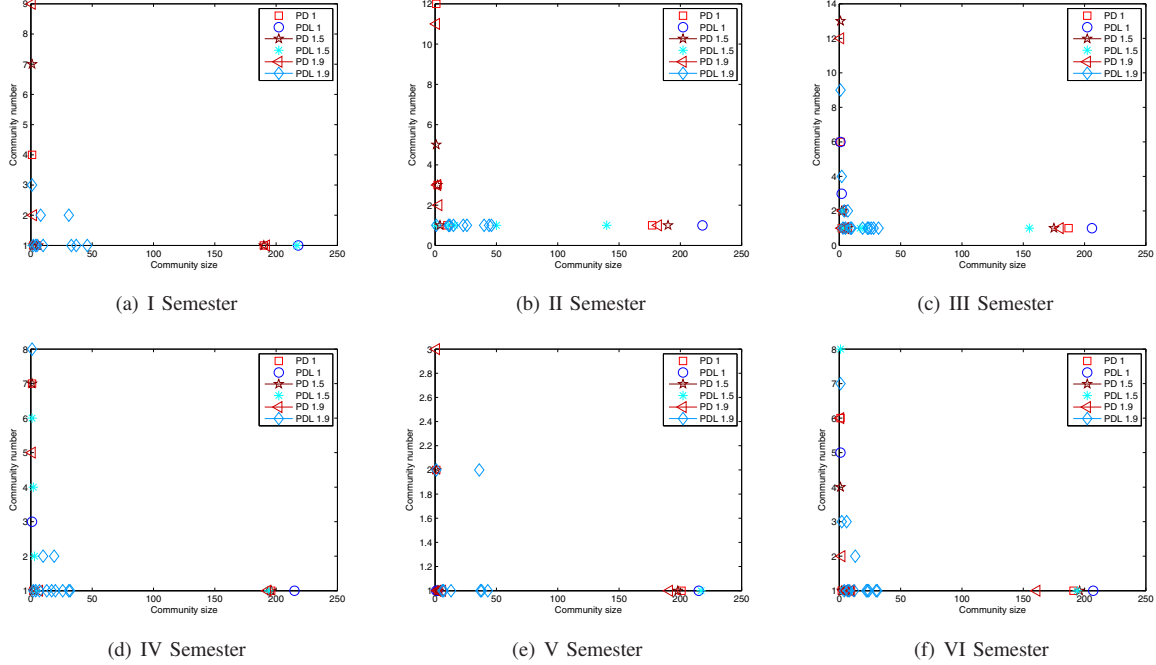


Figure 6. Number of communities in which the two main parties PDL and PD are split and respective size.

method with three different values of γ : 1, 1.5, 1.9. The latter two values have been chosen to analyze the existence of sub-communities inside those obtained with $\gamma = 1$ that cannot be found by optimizing modularity because of the resolution problem.

Figure 5 shows how modularity values vary during the seven semesters for all the three resolution parameters chosen. The figure clearly points out a sharp decrease of modularity in the 6th period and a drastic reduction in the 7th one. In order to better analyze the community structure detected by the algorithm, Figure 6 shows the number of communities in which the two main parties PDL and PD have been split. We do not report the results for the other parties because their behavior is analogous to the coalition they belong. Since the size of the largest community is 218 (i.e. the number of PDL members), the first coordinate varies between 1 and 218. The second coordinate, for each value of γ , reports the number of subgroups of that size obtained by the algorithm. Figure 6(a) shows that, with $\gamma = 1$ PDL is grouped in a unique community, while PD is clustered in a big community of 190 members and other 14 members are split in 7 small communities. When $\gamma = 1.5$ the situation is almost the same. However, when $\gamma = 1.9$, PD continues to have a big community of size 192, while PDL is split in 14 communities of size varying between 1 and 46. The very interesting result is that this behavior is maintained for all the semesters. Thus, while PD remains cohesive for all the semesters, independently of the γ value, PDL is divided in many subgroups since the first semester, when its degree of

aggregation was considered very high, and as obtained with the other approaches described in the previous sections.

Thus modularity allows a more deep analysis of the internal agreement of parties and can provide insights of early and unexpected changes a political party could encounter. Moreover, it affords an explicit and clear view of the steady fragmentation of the coalition endorsing the center-right government that culminated in its fall.

V. THE 7TH SEMESTER

The analysis described in the previous sections mainly considered the first six semesters. We decided to separate the last semester because the voting behavior of Parliamentarians had an abrupt alteration, as testified also by the results obtained by all the employed methods. First of all, the number of voted measures is less than the fifth part of the other semesters. Furthermore, it happened that the political party organization completely disappeared, and each Parliamentarian voted independently of his group.

Figure 7 gives a clear representation of this situation. In fact, the application of *SVD* on this semester (Figure 7(a)) shows a polarization of all the parties on the first coordinate, and distinguishes between center-left and center-right only on the bipartisan coordinate. Hierarchical clustering returns a unique cluster including all the parties (Figure 7(b)), and the visualization of the voting matrices (Figure 7(c)) depicts high fragmentation. Finally, Figure 7(d) shows that modularity optimization with $\gamma = 1$ extracts a group of 156 and another of 19 members from PD, and two groups of

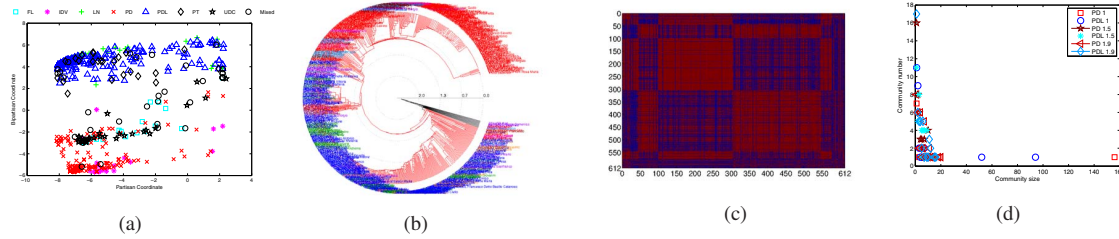


Figure 7. Results obtained by applying SVD (a), hierarchical clustering (b), visualization of the similarity matrix (c), and community detection (d) on the 7th semester.

94 and 52 members from PDL. However these groups are clustered together, thus confirming the results of the other approaches. For higher values of γ , both parties are split in small groups of at most 20 Parliamentarians, and the communities found are constituted by members of almost all the political parties.

It worth to note that, as already pointed out, Figure 5 indicates an abrupt lowering of modularity value in the 7th semester that explains the loss of community structure.

VI. CONCLUSIONS

The paper presented an investigation of the voting behavior of Italian Parliament in the last years by employing different computational tools. Though studies of this kind exist for different political institutions from US and Europe, as far as we know, this is the first tentative of exploring Italian Parliament with data mining and network analysis methods. We generated networks among the Parliamentarians at consecutive time periods and investigated community structure at multiple scales. By delving the voting records of Representatives, we were capable of characterizing the organizational structure of Parliament, and to discover latent information contained. All the methods used showed to be effective at identifying political parties, and at providing insights on the temporal evolution of groups and their cohesiveness. Future work aims at applying overlapping community detection methods to better uncover hidden collaborations among Parliamentarians of different political membership.

REFERENCES

- [1] A. Jakulin and W. Buntine, "Analyzing the US Senate in 2003: Similarities, networks, clusters and blocs," 2004. [Online]. Available: <http://eprints.fri.uni-lj.si/archive/00000146/>.
- [2] A. Pajala, A. Jakulin, and W. Buntine, "Parliamentary group and individual voting behavior in Finnish Parliament in year 2003 : A group cohesion and voting similarity analysis," 2004. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.103.2295>.
- [3] M. A. Porter, P. Mucha, M. E. J. Newman, and A. J. Friend, "Community structure in the United States House of Representatives," *Physica A: Statistical Mechanics and its Applications*, vol. 386, no. 1, pp. 414–438, Dec. 2007.
- [4] Y. Zhang, A. Friend, A. L. Traud, M. A. Porter, J. H. Fowler, and P. J. Mucha, "Community structure in Congressional cosponsorship networks," *Physica A: Statistical Mechanics and its Applications*, vol. 387, no. 7, pp. 1705–1712, Mar. 2008.
- [5] A. S. Waugh, L. Pei, J. H. Fowler, P. J. Mucha, and M. A. Porter, "Party polarization in Congress: A network science approach," 2009. [Online]. Available: <http://arxiv.org/abs/0907.3509>.
- [6] K. T. Macon, P. J. Mucha, and M. A. Porter, "Community structure in United Nations General Assembly," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 1-2, pp. 343–361, Jan. 2012.
- [7] S. Hix, A. Noury, and G. Roland, "Power to the parties: Cohesion and competition in the European Parliament," *British Journal of Political Science*, vol. 35, no. 2, 2005.
- [8] G. Strang, *Linear Algebra and its Applications*. IV edition, Thomson Brooks/Cole, 2005.
- [9] T. J. Brazill and B. Grofman, "Factor analysis versus multi-dimensional scaling: binary choice roll-call voting and the us supreme court," *Social Networks*, vol. 24, no. 3, pp. 201 – 229, 2002.
- [10] A. Jain and R. C. Dubes, *Algorithms for Clustering Data*. Prentice Hall, 1988.
- [11] P. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Pearson International Edition, 2006.
- [12] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review*, vol. E69, p. 026113, 2004.
- [13] C. Granell, S. Gómez, and A. Arenas, "Unsupervised clustering analysis : A multiscale complex network approach," *Journal of Bifurcation and Chaos*, in press, 2012.
- [14] S. Fortunato and M. Barthélemy, "Resolution limit in community detection," *Proc. National Academy of Science, USA*, vol. 104, no. 36, 2007.
- [15] C. Pizzuti, "Boosting the detection of modular community structure with genetic algorithms and local search," in *Proc. of the 27th Symposium on Applied Computing (SAC 2012)*, 2012, pp. 226–231.