POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

Twitter communities of users during 2022 Italian political elections: a network analysis

Tesi di Laurea Magistrale in
Mathematical Engineering - Ingegneria Matematica

Ilaria Saini, 969334

Advisor:
Prof. Carlo Piccardi

Co-advisors:
Francesco Pierri

Academic year:
2022-2023

Abstract: In this study, we conduct an analysis of a Twitter dataset focusing on the 2022 Italian political elections in the context of complex networks analysis. The dataset is aggregated on a weekly basis, consisting in 15 weeks of data, and the investigation primarily focuses around two sets of networks: one being the retweet networks encompassing all users, and the other considering the retweets only involving elected members. In both networks, nodes represent Twitter users, connected by undirected links with weights corresponding to the number of interactions, such as retweets or mentions, between users. Our primary objective is to examine the topological characteristics that unveil the dynamics of information propagation and patterns of social influence within the Twitter networks. To accomplish this, we employ the Louvain Community Detection Algorithm to identify communities in each network, defined as cohesive groups of users, and link them to the different political coalitions, such as CDX, CSX, M5S and CEN. Furthermore, a network analysis approach enables us to determine the different roles played by individual users, highlighting the major contributors to the political discourse surrounding the elections. Lastly, we employ a time series analysis in order to explore various network characteristics over the entire duration of interest. The analysis of the networks and their properties is conducted using Python programming language, specifically the NETWORKX package.

Key-words: Network Analysis, Twitter data, Italian political elections, Community detection, Role analysis, Time series

1. Introduction

In recent years, the advent of social media platforms has revolutionized communication and information exchange. Twitter, in particular, with its 368 million monthly active users worldwide [26], has emerged as a powerful platform for news diffusion, opinions sharing and ideas propagation among its vast network of interconnected users. Consequently, the analysis of Twitter data has become a subject of significant interest for researchers, offering valuable insights into various phenomena due to the easily accessible information.

The study of complex networks has provided a powerful framework for understanding and analyzing the structures and dynamics underlying complex systems. In particular, by considering Twitter data as a complex network, where users represent nodes and interaction represent edges, we can explore the flow of information between users. Remarkably, 48% of Twitter users employ the platform as a source of news [26], thus increasing the interest in studying Twitter data. Such studies could lead to deeper insights into the interplay between social interactions and information diffusion.

This thesis aims to explore the fundamental principles of complex networks theory within the context of Twitter's networked ecosystem. We will employ fundamental concepts of network analysis such as centrality metrics and community structures to unravel the dynamics at play. Specifically, we will investigate the specifics of Twitter as a complex network with its unique features, such as retweets and other user interactions. Through this investigation, our objective is to reveal the dynamics of information propagation and patterns of social influence within the Twitter network.

In particular, this thesis focuses on analysing the complex networks associated to the Italian political elections held on September 25th, 2022, which resulted in the ascent to power of right-wing parties with more then 40% of votes[1]. In this context, we will examine the influence of key political actors within the political discourse surrounding this event and their connections with other users. Furthermore, through network analysis, we will study the formation and evolution over time of communities linked to political coalitions.

As a starting point, in Section 2 we conducted a literature review to contextualise this work within the context of network analysis based on Twitter data, particularly in relation to political subjects, both in Italy and the United States. Subsequently, in Section 3 we described the dataset and the methods employed for our analysis. We conducted a comprehensive network analysis for each obtained network, the results are discussed in Section 4. The results of community detection are commented in Section 5 and a role analysis is presented in Section 6. Furthermore, a time series analysis in discussed in Section 7. Finally, we concluded this work by highlighting the major findings and suggesting potential further analysis.

2. Literature review

The analysis of complex networks based on Twitter data is not a novelty. Over the past 15 years, numerous studies have been conducted in this field. For example, in [9], the authors investigated the phenomenon of homophily on Twitter using a dataset of political tweets from users following Democratic or Republican accounts in 2009. The study aimed to examine whether Twitter functions as an echo chamber or a public sphere in terms of users' political orientations. The findings indicated a prevalence of echo chamber dynamics, where users interacted mainly with individuals who share similar opinions.
In [10] and [11] the primary focus was the political polarization on Twitter and predicting the political alignment of users during the six weeks leading up to the 2010 U.S. congressional midterm elections. The studies found that Twitter exhibited a high degree of political polarization with limited connectivity between left-leaning and right-leaning users. Furthermore, it highlighted the potential for accurately predicting users' political orientations using content-based and interaction-based approaches, with the application of network clustering techniques.
Examining the 2016 U.S. presidential elections, [5] and [6], investigated the impact of fake news on Twitter and validated opinion trends on the platform by comparing them with national polling aggregates. The use of a Collective Influence Algorithm revealed the widespread dissemination of false or ambiguous news. By classifying tweets as supportive or opposing to top candidates, the analysis demonstrated a significant correlation between Twitter data and national polls, suggesting its potential for providing valuable insights into public opinion.

Similar studies have been conducted in Italy. In [7], an analysis of user behaviour on Twitter during the Italian political elections in February 2013 demonstrated that the volume of tweets related to political party leaders served as a reliable indicator of elections outcomes. Additionally, the study provided a geographical analysis illustrating how political disparities among different regions of Italy were reflected in Twitter data.
The research presented in article [21] focused on Twitter semantic networks during the Italian elections in March 2018. It examined election-related tweets to understand communication dynamics on the platform, highlighting the existence of user communities discussing specific topics and the influence of key actors in information dissemination. The authors also analyzed the evolution of semantic networks over time studying changes during the election campaign.
The phenomenon of disinformation diffusion on Twitter during the 2019 European elections in Italy was analyzed in [19]. The study aimed to examine the existence and characteristics of disinformation campaigns and their impact on election-related debates. The collected data included tweets containing explicit URLs associated with Italian disinformation websites published in the five months preceding the elections. The research revealed the presence of disinformation confined to a specific community associated to the Italian conservative political

[1] https://elezionistorico.interno.gov.it

sphere.
In [23], a comprehensive overview on the impact of social media platforms, specifically Twitter and Facebook, on political elections in Italy from 2013 to 2020 was provided. The authors aimed to give a survey of the evolving research questions, data collection methods, and analytical approaches employed in this field of study. The research highlighted the role of Twitter as a prominent platform for political information diffusion. Many studies have focused on this specific social media platform due to its easy accessibility to data and interpretability.

These methodologies are not limited to politics and have been applied to other areas of public debate as well. For example, in [13], researchers applied community detection to data collected during the initial months of the COVID-19 outbreak. The work in article [25] focused on measuring user engagement with low credibility media sources, analyzing how users interacted with these sources and whether there were differences compared to reliable sources. Finally, the analysis of coordinated networks and their influence on social media platforms users was explored in [17].

3. Dataset and methodology

3.1. Data collection

In this thesis, we employed a data collection of Italian-language tweets related to the Italian political elections held on September 25th, 2022. The dataset covers the period form July 1st to October 20th, 2022, a time interval covering the period of the electoral campaign and one month after the Election Day. To collect tweets, a snowball sampling approach was employed, based on a list of keywords, such as "elezioni2022", "elezioni" and the names of the prominent political figures. The collected dataset consists of 19,087,594 tweets shared by 618,089 unique users, identified by their user IDs. For further details on the data collection process refer to [20]. Additionally, we manually compiled a list of user IDs corresponding to each elected Parliament member active on Twitter based on the official lists released by the Senate[2] and the Chamber[3] of Deputies. We also included the user IDs of the main political parties' official accounts in our list.

3.2. Data cleaning

On Twitter, there are several types of user interactions, which can be divided into four distinct types of tweets:
- Original tweet: This type of tweet is created and authored by the user, it contains original content;
- Reply: When a user directly responds to another user's tweet;
- Retweet: When a user shares a tweet from another account without adding any additional commentary;
- Quote tweet: When a user retweets a tweet from another user while appending their own comment or perspective.

In addition to these tweet types, each tweet can also include mentions to other users, indicating their inclusion or reference within the content of the tweet. An example of a retweet with mentions, identified by @user, can be seen in Figure 1a.

In line with the objective of this study, the analysis focused exclusively on retweets and their associated mentions as they play a significant role in the diffusion of political discourse on Twitter. Previous researches, as shown in [11], have demonstrated that retweets are a valuable indicator for predicting users' political alignment. Retweets allow users to share content generated by others, indicating a consensus or agreement with the content of the original tweet.

To attenuate the fluctuations inherent in daily recordings, the collected data for this study was aggregated into weeks, going from Monday to Sunday, covering a total of 15 weeks from July 4th to October 16th, 2022. This aggregation facilitates the identification of key periods, such as the election week (week 12) and the week immediately following the elections (week 13). These aggregated datasets serve as the basis for constructing the social networks for each week which are then analyzed within the scope of this research.

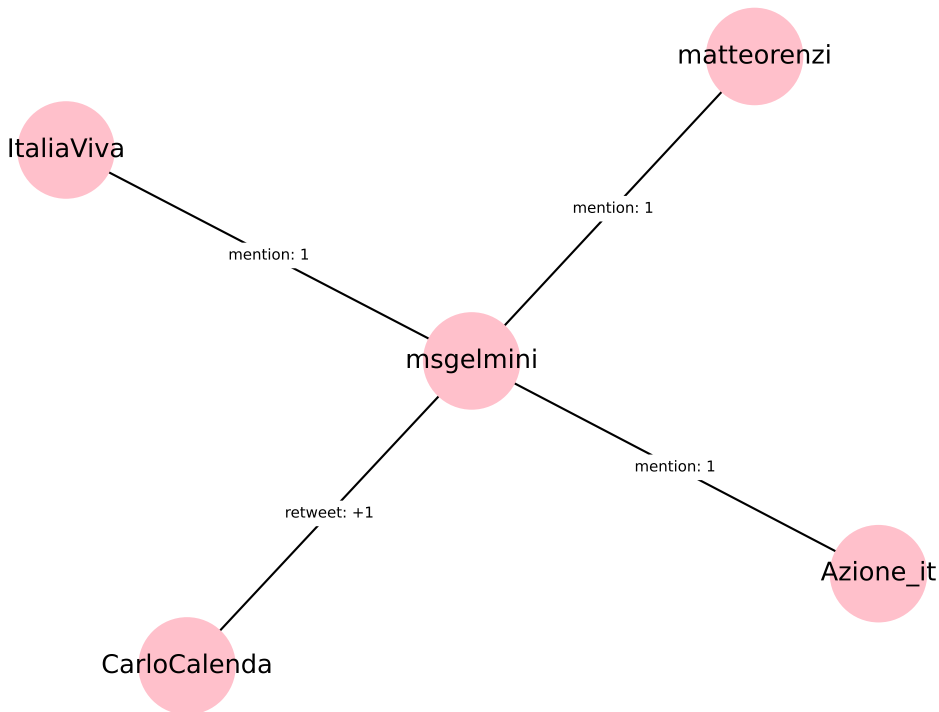
3.3. Methods

We constructed social networks using the aggregated data by assigning a node to each unique user. Users who shared retweets were connected to the users whose content was retweeted through undirected links. In

[2] https://www.senato.it/leg/19/BGT/Schede/Attsen/Sena.html
[3] https://www.camera.it/leg19/28

this context, nodes represent individual users, while links or edges depict connections and interactions between these users. The networks were weighted, taking into account the number of interactions between two users to determine the strength of the corresponding link. Additionally, we considered mentions within tweets, which created connections between the retweeting node and the mentioned user. However, mentions did not contribute to the weight of an already existing link. An example of a graph created from a retweet with mentions can be found in Figure 1b.

(a) Example of a retweet with mentions.



(b) Example of the graph created from a retweet with mentions.

Figure 1: (a)(Top): *msgelmini* retweets from *CarloCalenda*, the original tweet contains three mentions: *matteorenzi*, *ItaliaViva* and *Azione_it*. (b)(Bottom): the central node, corresponding to *msgelmini*, is connected to the other four nodes. Three connections are through mentions and do not add any weight to the edge (fixed weight 1). The connection between *msgelmini* and *CarloCalenda* is through a retweet thus it adds +1 weight to the link.

In our analysis we employed various techniques from the network science theory [1], including community detection and centrality measures. We also implemented a role analysis to highlight the different roles played by nodes within the network. Furthermore, we conducted time series analysis to examine temporal differences in network and community structures on a weekly basis. For network analysis we employed the NETWORKX

Python package [15].
Community detection is used to identify groups of nodes with strong internal connectivity allowing to understand the structure and function of the network. In this study, we employed the *Louvain Community Detection Algorithm* [4], which is based on modularity optimization. This algorithm enables to measure how much the community structure of the network deviates from a random arrangement using graph randomization techniques.
Role analysis is employed to uncover the role of each node within the network. In particular centrality measures are an important metric to quantify the significance of a node within a network. We used two centrality measures: *strength centrality* s_i [1], which counts the number of weighted edges a node is connected to, and *eigenvector centrality* γ_i, which is defined by taking into account the importance of a node's neighbors [16]. These measures provide insights into the relative importance and influence of nodes within the network structure. Furthermore, we used a *z-P analysis*, which has been previously applied in [14] and [8], where z is the *within-community strength* and P the *participation coefficient*. With this analysis we wanted to determine the extent to which a node is connected within its community (z) or outside of it (P). We compared the results with the centrality measures to verify if specific roles are associated to high strength and/or eigenvector centrality.
The time series analysis aimed to identify differences and similarities in the network structure across the considered period of time. We calculated the distance between each week using the *Weighted Jaccard distance* to compare networks [24] and detect any particular event. We visualized the flow of nodes between different communities, always identified by the Louvain Algorithm, in consecutive weeks using a visual tool called *Sankey diagram*. Finally, we computed the overall share of each political coalition until the Election Day.

4. Twitter Network

4.1. Basic network metrics

Using the methodology described in Section 3.3 we generated 15 weighted and undirected networks, each corresponding to an aggregated week of data. From these networks, the giant connected component was extracted, which consistently contained at least 94% of the total nodes. Figure 2 presents basic network metrics as functions of time. These data reveal an increasing level of interest during the election weeks (week 12 and 13) with a higher number of nodes N and edges L but a lower average strength[4] $\overline{S}$, particularly in week 13. This observation suggests that a substantial number of less active users engaged in the discussion on the elections during the days immediately preceding and following Election Day. The density ρ, defined for an undirected graph as

$$\rho = \frac{2L}{N(N-1)},$$

is consistently below 0.0003, indicating sparse networks with relatively few connections. The average clustering coefficient $< C >$, defined as

$$< C >= \frac{1}{N} \sum_{i=1}^{N} C_i,$$

where C_i indicates the local clustering coefficient[5] of node i, is consistently under 0.002, suggesting a network structure resembling that of a tree with a low density of triangles. The K-core of a graph is the maximal subgraph that contains nodes of internal degree K or more, then the K-core number K_c is defined as the highest value of K in the network [3], in Figure 2 it shows almost constant values for central weeks.

[4]The strength of a node is the sum of the weights of the links incident in the node [1].
[5]The clustering coefficient of node i is the number of direct connections between the neighbours of i with respect to the maximum possible number of such connections [1].

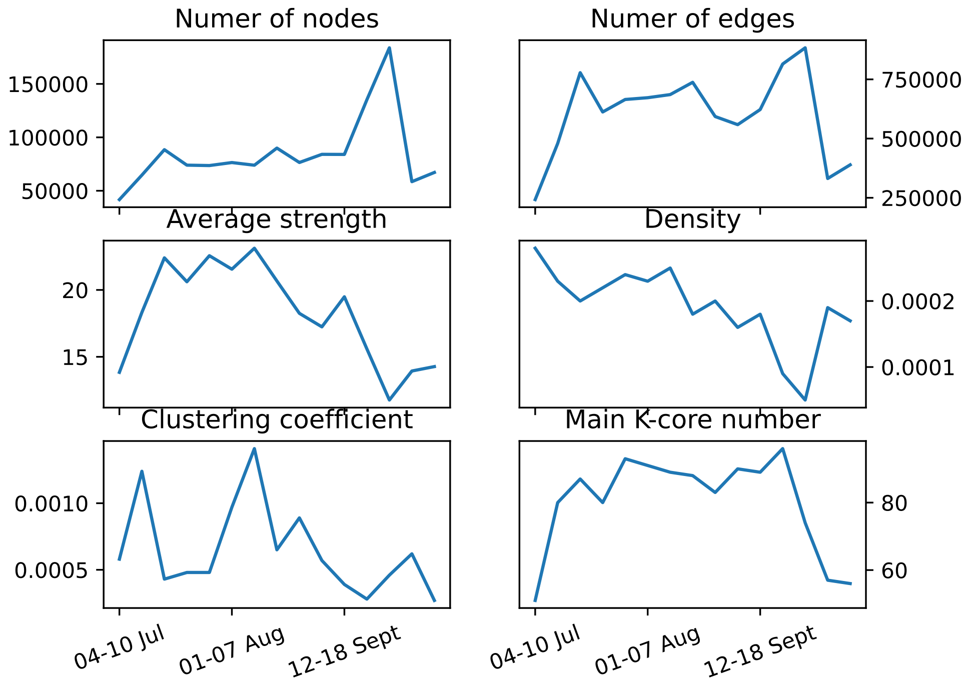
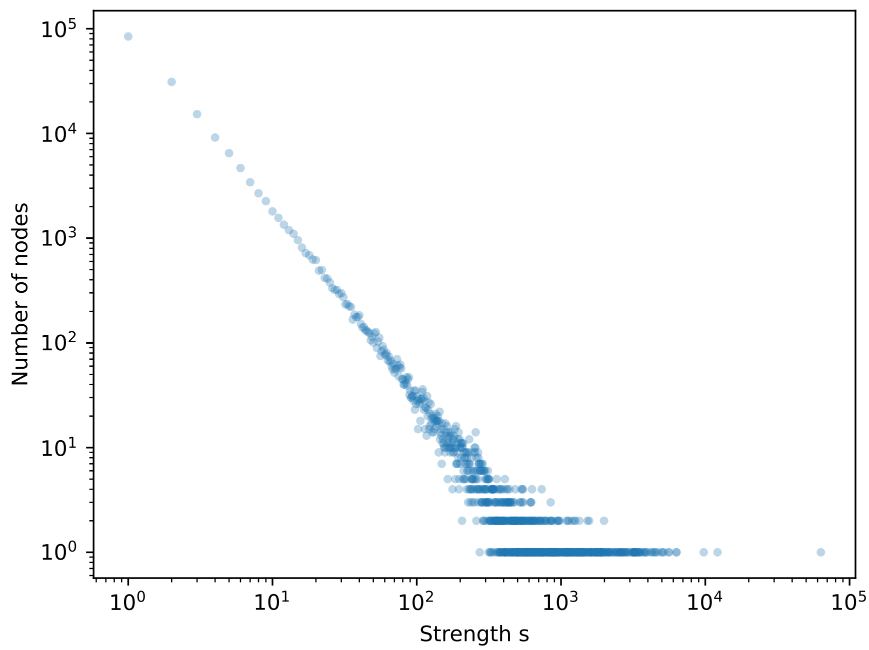
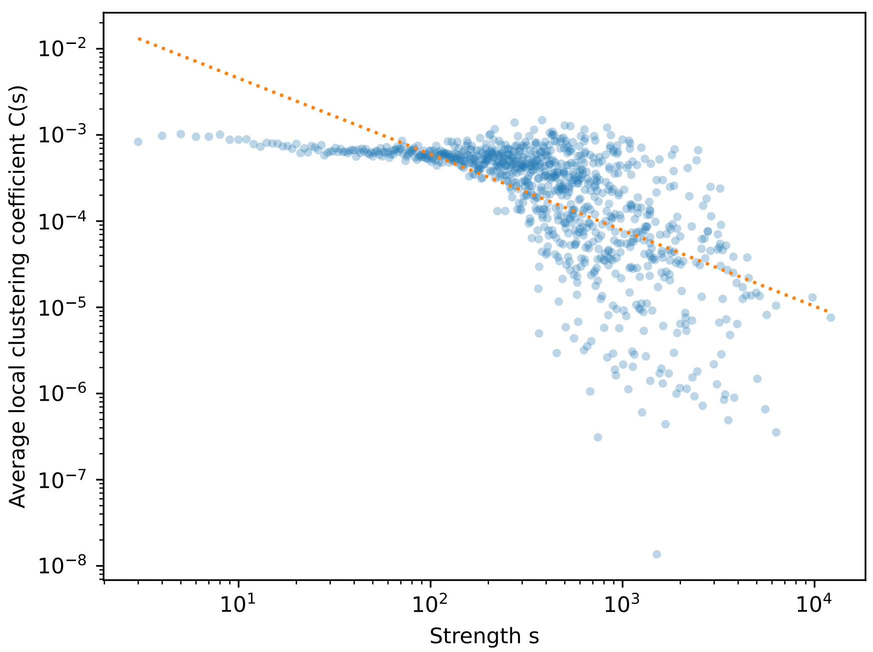


Figure 2: Number of nodes N, number of edges L, average strength $\overline{S}$, density ρ, average clustering coefficient $< C >$ and main K-core number K_c as functions of time.

Plots illustrating the distribution of strength and the local clustering coefficient as a function of strength were generated for each week. Figure 3 presents the plots related to week 13. In Figure 3a, a limited number of highly connected nodes, known as hubs [1], is observable. Additionally, there exists a wide spectrum of node strength, ranging from 1 to 10^5. A more detailed analysis of node strength can be found in Section 6.1. Figure 3b demonstrates the decreasing trend of $C(s)$, i.e. the average C_i for all nodes with same strength s, as a functions of s, indicating a hierarchical organization in the network's structure [22]. This suggests that nodes with lower strength tend to be situated within dense subgraphs, while nodes with higher strength act as connectors for the overall network. These observed patterns remain consistent across all weeks, highlighting the absence of significant structural differences among the networks.



(a) Strength distribution.



(b) Local clustering coefficient as a function of strength.

Figure 3: (a)(Top): Plot of the strength distribution of the network relative to week 13. (b)(Bottom): In blue, plot of the average local clustering coefficient as a function of the strength of the network relative to week 13. In orange, the linear regression of the data.

4.2. Networks of the elected members

In order to gain a deeper understanding of the propagation of political consensus, a specific focus was placed on the retweet networks associated with the official accounts of elected individuals. Specifically, we only considered retweets, with mentions, where the sharing or shared user is an elected member. These networks contains both elected members and individuals connected with them through retweets or mentions. Similar to the previous networks, only the giant connected component of these weighted undirected networks was considered. Figure 4 provides, as shown before, basic network metrics as functions of time. As expected, the number of nodes and edges is significantly lower compared to the previous case; the average strength and the main K-core number are significantly lower too. Regarding the density and the average clustering coefficient, the analysis conducted on the networks associated with the official accounts of the elected individuals yielded similar results to the previous case. The density ρ of these networks consistently remained below 0.0006, indicating sparse networks with relatively few connections between nodes. The average clustering coefficient $< C >$ also remained below 0.008, suggesting a low density of triangles and a network structure resembling that of a tree.

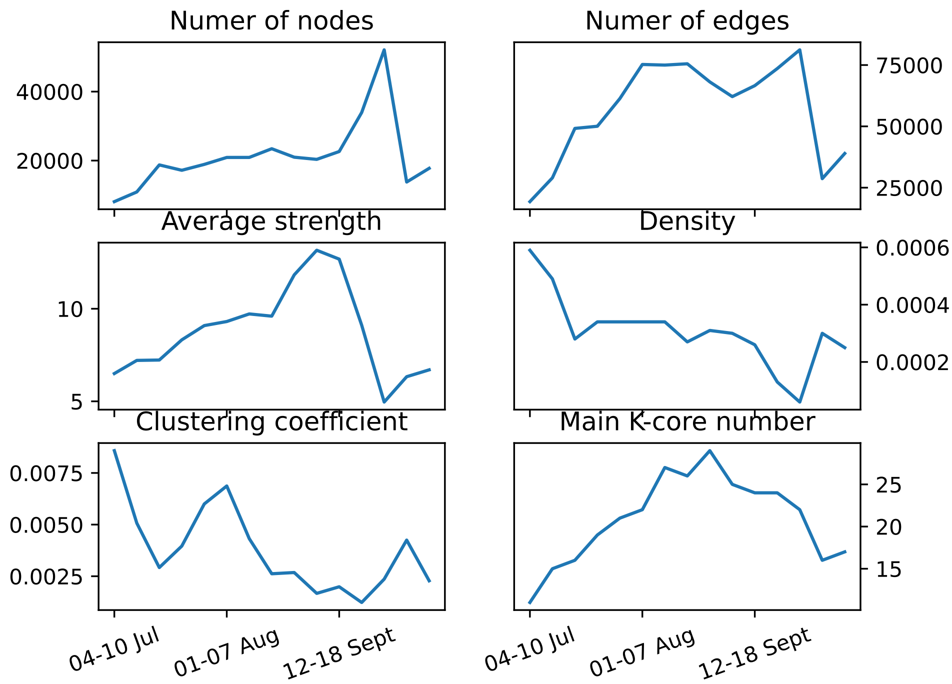
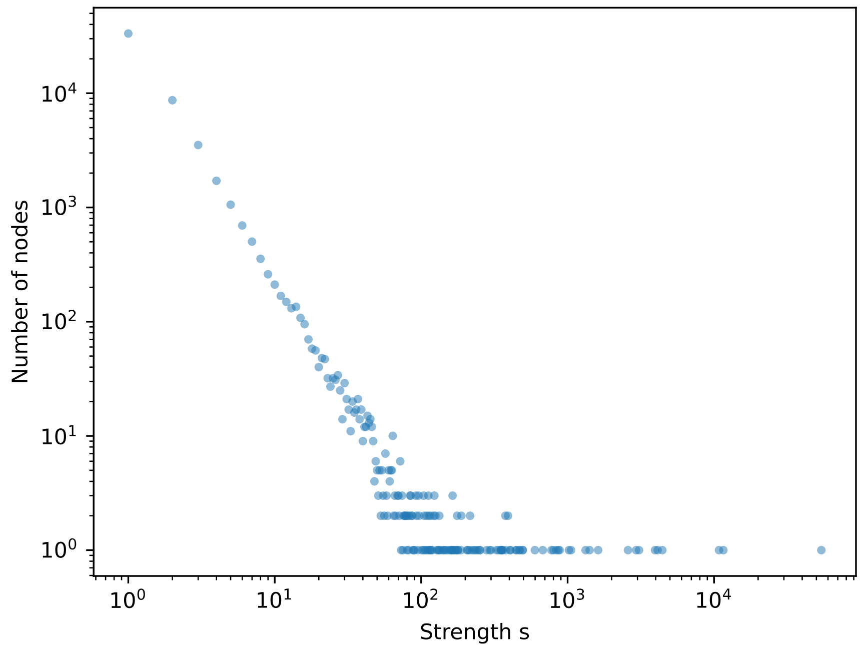


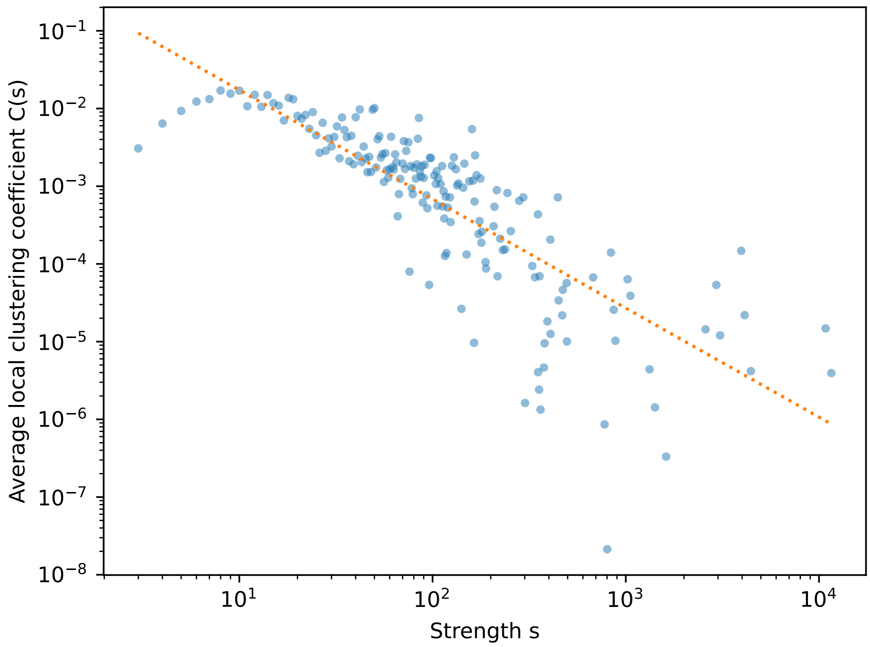
Figure 4: Number of nodes N, number of edges L, average strength $\overline{S}$, density ρ, average clustering coefficient $< C >$ and main K-core number K_c as functions of time for the network of elected members.

The plots in Figure 5, which illustrate the strength distribution and the local clustering coefficient as a function of strength, exhibit similarities to the previous case. The strength distribution shows a low number of highly connected nodes, who play a critical role in the connections within the network, and a wide range of node strengths. The local clustering coefficient demonstrates a decreasing trend, indicative of a hierarchical organization [22] also within this network.

In conclusion, we can affirm that no significant structural differences were observed within the analysis that was conducted on the two sets of networks.



(a) Strength distribution.



(b) Local clustering coefficient as a function of strength.

Figure 5: (a)(Top): Plot of the strength distribution of the network of the elected members relative to week 13. (b)(Bottom): In blue, plot of the average local clustering coefficient as a function of strength of the network of the elected members relative to week 13. In orange, the linear regression of the data.

5. Community detection

5.1. Political coalitions

Community detection using the Louvain Community Detection Algorithm [4] was performed on each week's network. From the resulting partition, only communities with more then 1% of the total number of nodes in the graph were considered. The objective was to assign each community to a specific political party or coalition, including:

- CDX: consisting of the right-wing parties *Fratelli d'Italia, Lega, Forza Italia* and *Noi Moderati*;
- CSX: consisting of the left-wing parties *Partito Democratico* and *Alleanza Verdi e Sinistra*;
- M5S: consisting of *Movimento 5 Stelle*;
- CEN: consisting of the coalition of *Azione* and *Italia Viva*.

Initially, we associated each elected member with their respective political party. Subsequently, to associate communities with political coalitions, the total strength associated with each party was computed within each community by summing the individual strengths of all elected members in the community. A community was considered associated with a particular political coalition if that coalition's strength exceeded 80% of the combined strength of all other coalitions and accounted for more than 20% of the total strength of the community. This approach ensured that only communities where the majority of links were connected to elected members of a certain political coalition were associated with it.

Table 1 provides information on the modularity of the partition and the number of communities identified for each week, categorized according to the political coalitions. The modularity Q serves as a metric for assessing the quality of a partition by measuring the deviation of edge densities within and between communities from what would be expected in a random distribution. Specifically, modularity Q is calculated as the difference between the fraction of edges within communities and the expected fraction of edges in a random distribution.

The Table indicates that every partition exhibits a modularity value greater than 0.54, indicating the presence of meaningful community structure. However, the identified communities are not associated with specific political coalitions, with the exception of a CDX community that is consistently present and a CSX community found only in specific weeks (4, 14 and 15). No communities are specifically linked to the M5S and CEN coalitions.

Communities in the networks

	Q	n°	CDX	CSX	$M5S$	CEN
week1	0.56	9	1	0	0	0
week2	0.56	5	0	0	0	0
week3	0.54	7	1	0	0	0
week4	0.54	8	1	1	0	0
week5	0.54	8	1	0	0	0
week6	0.55	7	1	0	0	0
week7	0.54	6	1	0	0	0
week8	0.55	7	1	0	0	0
week9	0.55	10	1	0	0	0
week10	0.59	9	0	0	0	0
week11	0.57	10	1	0	0	0
week12	0.61	10	1	0	0	0
week13	0.62	8	1	0	0	0
week14	0.57	10	1	1	0	0
week15	0.55	8	1	1	0	0

Table 1: The modularity of the partition, the number of communities found for each week and the number of communities associated with every political coalition.

This observation suggests two possibilities: either the discourse around the political elections extends beyond the considered political actors, as it involves other individuals and groups; or each coalition does not form a distinct and cohesive community. In the latter case, elected members from different parties may be grouped together within the same community, making it difficult to associate the community with a specific political coalition.

In the analysis of the networks of the elected members, the same community detection approach was applied and the results are presented in Table 2. This Table provides information on the modularity of the partition and the number of communities identified for each week, categorized by political coalitions. It is observed that the modularity values are consistently higher compared to the previous case, indicating better-divided communities. Furthermore, almost every community is associated with a specific political coalition.

Communities in the networks of the elected members

	Q	$n°$	CDX	CSX	$M5S$	CEN
week1	0.71	9	4	2	1	2
week2	0.68	7	2	1	1	1
week3	0.63	9	4	2	1	2
week4	0.66	9	4	2	1	2
week5	0.64	6	3	0	1	1
week6	0.62	8	3	2	1	1
week7	0.57	7	2	2	1	1
week8	0.60	8	2	3	1	2
week9	0.65	7	3	2	1	1
week10	0.67	6	2	2	1	1
week11	0.68	6	2	2	1	1
week12	0.71	7	3	2	1	1
week13	0.64	7	4	1	1	1
week14	0.69	7	4	1	1	1
week15	0.66	7	3	2	1	1

Table 2: The modularity of the partition, the number of communities found for each week and the number of communities associated with every political coalition for the network of the elected members.

For the CDX coalition, the number of communities varies between 2 and 4. Upon manual examination of the elected members in each community, it is observed that *Fratelli d'Italia* and *Lega* typically form two separated communities. Sometimes, *Forza Italia* is also distinct, while *Noi Moderati* consistently belongs to another CDX's community. It is also noteworthy that the larger parties *Fratelli d'Italia* and *Lega* are occasionally split into two communities, with one being closer to their leaders, respectively *GiorgiaMeloni* and *matteosalvinimi*.
Regarding the CSX coalition, it is observed that the communities of *Partito Democratico* and *Alleanza Verdi e Sinistra* are usually separated. In weeks 3 and 8, *Partito Democratico* is divided into two communities, with one being larger and the other centered around *CottarelliCPI*. In week 5, the larger community is associated with no political coalition as it contains the leaders of *Partito Democratico, Alleanza Verdi e Sinistra* and *Azione*. This week coincides with the announcement of the sudden end of the formerly announced alliance of *EnricoLetta* and *CarloCalenda*, who were respectively leaders of *Partitio Democratico* and *Azione*.
The M5S coalition remains consolidated into a single community across all networks and weeks. The community associated to this coalition is consistently the one with greater persistence probability u_{cc}, consistently $\geq$ 87%. For an undirected network, the persistence probability is defined as the ratio between the total internal strength and the total strength of a community C. It represents the fraction of the strength of the nodes of community C that remains within C [18]. High values of persistence probability indicate the tendency of nodes to maintain their interactions and connectivity within the community, therefore M5S coalition is a relatively closed community.
For the CEN coalition, a consistent pattern emerges with the presence of either 1 or 2 communities. Before week 6, there are two distinct communities representing *Italia Viva* and *Azione*. However, this pattern is not true for weeks 2 and 5, as only one community is associated to *Italia Viva* while *Azione* becomes part of an-

other community. From week 6, there is only a single, larger community including both party leaders, namely *matteorenzi* and *CarloCalenda*, with the exception of week 8 where they remain separated. This finding carries particular significance because in week 6, the two leaders made the announcement of their coalition.

Figure 6 gives a visual representation of the partitioning of communities within the network of the elected members in week 11, using Gephi software version 0.9.7 [2]. In this visualization, each community is colored based on its corresponding political coalition: blue represents CDX, red represents CSX, yellow represents M5S and pink represents CEN. The node sizes are proportional to their respective strengths and the names of nodes with higher strength are displayed.

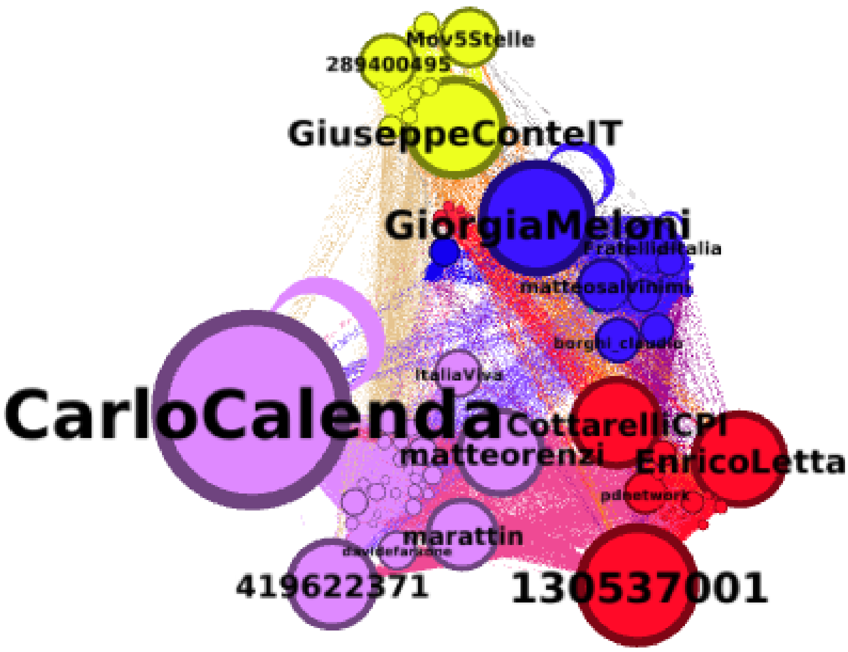


Figure 6: Communities identified in week 11 for the network of elected members. We identified 6 communities, 2 for CDX in blue, 2 for CSX in red, 1 for M5S in yellow and 1 for CEN in pink. It can be visually verified that the principal political actors belong to communities of their political coalition.

To gain a deeper understanding of the interconnections between various communities, in Figure 7 we have generated pie charts for each community depicting the proportion of weighted edges directed towards other communities. The figure specifically represents the outcomes for week 12, with the sizes of each pie chart being proportional to the number of nodes within the respective community. From the figure, we observe that the majority of links within each community are directed internally, indicating a strong intra-community cohesion. Additionally, there are some connections observed between CSX and CEN coalitions. For a more comprehensive analysis of the identified communities within each network of the elected members, refer to Appendix A.

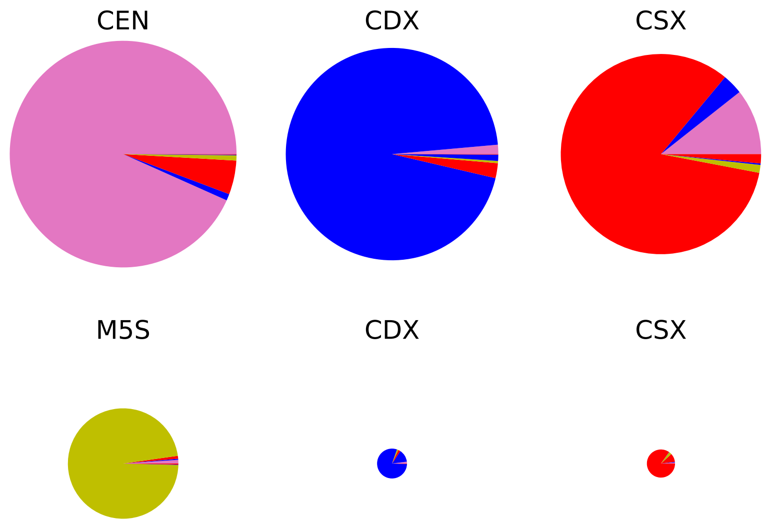


Figure 7: Pie charts of the communities identified in week 11 for the network of the elected members representing the interaction between communities. We identified 6 communities, 2 for CDX in blue, 2 for CSX in red, 1 for M5S in yellow and 1 for CEN in pink ordered by number of nodes. The majority of links within each community are directed internally, indicating a strong intra-community cohesion.

5.2. Coloring nodes

To gain a deeper understanding of the connections within each network, we have employed a particular visual representation. Specifically, we have associated a pie chart of colors to each node, indicating the proportion of weighted interactions with elected members or other users. By computing the weighted edges of each node, we color the node based on the affiliation of the connected individuals. The color scheme is the following: blue for CDX members, red for CSX members, yellow for M5S members, pink for CEN members and grey for non-elected users. This process is illustrated in the example provided in Figure 8. For elected members, we have assigned a single color based on their respective political coalition.

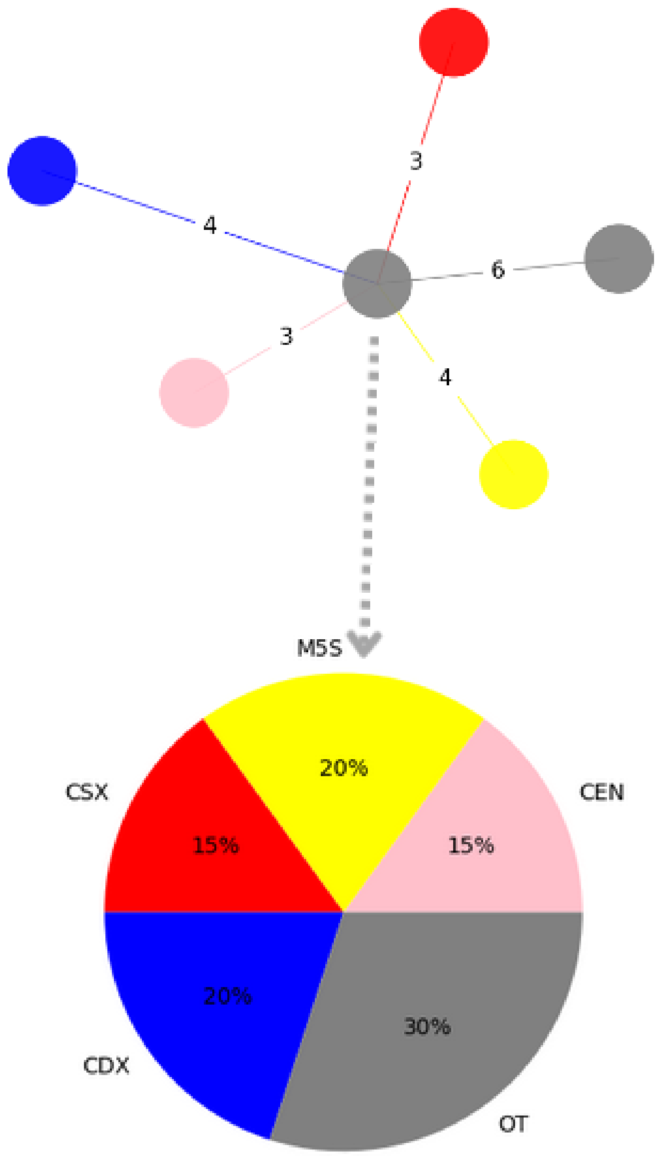


Figure 8: Example of a pie chart for a node considering its connections with other users. For example, the node has 20% of yellow coloring because it is connected to a M5S elected member with weight 4/20=0.2. Gray nodes are non elected users.

At this stage, we have analyzed each network by considering the total number of nodes and their pie chart of colors representing their affiliations. Figure 9 presents a pie chart that illustrates the percentage distribution of colors, considering each node with its strength, within the network for week 12, similar figures are obtained for other weeks. From this figure, it is evident that the majority of nodes are mostly connected to non-elected users. The number of interactions with elected members constitutes only a minority of the overall retweets. This observation suggests that the political discourse extends beyond politicians, involving a wider range of individuals. This fact also explains why, in Section 5, the community detection algorithm found only a few communities associated with a political coalition.

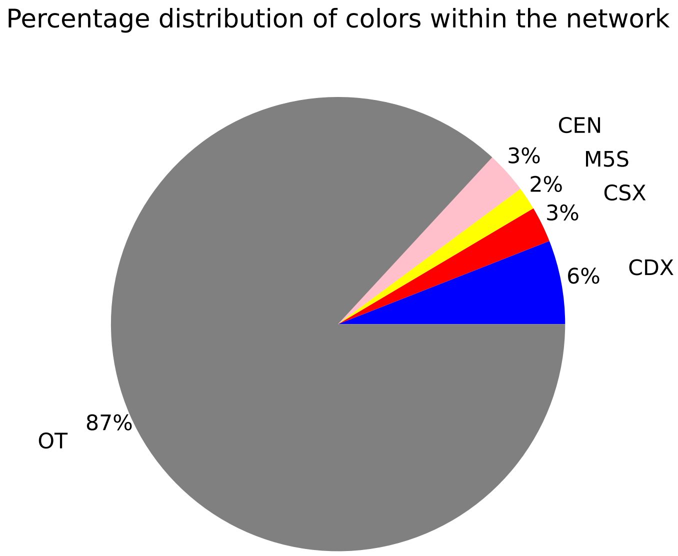


Figure 9: Percentage distribution of colored nodes for week 12. OT denotes connections with non elected users.

The same analysis is repeated for each network of the elected members. As expected, the results in this case are significantly different. Figure 10 visually illustrates, for week 12, that the majority of nodes are consistently connected to elected users, similar results hold for other weeks. In this scenario, the number of interactions with elected members constitutes the majority of the connections, making it easier to assign communities to a political coalitions. This indicates that the political discourse within these networks is primarily centered around elected politicians, leading to clearer affiliation and political community structures.

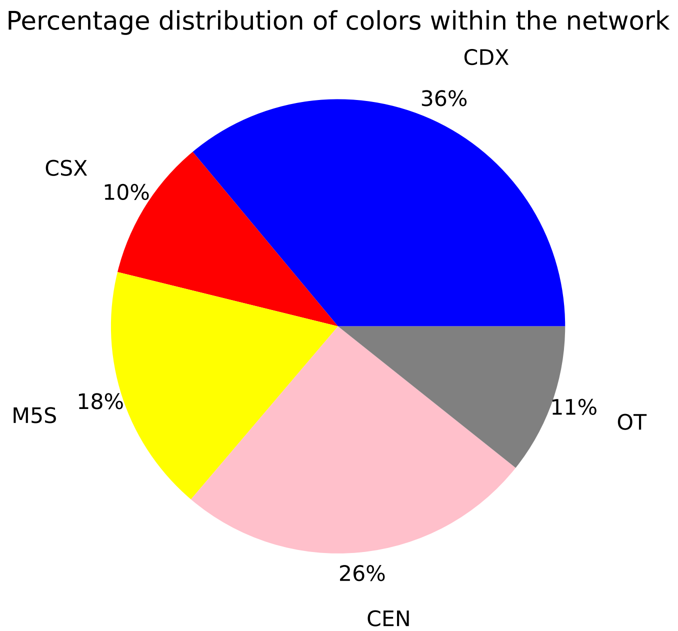


Figure 10: Percentage distribution of colored nodes for week 12 for the networks of the elected members. OT denotes connections with non elected users.

6. Role analysis

6.1. Centralities

For each network, we conducted a centrality analysis, specifically focusing on the strength s_i and eigenvector centrality γ_i of each node. In each network, we observed that the principal party leaders, such as *GiorgiaMeloni, GiuseppeConteIT, CarloCalenda, matteosalvinimi* and *EnricoLetta*, along with party pages, like *pdnetwork* and *FratellidItalia*, consistently exhibited higher strength. However, when considering eigenvector centrality, we found that mostly non-elected users achieved higher scores. Among these high-scoring accounts, we discovered other politicians who had not been elected, such as *elio_vito*, satirical accounts, like *ilruttosovrano*, and journalists, such as *lucianocapone*. This findings confirm that the political discourse extends beyond the sphere of only politicians, involving a broader range of individuals. Figure 11 shows the plot of the strength and eigenvector centrality for each node for week 12, confirming that elected members tend to have higher strength and lower eigenvector centrality, while non-elected members tend to have higher eigenvector centrality and smaller strength. This trend is consistent across all weeks, highlighting the different roles played within the network by these two categories of users.

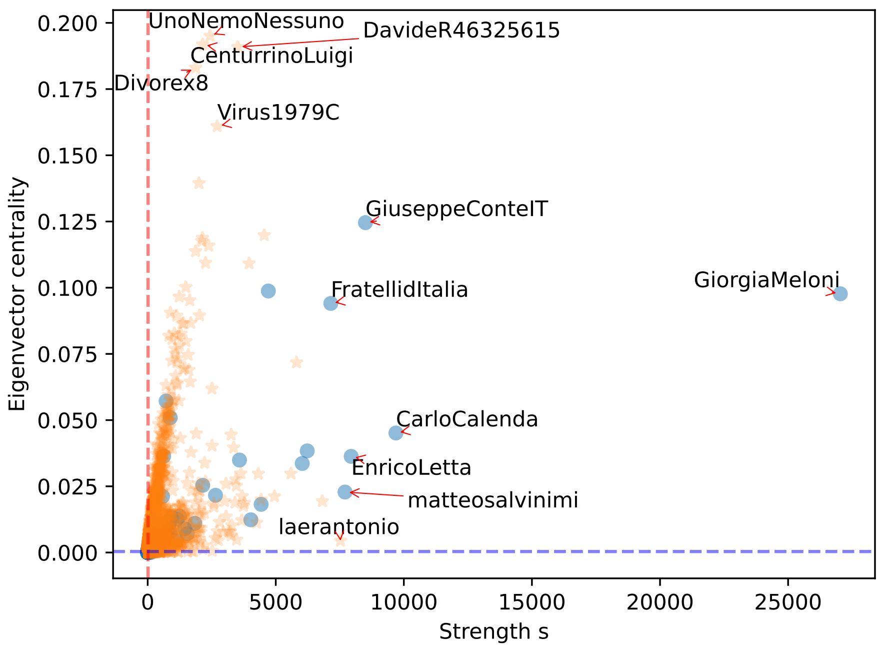


Figure 11: Plot of strength and eigenvector centrality for each node of week 12. In blue the elected member, in orange other users. The red and blue lines represent the average values, which are very close to zero.

When analyzing the networks of the elected members, we observed some differences in the results. Regarding strength, the accounts with higher values still belonged to principal political actors. In this case, one or two elected members also exhibits very high eigenvector centrality, above 0.5. Interestingly, here, the non-elected accounts having higher eigenvector centrality seemed to belong to particularly active users who were not directly associated with politics or journalism. This observations highlights that focusing exclusively on the network related to elected members results in the loss of numerous interactions, such as those involving journalists or other political actors. Figure 12 shows the plot of the strength and eigenvector centrality for each node for week 12. The central part of the plot is empty, the nodes of non-elected members are positioned along the y-axis, the elected members along the x-axis. Only two nodes, *Mov5Stelle* and *GiuseppeConteIT*, have both s_i and λ_i high. This trend is consistent across all weeks. For a more comprehensive analysis of these centralities, refer to Appendix B.

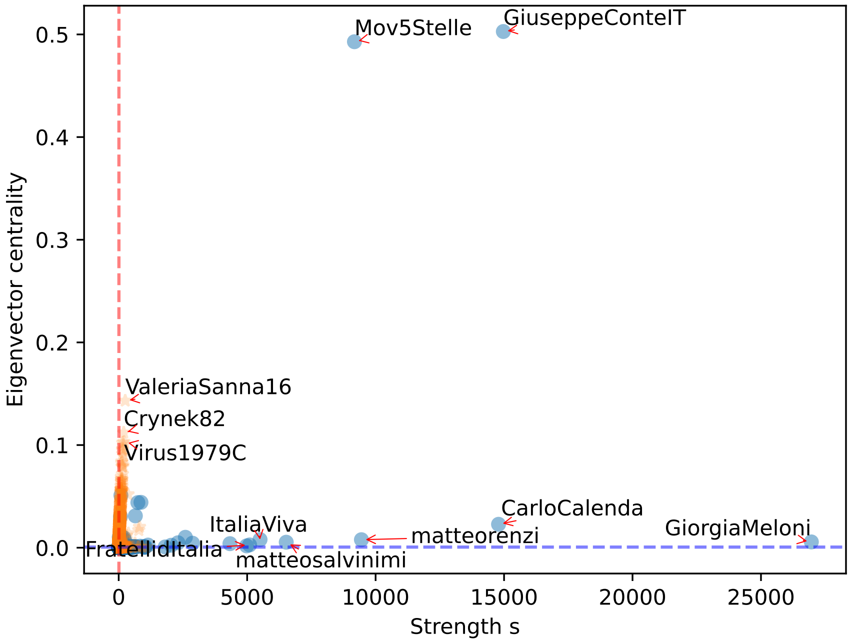


Figure 12: Plot of degree strength and eigenvector centrality for each node of week 12 of the networks of the elected members. In blue the elected members, in orange other users. The red and blue lines represent the average values, which are very close to zero.

6.2. z-P analysis

For each network, we conducted a role analysis for nodes using the z-P analysis method, [14] and [8], which is based on communities identified by Louvain Community Detection Algorithm. This analysis assigned a pair of indexes (z_i, P_i) to each node i, providing insights into the role of the node within the network. The two indexes, within-community strength z_i and participation coefficient P_i, were used to determine the significance of each node within its respective community.

To compute the within-community strength of a node i, we considered the community it belongs to, denoted as $c(i)$. The internal strength of node i, denoted as $s_i^{c(i)}$, is defined as the sum of the weights of the edges connecting node i to other nodes within the same community. In other words, it represents the strength of node i directed towards nodes within its own community. The within-community strength index, z_i, is then calculated as

$$z_i = \frac{s_i^{c(i)} - \overline{\mu}_{c(i)}}{\overline{\sigma}_{c(i)}}$$

where $\overline{\mu}_{c(i)}$ and $\overline{\sigma}_{c(i)}$ are the mean and standard deviation of $s_i^{c(i)}$ over all nodes $i \in c(i)$. Note that z_i can assume also negative values. The within-community strength z_i measures how strongly a node is connected within its own community.The participation coefficient P_i is defined as

$$P_i = 1 - \sum_{c=1}^{K} \left(\frac{s_i^c}{s_i} \right)^2$$

where $s_i^c = \sum_{j \in c} w_{ij}$ is the strength of node i directed towards nodes of community c, K is the total number of communities and s_i is the total strength of node i. The participation coefficient measures the extent to which a node is uniformly connected to all communities ($P_i \to 1$) rather than just its own community ($P_i \to 0$).

For each week we plotted the values (z_i, P_i) of each node on a z-P plane. Through this visualization, we observed that the majority of nodes had small values for both z_i and P_i. In contrast, only a few nodes had high values for either z_i or P_i, but not both simultaneously, this resulted in the central part of the plot being empty. Elected members had high z_i values, with small P_i values, indicating that they were strongly connected and played influential roles within their own communities but had limited connections to other communities.

Furthermore, a consistent number of nodes had small negative values of z_i suggesting that they had few connections within their community. Non elected members showed higher values of P_i indicating a uniform connection with different communities. To illustrates this analysis, Figure 13 shows an example of the z-P plot for week 12, highlighting the nodes with high z_i or P_i values.

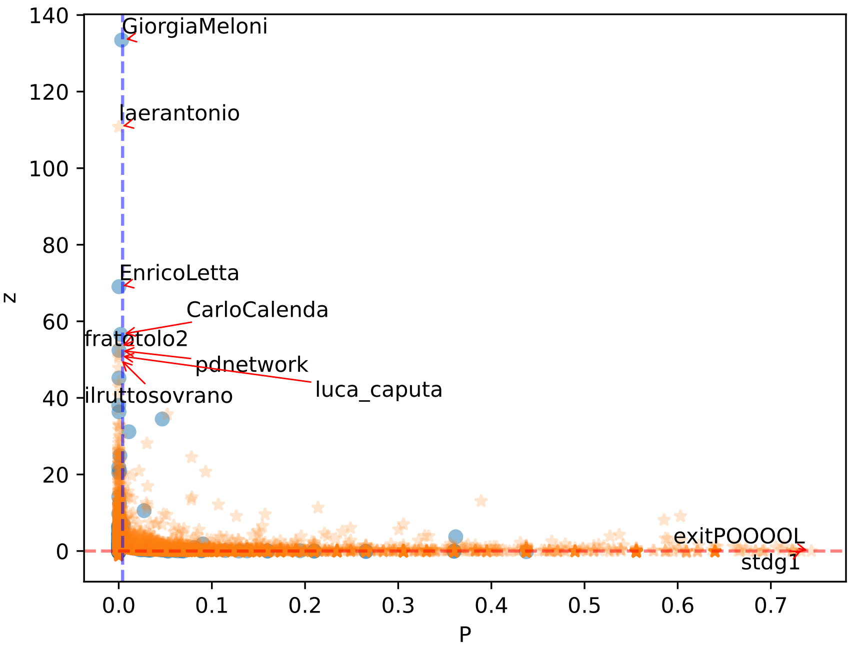


Figure 13: Plot of z_i and P_i for each node for week 12. In blue: elected members, in orange: other users. The red and blue lines represent the average values, which are very close to zero.

Considering that elected member exhibited both high values of strength centrality s_i and within-community strength z_i, we conducted a comparison between the results of the role analysis discussed earlier and the centrality analysis presented in Section 6.1. We observed a strong correlation between high strength centrality and high values of within-community strength z_i. This suggests that nodes with extensive connections within their communities tend to exhibit higher strength centrality. This relationship is depicted in Figure 14. However, this correspondence does not hold true for eigenvector centrality. There is no apparent alignment between these two measures.

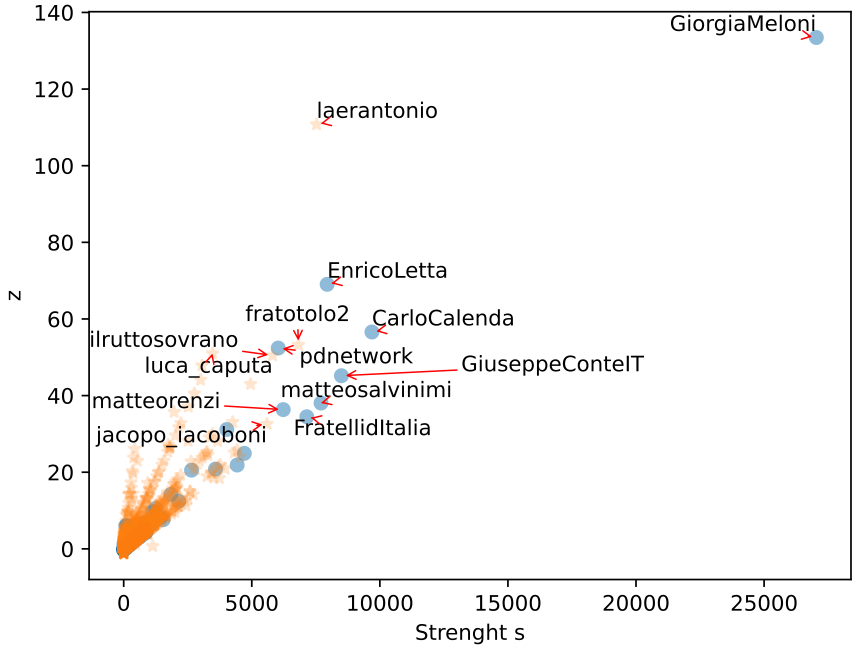
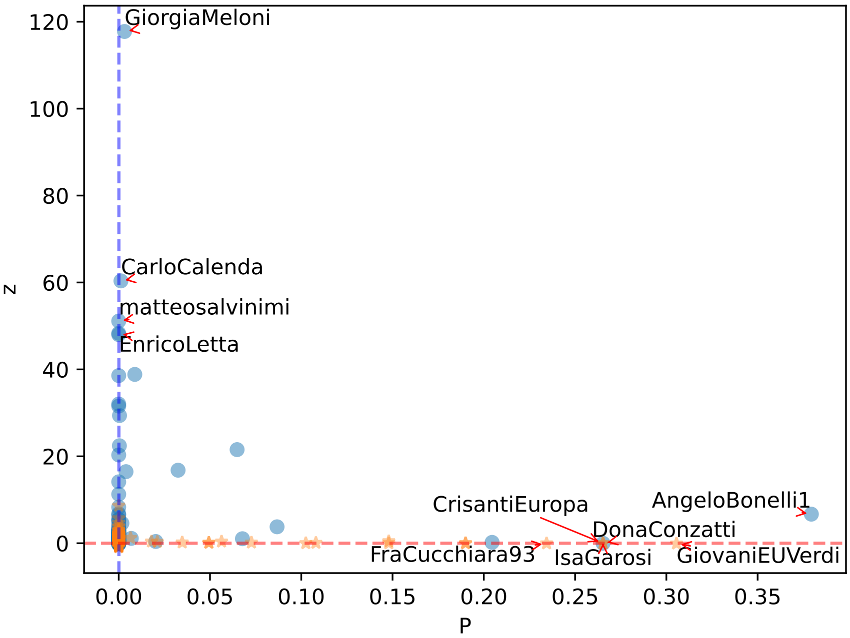


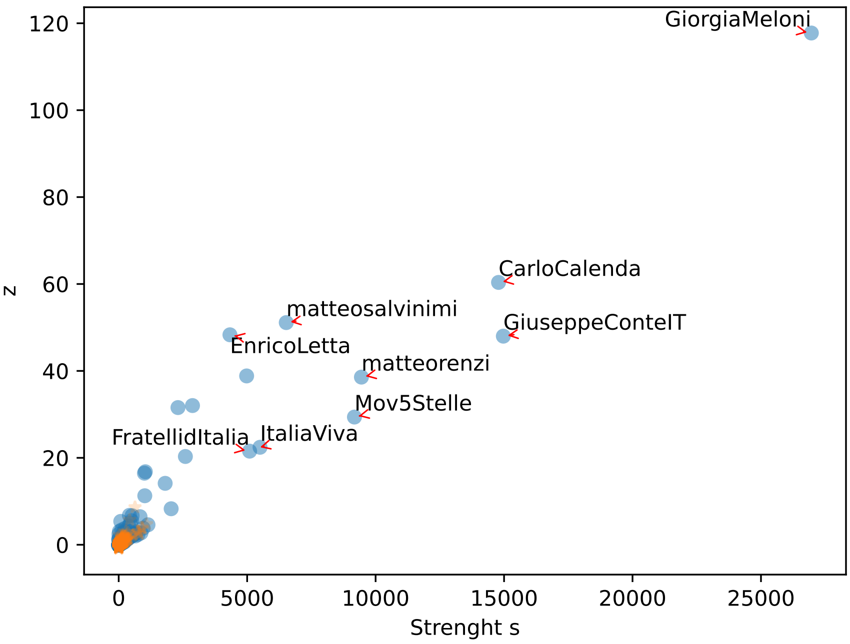
Figure 14: Plot of z_i and strength s_i for each node of week 12. In blue: elected members, in orange: other users.

We conducted the same role analysis for the networks of the elected members, the results were consistent with the previous analysis. Similar to before, nodes in these networks tend to cluster towards the edges of the $z - P$ plot with small values for both z_i and P_i. Additionally, similar to the previous analysis, the nodes in the networks with high z_i are elected members. Only few nodes, both elected and non-elected members, presented high values of P_i suggesting few connections between communities. Finally, the center of the plot is empty, highlighting the fact that each node plays only one role.

In these networks, the relationship between elected members and strong within-community connections is more pronounced. Elected members exhibit high values of z_i, indicating strong connections within their own communities. This underscores the role and influence of elected members within their respective communities. Figure 15a depicts the P-z plot of the network of the elected members in week 12 highlighting users with higher z_i values. Notably, all of these users are also among the top nodes ranked by strength centrality. This reinforces the association between high strength centrality and strong within-community connections among the elected members in these networks, as further demonstrated in Figure 15b.



(a) Plot of z_i and P_i



(b) Plot of z_i as a function of strength s_i

Figure 15: (a)(Top): Plot of z_i and P_i for each node relative to week 12 of the network of the elected members. In blue: elected members, in orange: other users. The red and blue lines represent the average values, which are very close to zero. (b)(Bottom): Plot of z_i and strength s_i for each node.

7. Time series analysis

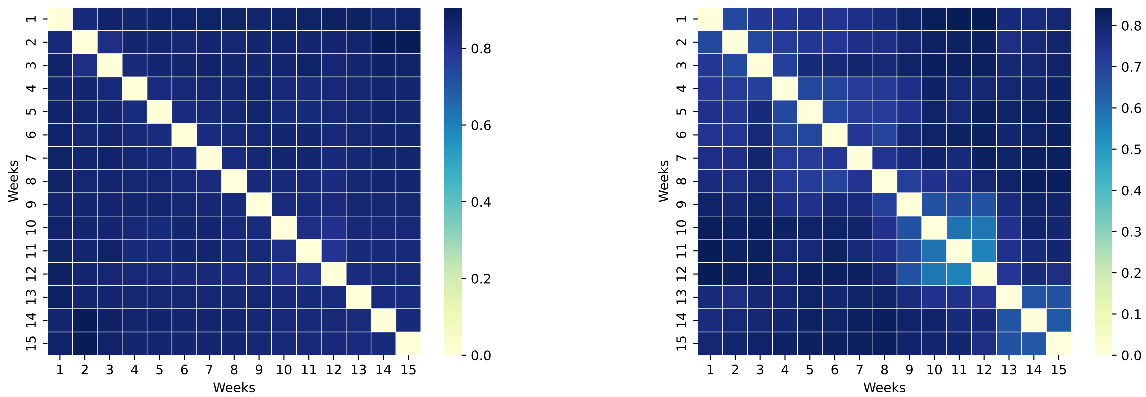
7.1. Distance between networks

With our time series analysis, we aimed to examine the temporal evolution of the networks from week 1 to week 15. Our primary objective was to identify any significant moment during the election period. We already observed a growing interest in political elections over the weeks, both in the total networks and the networks of the elected members, testified by an increasing number of involved users. Week 13 stood out as the week with the maximum number of nodes and edges, as shown in Figure 2 and 4.

To compare networks and identify important events, we computed the Weighted Jaccard distance between each couple of weeks. The Weighted Jaccard distance, considering the intersection of node sets V between graphs G_1 and G_2, is defined as $d_W(G_1, G_2) = 1 - J_W(A_1, A_2)$ where $A_1 = [a^1_{ij}]$, $A_2 = [a^2_{ij}]$ are the respective weighted adjacency matrices and

$$J_W(A_1, A_2) = \begin{cases} \frac{\sum_{i,j \in V} min(a^1_{ij}, a^2_{ij})}{\sum_{i,j \in V} max(a^1_{ij}, a^2_{ij})} & \text{if} \quad \sum_{i,j \in V} max(a^1_{ij}, a^2_{ij}) > 0 \\ 1 \quad \text{if} \quad \sum_{i,j \in V} max(a^1_{ij}, a^2_{ij}) = 0 \end{cases}$$

From this definition it is clear that $J_W \in [0, 1]$. Figure 16 illustrates the Weighted Jaccard distance over the entire time period for both the total networks and the networks of the elected members. Unfortunately, we observed that the distance remained uniform and consistently high, in both cases, with a slight decrease for weeks 10, 11 and 12 of the networks of the elected members. These high distance values suggest that there is substantial reorganization of connections between nodes in the networks. This means that there was no consistent pattern observed in the evolution of the networks over time. Therefore, based on this analysis, we were unable to identify any specific event within the network in the whole period of time.



(a)

(b)

Figure 16: Weighted Jaccard distance for both the networks of total retweets (a) and the networks of the elected members (b).

7.2. Flow of nodes

We created a Sankey Diagram to visualize the flow of nodes from one community to another, identified by Louvain Algorithm, during the weeks in the networks of the elected members. We focused on these networks because we were able to associate the majority of the identified communities with political coalitions, allowing us to study the transitions of nodes between communities. The diagram in Figure 17 revealed several interesting observations:

- the M5S coalition (in yellow) appears to be consistent in time and exhibits limited interactions with other communities;
- the CDX coalition (in blue) frequently split into several communities and merges within its communities. In week 12 and 13 the CDX communities exhibit a remarkably large number of nodes;
- the CSX coalition (in red) is often split into two communities, one bigger and one smaller, constantly associated to *Partito Democratico* and *Alleanza Verdi e Sinistra* respectively;

- the CEN coalition (in pink) primarily consists of one large community after week 5, excluding week 8, with some node transitions occurring with the CSX coalition;
- in week 5 we identified a community labeled 'OT', which was not associated with any political coalition. The Sankey Diagram reveals that this community is formed through the merger of a CSX and a CEN community in week 4. In week 6, some nodes from the OT community merged into another CEN community, while others split into two CSX communities. The weeks in question played a significant role in shaping the formation of CSX and CEN coalitions.

These observations highlight the dynamic nature of the networks and the fluidity of connections between different communities associated with political coalitions.

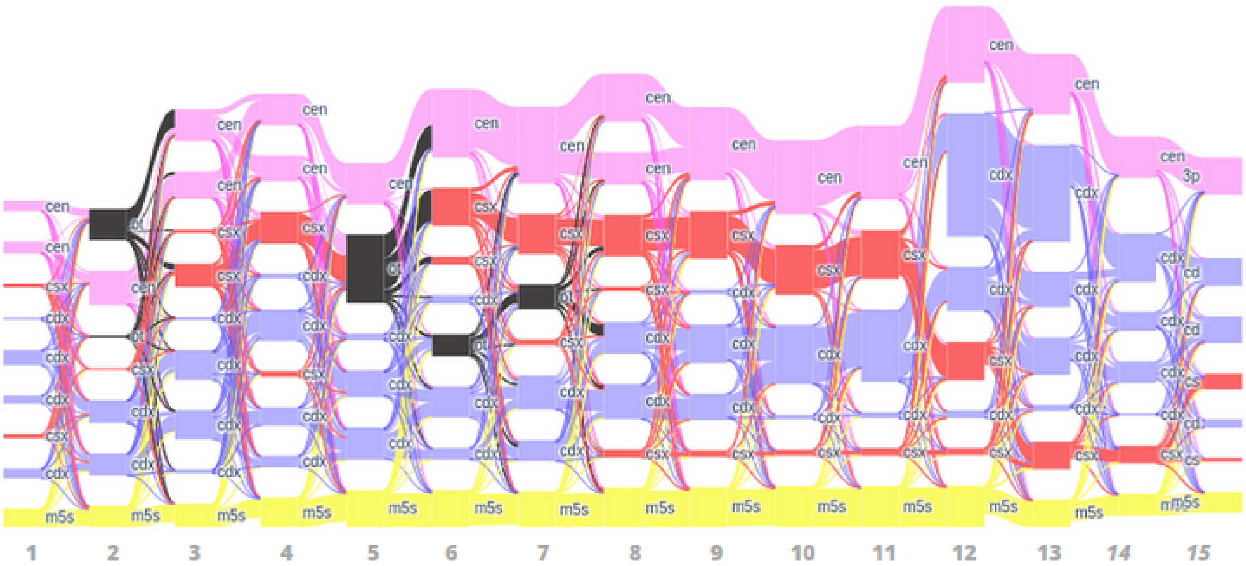


Figure 17: Sankey Diagram illustrating the flow of nodes from one community to another during the weeks in the networks of the elected members.

7.3. Political share

Finally, considering the overall networks, we calculated the share of each political coalition by aggregating the total weighted interactions received by each elected member within each political coalition. We compared these results with the elections outcomes. In Figure 18, the plot illustrates the evolving share over time. To mitigate fluctuations, we calculated the average share over two-week intervals, specifically from week 3 to week 12, which corresponds to the week of the elections.

The CDX coalition exhibits a slightly decreasing trend; its share is constantly below the political results.
CEN and M5S coalitions show relatively stable trends, with the latter slightly beyond its political outcomes. In contrast, the CEN coalition attains a much higher level of share compared to the election results.
The CSX coalition initially displays an increasing trend until weeks 7 and 8, after which it shows a decreasing trend. Throughout this period, the CSX coalition consistently maintains a share slightly above its political outcomes.

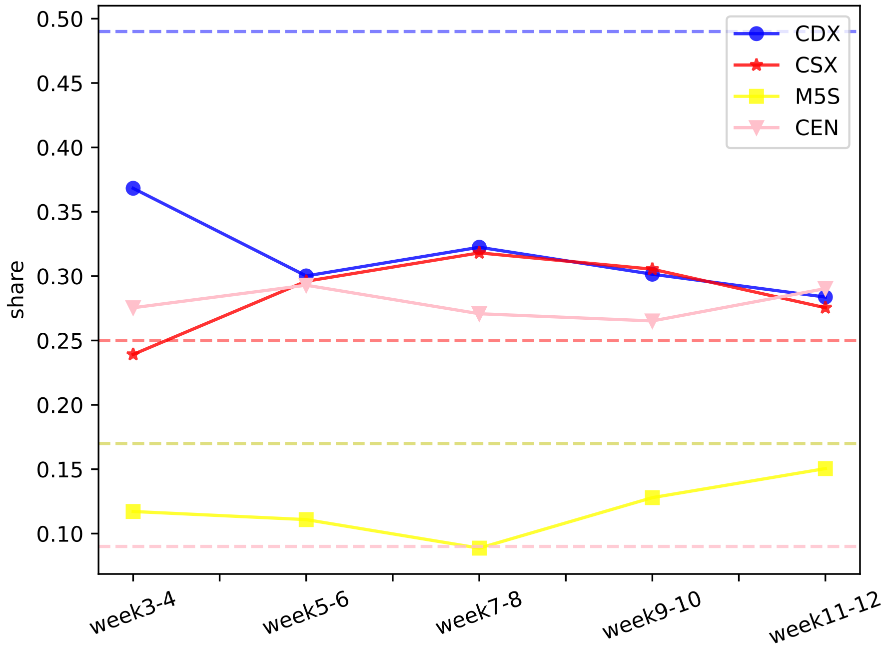
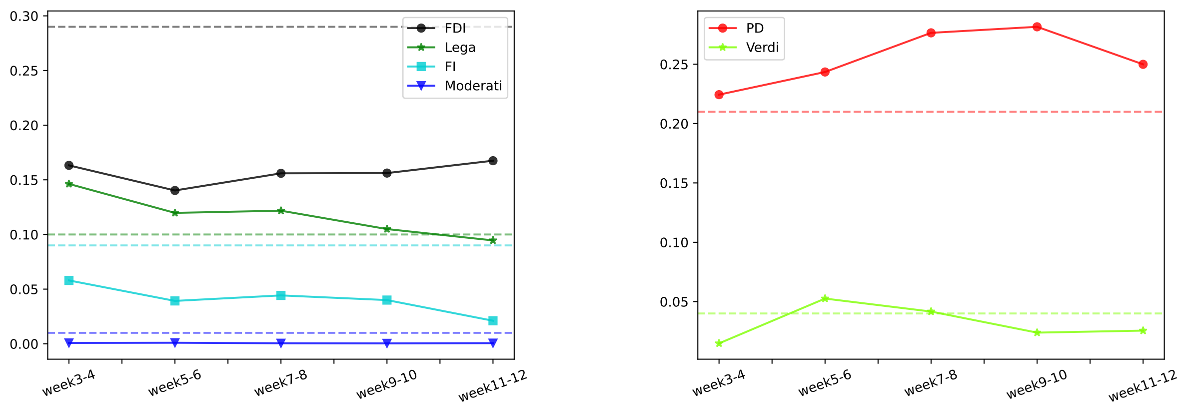


Figure 18: Plot of the share of each political coalition in time in the Twitter network. Dotted lines represent the political outcomes of each coalition, at the election of September 25th, 2022. The electoral results of the four coalitions, which collectively amounted to 90%, have been proportionally recalibrated to ensure a total sum of 100%.

Figures 19a and 19b illustrate the progressive evolution of the share held by the CDX and CSX coalitions, respectively, taking into account the different parties involved. Regarding the CDX coalition, there is an upward trend observed for the party *Fratelli d'Italia*, which is far below its actual political results in terms of total share. Conversely, both the *Lega* and *Forza Italia* experience a declining trend, with the former slightly exceeding its political outcome and the latter falling below it. The party *Noi Moderati* maintains a consistent trend and aligns with its political outcomes, with little to no impact on the total votes within its coalition.
As for the CSX coalition, the party *Partito Democratico* initially displays an increasing trend, followed by a subsequent decrease, while still maintaining a share above its political results. The party *Alleanza Verdi e Sinistra* exhibits a steady trend, aligning closely with its political outcomes.



(a) Share of CDX parties.

(b) Share of CSX parties.

Figure 19: Share of CDX coalition (a) divided into its four parties: *Fratelli d'Italia, Lega, Forza Italia* and *Noi Moderati.* Share do CSX coalition (b) divided into its two parties: *Partito Democratico* and *Alleanza Verdi e Sinistra.* The dottet lines

8. Conclusion

This thesis presents the findings of our network analysis conducted on a Twitter dataset, specifically examining the Italian political elections that occurred on September 25th, 2022. This dataset was generated using a snowball sampling approach, based on a list of pertinent keywords. Our analysis primarily focused on two groups of data: the overall retweet activity and the retweets involving elected members. By aggregating the data on a weekly basis, we obtained a total of 30 networks, 15 for each data group, covering the entire duration of the electoral campaign and one month following Election Day.

One of the primary objectives of our study was to analyze the partition of each network, using the Louvain Community Detection Algorithm, and to establish associations between the resulting communities and political coalitions. We focused on four major political coalitions: CDX, CSX, M5S, and CEN. Our findings revealed that for the network of total retweets, we did not observe any significant associations between the communities and political coalitions. However, while studying the networks involving elected members, we were able to successfully associate nearly every community with a specific political coalition.

Through our role analysis, we investigated the positions held by individual nodes within the network. Notably, we observed that elected members consistently occupied central roles in terms of strength centrality. On the other hand, when considering eigenvector centrality, we identified other nodes, including other politicians, satirical pages and journalists, assuming central positions. This finding indicates that the discourse surrounding the elections extends beyond elected members and encompasses various actors. Furthermore, our z-P analysis revealed that elected members exhibited high within-community strength, indicating strong connections within their respective communities.

In addition, we conducted a time series analysis to examine the distances between weeks in the networks, using the Weighted Jaccard distance. Our findings showed consistently high distances, indicating significant reorganization of connections between nodes in the networks from week to week. To provide further insights, we visualized a Sankey Diagram that illustrated the flow of nodes from one community to another throughout the entire time period under consideration. This diagram allowed us to identify the stability of certain political coalitions, such as M5S, as well as the division within other coalitions, like CDX. We were also able to detect specific political events, such as the separation of CSX and CEN coalitions in week 5. We lastly calculated the overall share obtained by each political coalition and compared it to their actual political outcomes. This analysis revealed that the CEN coalition had a significantly higher overall share compared to its electoral result, whereas the CDX coalition had a lower share than its result, with a decreasing trend over time.

The research offers valuable insights into network analysis using Twitter data in the context of political topics. Following these findings, several potential directions for further research can be identified. Firstly, alternative community detection algorithms could be employed to determine if the obtained results differ from those presented in this work. For instance, the use of a Label Propagation Algorithm [12] might be considered as an alternative approach. This would allow for a comparative analysis of different algorithms and their impact on community detection within political networks on Twitter.

Another avenue for further research is to incorporate a broader range of defined users into the analysis. As shown, non-political users can have a significant impact on the political discourse surrounding elections: by including a wider variety of users, such as journalists and citizens, a more comprehensive understanding of the internal dynamics within each network could be achieved. This expanded analysis might lead to a better understanding of the interactions between political and non-political actors, while also providing insights on how information flows influences the public opinion.

Finally, conducting a demographic study of Italian users on Twitter would be a valuable addition to this analysis. This study could provide information on the age, gender, location, occupation, and other relevant attributes of Twitter users who engage in political discussions and share content related to the elections. By analyzing these demographics in relation to the political affiliations and voting patterns, it would be possible to gain insights into the factors that contribute to the differences in the amount of share obtained by each political coalition and their respective political outcomes. This combined analysis of network dynamics and user demographics could contribute to a more comprehensive understanding of the role of Twitter in shaping political discourse and influencing electoral results.

References

[1] AL Barabási. Network Science. Cambridge University Press, Cambridge, 2016.

[2] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: an open source software for exploring and manipulating networks. In *Proceedings of the international AAAI conference on web and social media*, volume 3, pages 361–362, 2009.

[3] Vladimir Batagelj and Matjaž Zaveršnik. Generalized cores. *arXiv preprint cs/0202039*, 2002.

[4] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[5] Alexandre Bovet and Hernán A Makse. Influence of fake news in Twitter during the 2016 US presidential election. *Nature communications*, 10(1):7, 2019.

[6] Alexandre Bovet, Flaviano Morone, and Hernán A Makse. Validation of Twitter opinion trends with national polling aggregates: Hillary Clinton vs Donald Trump. *Scientific reports*, 8(1):8673, 2018.

[7] Guido Caldarelli, Alessandro Chessa, Fabio Pammolli, Gabriele Pompa, Michelangelo Puliga, Massimo Riccaboni, and Gianni Riotta. A multi-level geographical study of Italian political elections from Twitter data. *PloS one*, 9(5):e95809, 2014.

[8] Francesco Calderoni, Domenico Brunetto, and Carlo Piccardi. Communities in criminal networks: A case study. *Social Networks*, 48:116–125, 2017.

[9] Elanor Colleoni, Alessandro Rozza, and Adam Arvidsson. Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of communication*, 64(2):317–332, 2014.

[10] Michael Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. Political polarization on Twitter. In *Proceedings of the international aaai conference on web and social media*, volume 5, pages 89–96, 2011.

[11] Michael D Conover, Bruno Gonçalves, Jacob Ratkiewicz, Alessandro Flammini, and Filippo Menczer. Predicting the political alignment of Twitter users. In *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*, pages 192–199. IEEE, 2011.

[12] Gennaro Cordasco and Luisa Gargano. Community detection via semi-synchronous label propagation algorithms. In *2010 IEEE international workshop on: business applications of social network analysis (BASNA)*, pages 1–8. IEEE, 2010.

[13] Francesco Durazzi, Martin Müller, Marcel Salathé, and Daniel Remondini. Clusters of science and health related Twitter users become more isolated during the COVID-19 pandemic. *Scientific reports*, 11(1):19655, 2021.

[14] Roger Guimera and Luís A Nunes Amaral. Cartography of complex networks: modules and universal roles. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(02):P02001, 2005.

[15] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring Network Structure, Dynamics, and Function using NetworkX. In Gaël Varoquaux, Travis Vaught, and Jarrod Millman, editors, *Proceedings of the 7th Python in Science Conference*, pages 11 – 15, Pasadena, CA USA, 2008.

[16] M. E. J. Newman. *Networks: an introduction.* Oxford University Press, Oxford; New York, 2010.

[17] Diogo Pacheco, Pik-Mai Hui, Christopher Torres-Lugo, Bao Tran Truong, Alessandro Flammini, and Filippo Menczer. Uncovering coordinated networks on social media. *arXiv preprint arXiv:2001.05658*, 16, 2020.

[18] Carlo Piccardi. Finding and testing network communities by lumped Markov chains. *PloS one*, 6(11):e27028, 2011.

[19] Francesco Pierri, Alessandro Artoni, and Stefano Ceri. Investigating Italian disinformation spreading on Twitter in the context of 2019 European elections. *PloS one*, 15(1):e0227821, 2020.

[20] Francesco Pierri, Geng Liu, and Stefano Ceri. ITA-ELECTION-2022: A multi-platform dataset of social media conversations around the 2022 Italian general election. *arXiv preprint arXiv:2301.05119*, 2023.

[21] Tommaso Radicioni, Fabio Saracco, Elena Pavan, and Tiziano Squartini. Analysing Twitter semantic networks: the case of 2018 Italian elections. *Scientific Reports*, 11(1):13207, 2021.

[22] Erzsébet Ravasz and Albert-László Barabási. Hierarchical organization in complex networks. *Physical review E*, 67(2):026112, 2003.

[23] Luca Rossi, Nicola Righetti, and Giada Marino. (Nearly) Ten Years of Social Media and Political Elections in Italy: Questions, Platforms, and Methods. *Social Media+ Society*, 7(4):20563051211063460, 2021.

[24] Mattia Tantardini, Francesca Ieva, Lucia Tajoli, and Carlo Piccardi. Comparing methods for comparing networks. *Scientific reports*, 9(1):1–19, 2019.

[25] Salvatore Vilella, Alfonso Semeraro, Daniela Paolotti, and Giancarlo Ruffo. Measuring user engagement with low credibility media sources in a controversial online debate. *EPJ data science*, 11(1):29, 2022.

[26] Matthew Woodward. Twitter User Statistics: How Many People Use Twitter In 2023?, 2023. `https://www.searchlogistics.com/learn/statistics/twitter-user-statistics/`, Accessed 14 Jun 2023.

A. Appendix A

In the following we report the number of nodes N_C and the probability of persistence u_{CC} for each community, detected by Louvain Community Detection Algorithm, in the networks of the elected members.

Week 1

	N_C	u_{CC}
M5S	1694	0.94
CDX	1432	0.83
CEN	965	0.84
CDX	906	0.75
CDX	820	0.79
CSX	741	0.86
CEN	730	0.91
CDX	446	0.87
CSX	309	0.8

Table 3: Number of node and probability of persistence of each community in week 1.

Week 2

	N_C	u_{CC}
CEN	2786	0.89
OT	2369	0.83
CDX	1915	0.87
M5S	1771	0.96
CDX	1640	0.87
OT	215	0.63
CSX	143	0.74

Table 4: Number of node and probability of persistence of each community in week 2.

Week 3

	N_C	u_{CC}
CEN	4401	0.8
CDX	3096	0.81
CSX	2718	0.72
CEN	2493	0.79
CDX	2366	0.83
M5S	1744	0.88
CSX	798	0.58
CDX	738	0.7
CDX	250	0.55

Table 5: Number of node and probability of persistence of each community in week 3.

Week 4

	N_C	u_{CC}
CDX	3611	0.82
CSX	3384	0.82
CEN	2371	0.73
CEN	2152	0.83
M5S	1958	0.93
CDX	1454	0.78
CDX	1066	0.68
CSX	656	0.81
CDX	527	0.65

Table 6: Number of node and probability of persistence of each community in week 4.

Week 5

	N_C	u_{CC}
OT	6485	0.87
CEN	3188	0.87
CDX	2992	0.84
M5S	2972	0.88
CDX	2679	0.72
CDX	551	0.58

Table 7: Number of node and probability of persistence of each community in week 5.

Week 6

	N_C	u_{CC}
CEN	5321	0.87
CSX	4087	0.78
CDX	3183	0.76
OT	2874	0.69
M5S	2724	0.93
CSX	984	0.69
CDX	872	0.82
CDX	825	0.54

Table 8: Number of node and probability of persistence of each community in week 6.

Week 7

	N_C	u_{CC}
CEN	5937	0.85
CSX	3825	0.69
CDX	3130	0.77
OT	3008	0.65
M5S	2581	0.87
CDX	1786	0.66
CSX	492	0.68

Table 9: Number of node and probability of persistence of each community in week 7.

Week 8

	N_C	u_{CC}
CSX	4952	0.75
CEN	3809	0.76
CDX	3456	0.69
M5S	3274	0.92
CDX	3165	0.85
CEN	2963	0.62
CSX	984	0.76
CSX	767	0.53

Table 10: Number of node and probability of persistence of each community in week 8.

Week 9

	N_C	u_{CC}
CEN	5535	0.92
CSX	4672	0.82
CDX	4014	0.79
M5S	2845	0.96
CDX	2473	0.83
CDX	719	0.69
CSX	600	0.75

Table 11: Number of node and probability of persistence of each community in week 9.

Week 10

	N_C	u_{CC}
CSX	5553	0.84
CEN	5446	0.94
CDX	5148	0.95
M5S	3073	0.97
CSX	567	0.85
CSX	549	0.77

Table 12: Number of node and probability of persistence of each community in week 10.

Week 11

	N_C	u_{CC}
CEN	6326	0.93
CDX	5928	0.95
CSX	5588	0.83
M5S	3082	0.97
CDX	834	0.8
CSX	785	0.85

Table 13: Number of node and probability of persistence of each community in week 11.

Week 12

	N_C	u_{CC}
CDX	14402	0.91
CEN	5876	0.96
CDX	4861	0.82
CSX	3657	0.87
M5S	3217	0.98
CDX	1141	0.72
CSX	705	0.87

Table 14: Number of node and probability of persistence of each community in week 12.

Week 13

	N_C	u_{CC}
CDX	32246	0.91
CEN	5121	0.96
CDX	4352	0.7
CSX	4235	0.92
CDX	3278	0.78
M5S	1905	0.97
CDX	944	0.67

Table 15: Number of node and probability of persistence of each community in week 13.

Week 14

	N_C	u_{CC}
CDX	4098	0.85
CEN	3085	0.96
M5S	1808	0.97
CSX	1682	0.84
CDX	1580	0.85
CDX	1339	0.73
CDX	168	0.7

Table 16: Number of node and probability of persistence of each community in week 14.

Week 15

	N_C	u_{CC}
CDX	4308	0.8
CEN	4062	0.93
CSX	3485	0.83
CDX	2882	0.79
M5S	1649	0.94
CDX	898	0.69
CSX	428	0.72

Table 17: Number of node and probability of persistence of each community in week 15.

B. Appendix B

In the following we report the 5 users with higher strength and eigenvector centrality and respective values for both the total networks and the networks of the elected members.

Week 1

	ID	s_i	ID	γ_i
1°	*MarcoRizzoPC*	6543	*DavideR46325615*	0.36
2°	*DavideR46325615*	4507	*serebellardinel*	0.18
3°	matteosalvinimi	4117	*erretti42*	0.16
4°	GiorgiaMeloni	4021	*InfinitoIsacco*	0.15
5°	GiuseppeConteIT	3297	*dukana2*	0.15
1°	matteosalvinimi	3227	matteosalvinimi	0.63
2°	GiorgiaMeloni	2581	LegaSalvini	0.28
3°	GiuseppeConteIT	1786	GiorgiaMeloni	0.16
4°	CarloCalenda	1748	Noiconsalvini	0.15
5°	Mov5Stelle	1656	borghi_claudio	0.14

Table 18: Top: the 5 users with higher strength and eigenvector centrality for total network in week 1. Bottom: for the network of the elected members.

Week 2

	ID	s_i	ID	γ_i
1°	GiuseppeConteIT	5573	*dariodangelo91*	0.19
2°	*lefrasidiosho*	5209	*lucianocapone*	0.13
3°	*lucianocapone*	4791	*DavideR46325615*	0.13
4°	CarloCalenda	4147	*HSkelsen*	0.12
5°	*DavideR46325615*	4085	*Paroledipaola*	0.11
1°	CarloCalenda	3448	borghi_claudio	0.65
2°	GiorgiaMeloni	2913	LegaSalvini	0.18
3°	borghi_claudio	2822	*marcoranieri72*	0.17
4°	matteorenzi	2229	GiorgiaMeloni	0.17
5°	ItaliaViva	1952	*Marko_Morandi*	0.16

Table 19: Top: the 5 users with higher strength and eigenvector centrality for total network in week 2. Bottom: for the network of the elected members.

Week 3

	ID	s_i	ID	γ_i
1°	CarloCalenda	10475	*dariodangelo91*	0.3
2°	*jacopo_iacoboni*	9589	*jacopo_iacoboni*	0.24
3°	*LiveSpinoza*	8931	*HSkelsen*	0.17
4°	GiorgiaMeloni	7965	*danieledv79*	0.14
5°	EnricoLetta	7601	*PaoloBorg*	0.13
1°	CarloCalenda	10478	borghi_claudio	0.64
2°	borghi_claudio	5074	LegaSalvini	0.21
3°	GiorgiaMeloni	4578	CarloCalenda	0.18
4°	matteorenzi	2907	*ST09972061*	0.14
5°	matteosalvinimi	2814	*Marko_Morandi*	0.13

Table 20: Top: the 5 users with higher strength and eigenvector centrality, with values, for total network in week 3. Bottom: for the network of the elected members.

Week 4

	ID	s_i	ID	γ_i
1°	GiorgiaMeloni	11472	*erretti42*	0.13
2°	EnricoLetta	10104	*LuciaLaVita1*	0.13
3°	pdnetwork	8830	*DavideR46325615*	0.13
4°	matteosalvinimi	8002	*Moonlightshad1*	0.12
5°	*elio_vito*	7323	*elio_vito*	0.12
1°	matteosalvinimi	5789	matteosalvinimi	0.56
2°	GiorgiaMeloni	5656	LegaSalvini	0.33
3°	borghi_claudio	5566	borghi_claudio	0.29
4°	CarloCalenda	3883	*Lega_Massa*	0.17
5°	marattin	2751	Noiconsalvini	0.16

Table 21: Top: the 5 users with higher strength and eigenvector centrality for total network in week 4. Bottom: for the network of the elected members.

Week 5

	ID	s_i	ID	γ_i
1°	CarloCalenda	12017	*AlvisiConci*	0.13
2°	EnricoLetta	10701	*DavideR46325615*	0.13
3°	pdnetwork	10105	*danieledv79*	0.12
4°	GiorgiaMeloni	9578	*_marlene1265*	0.1
5°	matteosalvinimi	6986	*GarauSilvana*	0.1
1°	CarloCalenda	6046	matteosalvinimi	0.67
2°	matteosalvinimi	5849	LegaSalvini	0.24
3°	GiorgiaMeloni	5562	Noiconsalvini	0.21
4°	GiuseppeConteIT	5559	*mogicrz*	0.17
5°	marattin	3872	*MarioSavioDe*	0.16

Table 22: Top: the 5 users with higher strength and eigenvector centrality for total network in week 5. Bottom: for the network of the elected members.

Week 6

	ID	s_i	ID	γ_i
1°	CarloCalenda	11663	*DavideR46325615*	0.18
2°	GiorgiaMeloni	11330	*CenturrinoLuigi*	0.14
3°	pdnetwork	10487	*GarauSilvana*	0.14
4°	EnricoLetta	9090	*serebellardinel*	0.13
5°	matteosalvinimi	8009	*ilruttosovrano*	0.12
1°	CarloCalenda	7759	matteosalvinimi	0.63
2°	GiorgiaMeloni	7050	LegaSalvini	0.3
3°	matteosalvinimi	5358	Noiconsalvini	0.2
4°	pdnetwork	3671	*SabryStefano*	0.17
5°	GiuseppeConteIT	3593	borghi_claudio	0.16

Table 23: Top: the 5 users with higher strength and eigenvector centrality for total network in week 6. Bottom: for the network of the elected members.

Week 7

	ID	s_i	*ID*	γ_i
1°	GiorgiaMeloni	12963	*GarauSilvana*	0.23
2°	pdnetwork	12839	*DavideR46325615*	0.17
3°	EnricoLetta	10957	*serebellardinel*	0.15
4°	CarloCalenda	10806	*francotaratufo2*	0.13
5°	matteosalvinimi	9320	*gioart79*	0.12
1°	CarloCalenda	8400	matteosalvinimi	0.61
2°	GiorgiaMeloni	7987	LegaSalvini	0.41
3°	matteosalvinimi	7507	Noiconsalvini	0.21
4°	GiuseppeConteIT	5215	*AgrilloAlex*	0.18
5°	EnricoLetta	4948	eugenio_zoffili	0.16

Table 24: Top: the 5 users with higher strength and eigenvector centrality for total network in week 7. Bottom: for the network of the elected members.

Week 8

	ID	s_i	*ID*	γ_i
1°	GiorgiaMeloni	16340	*DavideR46325615*	0.21
2°	CarloCalenda	13265	*Lorellastelle*	0.16
3°	EnricoLetta	12923	*GarauSilvana*	0.16
4°	pdnetwork	9665	*laura_maffi*	0.12
5°	GiuseppeConteIT	7470	*erretti42*	0.12
1°	CarloCalenda	11997	matteosalvinimi	0.5
2°	GiorgiaMeloni	7660	LegaSalvini	0.47
3°	borghi_claudio	6672	*SabryStefano*	0.29
4°	matteosalvinimi	5782	*Lega_Massa*	0.28
5°	GiuseppeConteIT	5537	Noiconsalvini	0.27

Table 25: Top: the 5 users with higher strength and eigenvector centrality for total network in week 8. Bottom: for the network of the elected members.

Week 9

	ID	s_i	ID	γ_i
1°	CarloCalenda	11937	Mov5Stelle	0.29
2°	GiorgiaMeloni	11889	GiuseppeConteIT	0.24
3°	EnricoLetta	10442	*DavideR46325615*	0.15
4°	pdnetwork	9379	CarloCalenda	0.15
5°	GiuseppeConteIT	8400	*CenturrinoLuigi*	0.15
1°	CarloCalenda	13662	Mov5Stelle	0.57
2°	GiuseppeConteIT	10151	GiuseppeConteIT	0.38
3°	Mov5Stelle	7543	CarloCalenda	0.15
4°	GiorgiaMeloni	6707	*Crynek82*	0.15
5°	matteosalvinimi	6036	*MassimoChiaram7*	0.14

Table 26: Top: the 5 users with higher strength and eigenvector centrality for total network in week 9. Bottom: for the network of the elected members.

Week 10

	ID	s_i	ID	γ_i
1°	EnricoLetta	10946	*puntualeh*	0.26
2°	CarloCalenda	10602	*laura_maffi*	0.22
3°	GiorgiaMeloni	9833	*DavideR46325615*	0.21
4°	GiuseppeConteIT	7321	*CenturrinoLuigi*	0.19
5°	pdnetwork	7276	*Giancar70336148*	0.16
1°	CarloCalenda	18746	CarloCalenda	0.63
2°	GiuseppeConteIT	13196	ItaliaViva	0.23
3°	Mov5Stelle	9296	*GabboAntoninoN1*	0.14
4°	ItaliaViva	5763	*Mirarch3*	0.13
5°	GiorgiaMeloni	5551	*NuvolettaZen*	0.12

Table 27: Top: the 5 users with higher strength and eigenvector centrality for total network in week 10. Bottom: for the network of the elected members.

Week 11

	ID	s_i	ID	γ_i
1°	CarloCalenda	12355	*DavideR46325615*	0.19
2°	GiorgiaMeloni	11031	*laura_maffi*	0.17
3°	EnricoLetta	10703	*CenturrinoLuigi*	0.15
4°	GiuseppeConteIT	9223	*Moonlightshad1*	0.14
5°	*jacopo_iacoboni*	6800	*GarauSilvana*	0.13
1°	CarloCalenda	18612	Mov5Stelle	0.53
2°	GiuseppeConteIT	15225	GiuseppeConteIT	0.46
3°	Mov5Stelle	9921	*ValeriaSanna16*	0.15
4°	GiorgiaMeloni	7154	*Divorex8*	0.15
5°	ItaliaViva	6255	*Crynek82*	0.13

Table 28: Top: the 5 users with higher strength and eigenvector centrality for total network in week 11. Bottom: for the network of the elected members.

Week 12

	ID	s_i	ID	γ_i
1°	GiorgiaMeloni	27040	*UnoNemoNessuno*	0.2
2°	CarloCalenda	9691	*CenturrinoLuigi*	0.19
3°	GiuseppeConteIT	8500	*DavideR46325615*	0.19
4°	EnricoLetta	7943	*Divorex8*	0.18
5°	matteosalvinimi	7703	*Virus1979C*	0.16
1°	GiorgiaMeloni	26961	GiuseppeConteIT	0.5
2°	GiuseppeConteIT	14972	Mov5Stelle	0.49
3°	CarloCalenda	14785	*ValeriaSanna16*	0.14
4°	matteorenzi	9448	*Crynek82*	0.11
5°	Mov5Stelle	9176	*Virus1979C*	0.1

Table 29: Top: the 5 users with higher strength and eigenvector centrality for total network in week 12. Bottom: for the network of the elected members.

Week 13

	ID	s_i	ID	γ_i
1°	GiorgiaMeloni	63241	GiorgiaMeloni	0.25
2°	FratellidItalia	12157	*Divorex8*	0.16
3°	CarloCalenda	9755	*Virus1979C*	0.15
4°	*_aquiloni*	6315	FratellidItalia	0.15
5°	matteosalvinimi	6302	*Moonlightshad1*	0.14
1°	GiorgiaMeloni	54044	GiorgiaMeloni	0.57
2°	CarloCalenda	11580	FratellidItalia	0.52
3°	FratellidItalia	10816	*julio_martinezp*	0.08
4°	matteosalvinimi	4437	*AgrilloAlex*	0.07
5°	borghi_claudio	4122	*m88660092*	0.07

Table 30: Top: the 5 users with higher strength and eigenvector centrality for total network in week 13. Bottom: for the network of the elected members.

Week 14

	ID	s_i	ID	γ_i
1°	GiorgiaMeloni	12048	*DavideR46325615*	0.32
2°	CarloCalenda	5184	*dukana2*	0.29
3°	*ilruttosovrano*	4234	*InfinitoIsacco*	0.26
4°	matteosalvinimi	3746	*Divorex8*	0.14
5°	GiuseppeConteIT	3719	*Virus1979C*	0.13
1°	GiorgiaMeloni	7037	CarloCalenda	0.69
2°	CarloCalenda	5304	*Mirarch3*	0.16
3°	FratellidItalia	2765	*FItaliasulserio*	0.16
4°	matteosalvinimi	2687	ItaliaViva	0.16
5°	Mov5Stelle	2431	*mauriziosantin*	0.12

Table 31: Top: the 5 users with higher strength and eigenvector centrality for total network in week 14. Bottom: for the network of the elected members.

Week 15

	ID	s_i	ID	γ_i
1°	GiorgiaMeloni	15433	*ilCoperchio*	0.996
2°	CarloCalenda	7123	*leopadanofe*	0.03
3°	EnricoLetta	5598	*antokindness*	0.03
4°	FratellidItalia	4919	*EmyRoyaleagle*	0.02
5°	*ultimora_pol*	4813	*adrianobusolin*	0.02
1°	GiorgiaMeloni	9596	CarloCalenda	0.67
2°	CarloCalenda	7765	*Profilo3Marco*	0.13
3°	FratellidItalia	4363	*GuadagnoRaffae2*	0.12
4°	GiuseppeConteIT	3832	*FItaliasulserio*	0.11
5°	matteosalvinimi	2531	*Mirarch3*	0.11

Table 32: Top: the 5 users with higher strength and eigenvector centrality for total network in week 15. Bottom: for the network of the elected members.

Abstract in lingua italiana

In questo studio, condotto con le metodologie per lo studio delle reti complesse, analizziamo un dataset di Twitter incentrato sulle elezioni politiche italiane del 2022. Il dataset è aggregato su base settimanale, coprendo un periodo di 15 settimane; l'indagine si focalizza sia sulle reti di retweet che coinvolgono tutti gli utenti sia sulle reti dei soli retweet associati ai membri eletti. Nel modello a rete, i nodi corrispondono agli utenti di Twitter, collegati da link non diretti con pesi corrispondenti al numero di interazioni tra gli utenti, come retweet o menzioni. L'obiettivo principale è quello di esaminare le caratteristiche topologiche evidenziate dalle dinamiche di propagazione delle informazioni, per poi analizzare il loro impatto all'interno delle reti di Twitter. A questo scopo, viene utilizzato l'algoritmo di Louvain per identificare la partizione in comunità della rete, definite come gruppi coesi di utenti, in modo da associarle alle diverse coalizioni politiche, come CDX, CSX, M5S e CEN. L'analisi delle reti, ci consente poi di determinare i diversi ruoli degli utenti, evidenziando chi ha contribuito in modo significativo al dibattito politico legato alle elezioni. Infine, viene effettuata un'analisi delle serie temporali al fine di esplorare le diverse caratteristiche delle reti nell'intero periodo di interesse. L'analisi delle reti e delle loro proprietà viene condotta utilizzando il linguaggio di programmazione Python, nello specifico il pacchetto Networkx.

Parole chiave: Teoria delle reti, Dati Twitter, Elezioni politiche italiane, Analisi di comunità, Analisi dei ruoli, Analisi temporale

Acknowledgements

La fine di questo lungo percorso è finalente arrivata. E' stato difficile ma pieno di soddisfazioni. Se sono arrivata fin qui devo ringraziare sicuramente me stessa, ma non solo. Innanzitutto, ringrazio il professore Carlo Piccardi e Francesco Pierri per avermi seguito in questo lavoro finale, sempre disponibili e puntuali. Ogni incontro era pieno di idee e stimoli. Ringrazio poi la mia famiglia tutta, in particolare mio papà per avermi dato la testa per affrontare questo percorso, mia mamma per avermi dato la sensibilità necessaria per arrivare fino alla fine, mio fratello Federico, una sfida costante (ho vinto io). Ringrazio anche il resto della famiglia, cugine, zii, nonni e ultimi, ma non ultimi, i gatti. Ringrazio Patrick per avermi distratto, un po' troppo, in questi anni. Infine ringrazio tutti gli amici, quelli che c'erano già, vicini anche se lontani, quelli conosciuti durante questo percorso, che più di tutti mi hanno aiutato, quelli della birra del lunedì e del mercoledì, sfogo necessario.