SCUOLA ALTI STUDI LUCCA
SCUOLA NORMALE SUPERIORE
Consiglio Nazionale delle Ricerche

Scuola Normale Superiore

Classe di Scienze

Tesi di perfezionamento in Data Science

All the ties that bind. A socio-semantic network analysis of Italian political discussions on Twitter

Supervisors
Guido Caldarelli
Elena Pavan
Tiziano Squartini

Candidate
Tommaso Radicioni

Academic Year 2019-2020 (XXXIII cycle)

Chiediamo soltanto un po' di ordine per proteggerci dal caos [...] Ma l'arte, la scienza, la filosofia esigono di più: esse costruiscono dei piani sul caos. Queste tre discipline non sono come le religioni che invocano delle dinastie di dei o l'epifania di un solo dio per dipingere sull'ombrello un firmamento [...] La filosofia, la scienza e l'arte vogliono che noi strappiamo il firmamento e che ci addentriamo nel caos. Lo vinceremo solo a questo prezzo. E ho, tre volte vincitore, attraversato l'Acheronte. Il filosofo, lo scienziato, l'artista sembrano ritornare dal paese dei morti.
(Che cos'è la filosofia?, G. Deleuze F. Guattari)

Thesis abstract

Social media play a crucial role in what contemporary sociological reflections define as a 'hybrid media system'. Online spaces created by social media platforms resemble global public squares hosting large-scale social networks populated by citizens, political leaders, parties and organizations, journalists, activists and institutions that establish direct interactions and exchange contents in a disintermediated fashion. In the last decade, an increasing number of studies from researchers coming from different disciplines has approached the study of the manifold facets of citizen participation in online political spaces. In most cases, these studies have focused on the investigation of direct relationships amongst political actors. Conversely, relatively less attention has been paid to the study of *contents* that circulate during online discussions and how their diffusion contributes to building political identities. Even more rarely, the study of social media contents has been investigated in connection with those concerning social interactions amongst online users. To fill in this gap, my thesis work proposes a methodological procedure consisting in a network-based, data-driven approach to both infer communities of users with a similar communication behavior and to extract the most prominent contents discussed within those communities. More specifically, my work focuses on Twitter, a social media platform that is widely used during political debates. Groups of users with a similar retweeting behavior - hereby referred to as *discursive communities* - are identified starting with the bipartite network of Twitter *verified users* retweeted by *non-verified users*. Once the discursive communities are obtained, the corresponding semantic networks are identified by considering the co-occurrences of the *hashtags* that are present in the tweets sent by their members.

The identification of discursive communities and the study of the related semantic networks represent the starting point for exploring more in detail two specific conversations that took place in the Italian Twittersphere: the former occured during the electoral campaign before the 2018 Italian general elections and in the two weeks after Election day; the latter centered on the issue of migration during the period May-November 2019. Regarding the social analysis, the main result of my work is the identification of a behavior-driven picture of discursive communities induced by the retweeting activity of Twitter users, rather than determined by prior information on their political affiliation. Although these communities do not necessarily match the political orientation of their users, they are closely related to the evolution of the Italian political arena. As for the semantic analysis, this work sheds light on the symbolic dimension of partisan dynamics. Different discursive communities are, in fact, characterized by a peculiar conversational dynamics at both the daily and the monthly time-scale. From a purely methodological aspect, semantic networks have been analyzed by employing three (increasingly restrictive) benchmarks. The k-shell decomposition of both filtered and non-filtered semantic networks reveals the presence of a core-periphery structure providing information on the most debated topics within each discursive community and characterizing the communication strategy of the corresponding political coalition.

Keywords: *Complex networks, Social networks analysis, Semantic networks, Null models, Network filtering, Bipartite networks projection, Twitter.*

Contents

List of Figures

List of Tables

List of Acronyms

Null models used in the bipartite networks projection

BiCM Bipartite Configuration Model
BiPCM Bipartite Partial Configuration Model
BiRGM Bipartite Random Graph Model

ERG Exponential Random Graph

Discursive communities recovered in the users' analysis

CDX Center-right and far-right
CSX Center-left

DX Far-right

FI 'Forza Italia'

LEU 'Liberi e Uguali'

M5S 'Movimento 5 Stelle'
MEDIA Media organizations
MINGOs Media, international governmental and non-governmental organizations

Chapter 1

Introduction

In January 2021, the global digital population active on social media amounted to 4.66 billion individuals [1] whereas, according to the Euro-barometer [2], the percentage of EU citizens who use social media on a daily basis has increased from 18% in 2010 to 48% in 2019. The unprecedented amount of data made available by the massive usage of social media has paved the way for the study of online human behavior in a data-driven fashion. This so-called 'data deluge' [3] poses several challenges: on the one hand, it provides a unique environment to test pre-existent theories and methodological approaches across traditional disciplinary boundaries; on the other hand, it calls for the definition of novel methods and resources - such as data analysis techniques and filtering procedures - to extract insightful information from the newly available data structures.

Most studies focus on the analysis of *networks of users* approaching the study of online political dynamics through the investigation of direct relations amongst actors of a different nature - individuals, organizations, institutions and even bots. Conversely, less attention has gone towards the *contents* that circulate during online political discussions and how these contribute to nurturing collective political identities which, in turn, drive political action and participation. Moreover, the social and semantic aspects have been studied so far independently and the exploration of the nexus between contents of online political conversations and the relational systems amongst users sustaining them is still in its infancy.

At the intersection of several disciplines such as network science, data science and sociology, the following PhD thesis is the result of a multidisciplinary research on the effect that the formation of online user communities has on both the social and the semantic dynamics within social media platforms. Although Twitter is not the only social media platform that hosts politically relevant discussions, the current study focuses on this platform because of the large avalaibility of data that can be harvested through open APIs and its prominence during societal and electoral discussions. In the present work, I selected two case studies representing two distinct online conversations that unfolded on the Italian Twittersphere, namely the discussion (i) in the run up to the 2018 Italian Elections and during the two weeks after Election day and (ii) on the issue of migration during the period May-November 2019.

This PhD work proposes a network-based and data-driven approach that can be delineated starting from these discussions in the longitudinal exploration of the twofold set of social and semantic structures. Following the approach outlined in [4–7], the present work identifies broader *discursive communities* that form around Twitter verified users who are considered as proxies of enlarged digital elites. Following attention flows that pass through the retweeting activity of non-verified users, the identification of such communities is grounded in the role of verified users in the Twitter communication processes within the peculiar space afforded by this platform [8, 9]. Accordingly, the discursive communities are identified as *groups of Twitter users who share significantly similar retweeting patterns* and the contents circulating within these online political communities are explored in connection to the relational structures of communities themselves.

The network approach proposed in this work provides a method for extracting relevant political information from a Twitter discussion without relying on any pre-existent information on users or media contents. In this sense, the present procedure of network inference offers two main advantages with respect to other methods adopted in the extant scientific literature. First, the inference method adopted for the present analysis is based on a very general principle rooted in statistical physics [10], i.e. entropy maximization under certain constraints, which guarantees that the procedure is unbiased. In fact, relevant information about the political affiliation of the discursive communities and the structures of semantic networks are extracted without resorting to any external information, i.e. without passing through any manual labelling of users or media contents (e.g. as performed in [11, 12]). As a consequence, this method is suitable for application to *any* Twitter discussions regardless of the language or the topic addressed. Second, the semantic networks characterizing each community are obtained by connecting any two hashtags if used a significantly large number of times by the users of that community, hence overcoming the limitations of other analyses [11, 13] where online political conversations are studied in an unrelated fashion with respect to the relational system amongst users sustaining them.

One of the main findings of the present work is that discursive communities mirror consistently the political coalitions running for the 2018 Italian general elections and that they are also sensitive to the dynamics characterizing the relationship within and between these coalitions. These common characteristics are a clear indication of a coherent coordination between the online communication behavior of the political actors present in each discursive community and their political strategies. A second core result of my analysis concerns the study of the mechanisms that shape the Twitter discussions characterizing the aforementioned discursive communities. By monitoring, on a daily and monthly basis, the structural evolution of the community-generated semantic networks, each discursive community reveals a significantly different conversational behavior inducing semantic networks with peculiar topological structures. Semantic results further confirm the fluid nature of online political dynamics which requires an investigation of its evolution across different time scales.

In this work, particular attention has been paid to the online polarization dynamics which here is understood as a relational phenomenon that is, at the same time, *collective* and *multidimensional*. In characterizing online polarization as a collective phenomenon, the proposed methodological approach acknowledges the multiplicity of actors jointly contributing in different forms and with different levels of power and authority to shape political discussions as oppositional spaces. However, more than a strict adherence to partisan membership, users partisanship configures itself in a discursive fashion and can be grasped by looking at the interactions between users interested in specific discussion topics (in the present case, migration and electoral issues). At the same time, this work aims at shifting away from more traditional views of polarization that see different agencies as bound to party logics and hierarchically ordered (typically, top-down). Consistently, it explores more in detail how polarization occurs within communities that arise from the joint communicative behaviors of both citizens and public figures. Additionally, in defining online polarization as a multidimensional phenomenon, this work expands cross-dimensionally the scope of analysis by linking online endorsement and framing dynamics with offline events and political arrangements. In fact, online dynamics are here conceived as networked discursive processes unfolding fluidly, together with conversational and attention flows, within the online space opened by social media platforms. Through these dynamics, attention is directed towards specific actors and particular frames are brought into prominence.

The present thesis is divided into six main chapters. Chapter 2 will present an overview of the state-of-the-art on the studies about newer and older features interacting in the sociotechnical environment of social media platforms. Against a multidisciplinary background, this Chapter provides a comparison between the main advances of this work and the extant scientific literature. In Chapter 3, the two selected case studies and the Twitter data sets associated with them will be presented. Chapter 4 will offer a detailed description of the methodological approach adopted for the construction of semantic and user networks which lie at the core of the present quantitative analysis. Finally, in Chapters 5 and 6, I focus my attention on the main results of the analysis characterizing both online user communities and the semantic aspects of their Twitter communications.

Chapter 2

Social media analysis: A multidisciplinary state-of-the-art

In the last decade, the emergence of social media platforms has deeply changed the way in which information is produced and consumed in our daily lives. As these platforms facilitate the rapid exchange of media contents and large-scale information cascades, what emerges is a shift from a *mediated* and *top-down* communication model heavily ruled by traditional mass media to a *disintermediated* and *horizontal* model in which citizens actively select, share and contribute to the production of politically relevant information, in turn affecting the social life of their countries [14]. Many researchers across traditional disciplinary boundaries have multiplied their efforts to uncover the many implications of users online behavior for political participation and democratic processes.

In the following Sections, these studies are going to be reviewed. Section 2.1 provides a brief description of the tight entwinement of social media platforms with traditional mass media in the multifaceted environment called 'hybrid media systems' [15]. Section 2.2 provides an overview of the general features of social media induced by the emergence of online collective phenomena, e.g. the formation of echo-chambers. Section 2.3 focuses on the phenomena of polarization and networked framing within the digital space of social media. Finally, Section 2.4 reviews recent studies about social and semantic aspects of social media platforms.

2.1 The new hybrid media logics

The radical transformation brought by social media platforms has permanently changed assumptions, norms, organizational forms and even the tangible technologies of the media. Traditionally, these characteristics are condensed in the concept of 'media logic', defined as «the imperative that shapes the particular attributes and ways of doing things within given media» [14]. The original idea of media logic appears to be limited when the power of traditional media institutions, such as newspaper or television companies, is played out in a diverse and polycentric media environment of digital communication. The idea of a single media logic pervading social and political life is thus disaggregated into different competing, yet interdependent, *media logics* which can be seen as a 'force' co-created by media institutions, political actors and the audience [15].

Within the current digital communication environment, digital contents are indeed not only created by traditional mass media but also by single users who behave like *produsers* [16]. The *produsage* process is not simply the usage or the production of new contents but their collaborative alteration which allows individuals to easily switch between the two roles of user and producer of information. In this sense, the traditional one-directional mass communication process is radically changed with the emergence of social media in a many-to-many *mass self-communication* [17]. The term *mass* highlights that posting or sharing in this new digital environment has the potential to reach a wide audience while *self-communication* means that the media contents are self-generated but also that the retrieved information is self-selected. The self-mass, the interpersonal (or *one-to-one*) and the mass (or *one-to-many*) communication co-exist within a digital and interactive space whose organizational structure is redefined by novel social (and power) relationships. This transformation has been made possible by the technological change which has affected the means of communication: the 'deep mediatization' [18] has facilitated the penetration of such means of communication into our everyday life, permanently altering the way we interact with information and communicate with people.

In fact, as pointed out by a recent Eurobarometer survey [2], the percentage of European people employing Internet to access information, on a daily basis, has increased from 18% in 2010 to 49% in 2018. Even though television is still the preferred source of news on national political affairs by 76% of EU citizens, these numbers are not able to reflect the intertwining nature of the media landscape. In the last decade, this new sociotechnical environment has become a complex and fluid space where digital media technologies and norms are able to interact dynamically with traditional mass media, i.e. the so-called *hybrid media system* as defined by Andrew Chadwick [15]. In his idea, hybridization in media is a constant process of integration of older means of communications with newer technologies that reconfigures social and power relationships among existing media companies, political actors and the audience by providing different terms of engagement for the interactions among these groups. As a consequence, this process generated a reconfiguration of older media, such as television and printed newspapers, which evolve towards new forms of interactivity in response to newer media as social media platforms.

As an example of media hybridization, political actors, newspapers and journalists have adopted communication practices that are typical of the social media environment thus giving rise to what Henry Jenkins has called *transmedia storytelling* [19]. Storytelling is traditionally viewed as the activity of creating and sharing narratives according to a specific media logic. A remarkable example of transmedia storytelling is the online commentary performed when television shows are aired: online users are both the subject, as their messages are displayed during the program, and the object, as the television program makers seek to create contents that will activate such connections across media. In this sense, the engagement and the interactivity of social media allow the audience to be not only a passive user but a ‘produser’ of media contents. Traditional media logics are able to exercise a substantial power also in political dynamics when online users are called on to express their views or interact directly with political candidates. For instance, the *hybrid media campaigning* of Donald Trump’s nomination [20] has been a successful experiment of this interaction between traditional mass media and social media platforms. Donald Trump’s efforts to court media attention has resulted in an electoral strategy where conventional media coverage of rallies, press conferences and interviews have been complemented by a massive presence of social media through which supporters are encouraged to extend his narratives. The recombination and mixing between media logics owning to traditional mass media and social media platforms is an essential aspect for studying the conversational dynamics between media institutions, political actors and the audience.

2.2 Echo chambers and filter bubbles

The new sociotechnical environment of social media platforms, while incorporating traditional mass media logics, presents innovative features that drive communication processes like the access to an unprecedented availability of information often selected through algorithmic mechanisms based on prior interests and choices [21]. These features together with the tendency of users to directly interact with a limited subset of media contents [22] are mainly responsible for online collective phenomena such as the formation of the so-called *echo-chambers* [23, 24]. Their existence and their role in shaping the social media environment have been extensively debated: while some researchers argue that their relevance is overstated [25], echo-chambers have been observed to characterize a wide variety of real-world social networks [23, 26].

The emergence of homophilic groups of online users has been seen as the effect of diverse social and technological mechanisms, such as the tendency of individuals to select information confirming their beliefs. This tendency leads to a shared common narrative [23] and online political fragmentation [26]. In fact, as several studies show, the tendency to share and interact with media contents which confirm pre-existing opinions, the so-called *selective exposure* [22, 23], is a collective phenomenon acting as a mechanism to reinforce an existing opinion within a group. As a result, the entire group moves towards more extreme opinions causing *group polarization* [27]. Thus, interactions between groups with more extreme positions occur less frequently than between groups with moderate positions leading to an online political fragmented scenario [26].

There is another mechanism that is rooted in the digital nature of social media: in fact, these platforms are not uniquely passive means of communication but also data generators when they are used for communication tasks. Social media data are used as a source for several forms of automated processing algorithms which are increasingly relevant in online dynamics. In contrast with traditional mass media, non-human actors, such as platform algorithms [21] and bots [28, 29], give an active contribution in shaping online political discussions and information sharing processes. Consideration for non-human actors follows from extant computer science studies, such as those about social bot detection [30], and social science approaches, such as the actor-network theory which proposes disanchoring agency from social actors and prefers a recognition for *actants*, that is, for any agent capable of intervening within social dynamics [31].

The pervasive diffusion of social media platforms in every domain of human activity has called greater attention to both platform materiality (i.e. the modes in which specific technological artifacts are constructed and function) and to actants such as algorithms and bots, starting from the premise that online dynamics are inherently sociotechnical and, thus, technology features stand in a mutual and co-creative relationship with their social understanding and uses [32]. Shrouded in invisibility, platform algorithms and social bots actively filter and/or push specific types of contents thus managing to manipulate users' opinions [21, 23] - in some cases acting as true agents of disinformation [30, 33]. In this sense, the emergence of a *filter bubble* [21] is an algorithmic consequence mainly driven by algorithmic biases induced by the web search engines and social media. In the same way in which traditional mass media can give relevance to news contents through agenda sorting and setting, web search engines and social media are primarily responsibles for the *threat of invisibility* [34], namely the possibility to cause media contents to disappear from the online feeds of other users. Moreover, web search engines and social media are the primary gatekeepers of online media contents with strong effects on online diversity and biases [35, 36].

2.3 Online polarization and networked framing

A wide variety of extant studies on social media focus on the analysis of social network structures associated with online political discussions. Within this conversational dynamics, the formation of homophilic groups of online users is usually associated with group polarization. Ironically, the very debate on the political role of social media communications has initially configured itself as highly polarized, opposing optimistic views on the potentialities of these tools to enhance democratic participation [37] to pessimistic opinions, depicting the fragmentation and tribalization of the public discussion, most notably within echo-chambers [38, 39]. More recently, sustained theoretical and empirical interest for mapping and addressing the consequences of homophilic communication dynamics, especially with respect to the diffusion of disinformation [40, 41], has been flanked by attempts to complicate the very concept of polarization, suggesting that it has a hybrid interactional, positional and affective nature [42]. While interactional polarization concerns the preferential conversational attachment of like-minded individuals as described in the previous Section, the other two aspects regard the increase of antagonistic and extreme political preferences [43] and the rising of hostility in political rhetoric addressed to opposing viewpoints [44]. As peculiar differences in social media platforms affect echo chamber formation [23], specific algorithmic and conversational features of each social media platform contribute to different polarization mechanisms [42] adding another layer of complexity to this debate.

Online political polarization is intertwined with several offline dynamics, such as the transformation of the elite party competition. In fact, *elite polarization* - that is, patterns of progressive spacing between parties and party representatives compensated by increasing internal proximity - has been shown to change the decision-making process of citizens [45] and to drive *mass polarization* [46]. While the studies on polarization based on social media leverage on the potentialities of large-scale data sets to advance our understanding of its effects, less attention has been paid to the existing interconnections between elite and mass polarization also online. Empirical analyses show that increasing elite polarization dramatically affects citizens' decisions by stimulating 'partisan motivated reasoning' which binds individual opinions to partisan endorsement rather than to substantial reasoning [45]. Relevantly, similar conclusions are made when elites are not restricted within party and ideology-related boundaries but open up to include partisan media [47] (i.e., media outlets that aim at advancing particular political agendas) and mainstream outlets [48–50] as well.

Polarization mechanisms emerge in the formation of homophilic groups of online users also when increasingly distant antagonistic narratives are built within them. As pointed out in [51, 52], the act of generating and sharing media contents is a constitutional element for the formation of political collective identities and framing processes. Common to both online and offline studies of polarization is an attention for different modes of understanding or, more broadly, framing issues that are of common concern. Even when conceptualized in affective terms [44], polarization remains anchored to situated understandings of reality which are found to relevantly affect voting and participation behaviors as well as opinions [45]. Also in the manifold cases in which online polarization is investigated in terms of concentration of homophilic relations, users' positioning is always derived with respect to their understanding of specific issues, which is nonetheless often subsumed within users' ideological placement [53] or expressed opinions [54]. Relevantly, besides polarization studies, sociological investigations of online political networks are shedding light on the framing power that characterizes social media practices, particularly *hashtagging*, which can epitomize the formation of counter-publics challenging mainstream narratives [55] but also channel affective dynamics of high political relevance [56].

While traditional news organizations remain privileged actors in determing how major events are presented and interpreted, within the contemporary hybrid media system [15] new practices of framing like hashtagging emerge from the collective interactions between users across various social media platforms. Thus, *networked publics* shaping the news topics, views or interpretations have been shown to be able to induce a *networked framing* resulting from users' social media activities such as sharing, liking and commenting [57]. In this framework, networked framing refers to the collective behavior within social media platforms through which news is presented and interpreted by non-elite actors and transformed into popular narratives [58]. The emergence of a greater plurality of views does not necessarily reflect an increased diversification of the public debate: as mentioned above, social media communications may lead to a political polarization where users are more likely to form homophilic groups where partisan views are dominant [38]. In the networked framing perspective, polarized communities induce a highly partisan discussion through both the divergence of antagonistic framing [59] and the selective exposure which is strongly influenced by algorithmic mechanisms pushing media contents that preferentially adhere to particular frames [21]. In this sense, online polarization is a multifaceted issue which can be observed in the combination of a political scenario of fragmented online user communities and the distinctive semantic practices of networked framing shaping them.

2.4 Combining social and semantic analysis

The systematic investigation of online networks arising from social media use during relevant political discussions has been particularly helpful for studying the social and semantic dynamics within online user communities. Endorsing a view of online political activism as complementary to - and not as a substitution for - traditionally studied political participation dynamics [60], detailed and data-intensive explorations of online systems of interactions contributed to a more genuine and multilevel understanding of how social media relate to political participation processes.

On a macro (or global) level, research endeavors have focused on mapping the structural and processual features of online interaction systems in order to elaborate on the social media potential for fostering democratic and inclusive political debates. In this respect, as already pointed out, specific attention has been paid to assessing grades of polarization and closure [27] of online discussions within echo-chambers [23] with a view to connecting such features with the progressive polarization of political dynamics [39, 61, 62]. Online dynamics are also directly connected with the study of collective action [63] which is an interaction pattern widely investigated to gain insights into the explanation of collective phenomena, such as industrial innovation [64], offline and digital mobilizations [51, 60] and political participation [65]. In several environments, social actors have been shown to possess a *collective identity*, namely a consistent aggregated behavior displayed in interaction networks despite a lack of formal organization [66]. For instance, in a recent paper [67], the collective identity on social media platforms such as Twitter has been defined as «cohesive and coordinated communicative interaction networks». On a micro (or local) level, research has focused on disambiguating the different roles that social media users may play within online networks - particularly, to identify influential spreaders [68] responsible for triggering the pervasive diffusion of certain types of information, but also to elaborate on the redefinition of political leadership in comparison to more traditional offline dynamics [69]. More specifically, accounting for users behavior has helped to characterize the different contributions delivered by political actors and citizens who exploit to various extents social media communication [70, 71] and networking potentials [17] to intervene and take an active stand within public protests, electoral campaigns and, more in general, within political dynamics they consider to be more relevant and worth participating. In this way, concepts like 'political relevance' and 'leadership' get redefined at the crossroads between actors attributes and their actual engagement in online political discussions.

In all its heterogeneity, the common feature shared by all these works is that they are grounded in the study of social media users interactions and, thus, approach the study of online networks by giving priority to the investigation of direct relations amongst actors of a different nature - individuals, organizations, institutions and even bots. Studies that focus on the circulation of media contents during online political discussions and their contribution to foster collective identities have been put forward less often, however embracing a multiplicity of political instances, from electoral campaigns to social movements and protests. For instance, by leaning exclusively on Twitter data as the present thesis, researchers have compared the content of tweets published by parties with the content of tweets sent by candidates [11], analyzed the contents of the 2017 French presidential electoral campaign [72] and the online media coverage in the run up of the 2018 Italian Elections [73] and looked at the keywords and hashtags related to the #MeToo movement [13]. Nonetheless, when the focus has been set on social media contents, only rarely have these been investigated in connection with systems of social relations established amongst users upon social media platforms [74]. Ultimately, social and semantic structures of online political conversations have been explored independently without bridging the study of the social media contents circulation with that of the relational systems amongst users sustaining them.

The present PhD thesis expands current analyses of online conversational dynamics by proposing an innovative operational procedure which brings prominence to semantic aspects without neglecting the social side of online dynamics. This research starts by applying recent approaches to the identification of groups of Twitter users who share a significantly similar retweeting behavior [5]. However, rather than simply investigating the social dynamics through methodological approaches applied in existing works [11, 12], the PhD thesis pushes these approaches one step further leaning on the operational definition of discursive communities as a starting point to infer the semantic structures. In fact, the semantic networks analyzed in this work follow from the identification of discursive communities based on an entropy-based framework[4–7]. Twitter has been considered as a privileged digital space strongly influenced by offline events where a complex online dynamics of social and semantic interactions occurs between a wide variety of political actors [75]. Temporal changes in social structures as discursive communities (or similarly echo chambers) are analyzed in relation to online polarization, here understood as a collective and multidimensional relational phenomenon. Additionally, networked framing practices occurring within each discursive community are examined by studying the semantic networks they induce via hashtagging practices, recognized as powerful framing procedures [56, 58, 76].

Chapter 3

Data and case studies

The present Section provides an overview of two selected case studies and the corresponding data sets. The present work builds upon two data sets concerning distinct conversations of political and societal nature in the Italian Twittersphere. The first data set concerns the discussion occurred during the weeks of the electoral campaign before the 2018 Italian general elections and the two weeks after Election day, i.e. March 4^{th}. The second data set is retrieved by considering a broader discussion on the issue of migration and the refugee crisis taking place over a wider time period spanning May and November 2019.

A brief review of the political background of these two Twitter data sets is reported in Section 3.1 while the description of the data acquisition process is provided in Section 3.2.

3.1 Overview of the political background

3.1.1 2018 Italian general elections

The 2018 Italian general elections were a relevant electoral event where the tripolar competition of the main political coalitions consolidated [77]. The balance of power among these political poles radically changed, representing a crucial tipping point and overturning the traditionally bipolar electoral competition of the so-called 'Italian Second Republic'. The first electoral earthquake of the 2013 general elections led to the collapse of the bipolar system due to the decreased number of votes of both the center-right and center-left coalitions and the unexpected electoral success of the populist party 'Movimento 5 Stelle' (M5S, *Five Stars Movement*). No clear majority emerged from these general elections and the subsequent governments have been supported by a heterogeneous parliamentary majority consisting of center-left, center and the center-right parties.

After the formation of three governments headed by three members of the major center-left party *Partito Democratico* (PD, *Democratic Party*) in succession (Enrico Letta, Matteo Renzi and Paolo Gentiloni), the novel Italian political scenario before the 2018 general elections was composed of three poles of electoral attraction. The first pole, the center- and far-right coalition, eventually won the elections with 37% of the votes. The coalition was composed of the post-fascist area represented by 'Fratelli d'Italia' (FDI, *Brothers of Italy*), the post-Christian Democratic 'Noi con l'Italia-Unione di Centro' (NCI-UDC, *Us with Italy-Union of the Centre*), Silvio Berlusconi's center-right party 'Forza Italia' (FI, *Go Italy*) and the populist radical right-wing party 'Lega' (*League*). The latter, a former northern regionalist party called 'Lega Nord' (*Northern League*) turned into a right-wing party with nationwide appeal under Matteo Salvini's leadership, received the highest number of votes in the center-right coalition with 17,4% of preferences. The second pole was represented by a center-left coalition guided by the 'Partito Democratico' and its secretary and Prime Minister candidate Matteo Renzi. A joint list of other parties to the left of the 'Partito Democratico' called 'Liberi e Uguali' (LEU, *Free and Equals*) ran for elections separately from the 'Partito Democratico'. The third pole represented by the populist party 'Movimento 5 Stelle' was the most voted party with 32.7% of the vote share. Since September 2017, Luigi di Maio has been the formal political leader of the 'Movimento 5 Stelle', after having won the online primaries of the party.

The Italian political actors campaigned by taking up a position on policies mainly regarding economics and migration [77]. While the center-right coalition pushed for tougher immigration policies and for a flat taxation system, the most relevant proposal of the 'Movimento 5 Stelle' concerned the 'reddito di cittadinanza' (i.e., *citizens' income*), a public-funded monthly basic income for people below the poverty threshold. At the same time, the 'Movimento 5 Stelle' maintained an ambiguous position on migration. Finally, the center-left coalition campaigned for the continuity of the former government's economic policies and for a reduction in immigration flows. Even though the 2018 hybrid campaign has been characterized by the combined communication strategy at the intersection between traditional and digital media logics [78], the social media communication habits of the main Italian leaders and parties are profoundly grounded in the use of social media platforms. In particular, in the last decade, Twitter has significantly increased its prominence in affecting the political discussions in the electoral process [75, 79]. In fact, since the 2013 Italian general elections which have been considered the first 'Twitter Italian general elections' [80], most Italian politicians flocked to this platform to increase the audience of their electoral campaigns and to give more resonance to their statements.

3.1.2 2019 Italian discussion on migration

Since the beginning of 2014, when the number of migrants attempting to enter the European Union has massively increased, the refugee crisis has been a widely discussed hot topic in the European mainstream media [81]. This topic has gained prominence due not only to a proliferation of messages with negative connotations on refugee issues broadcast by mass media and published upon social media platforms [57] but also to European political parties fuelling xenophobic or anti-immigration policies or narratives and building on social media plaforms a rhetorical construction of migrants and refugees as a dangerous threat for national security [82].

After the 2018 Italian general elections, described in detail in the previous Subsection, in June 2018 the two initially competing parties, the ‘Movimento 5 Stelle’ and the ‘Lega’, emerged as winners to then form a new government. In the self-proclaimed ‘government of change’, the ‘Lega’, albeit being the minority partner in the coalition, and Matteo Salvini, as Minister of Interior and vice-Prime Minister, successfully succeeded in influencing the political agenda of the government on several topics at the center of its political project, particularly migration policies [83]. One of Matteo Salvini’s most debated decisions was the closure of the Italian ports to the NGOs boats which were rescuing migrants in the Mediterranean Sea. This political initiative not only took the tangible form of entry bans into Italian ports or seizure of several rescue boats but also in the approval of two ‘security decrees’ (the second called ‘Decreto Sicurezza Bis’ or ‘Second Security Act’ approved in June 2019) that made it more difficult for asylum seekers to request the residence permits issued for humanitarian reasons and that introduced fines for NGOs active in migrant rescue activities. On 19 August, approximately 14 months after its formation, the governement fell after Matteo Salvini and his party submitted a no-confidence motion against the Prime Minister Giuseppe Conte. A few weeks later, the ‘Movimento 5 Stelle’ formed a new government with the ‘Partito Democratico’ and other minor forces of Italian left-wing parties while the ‘Lega’ became an opposition party. The new government immediately presented a different attitude towards the NGOs involved in ‘search and rescue’ operations: on 26 October, the new Minister for Internal Affairs, Luciana Lamorgese, met the organizations to discuss their activities.

Other major political events of interest for the current analysis, occuring during the data acquisition period are mainly two: the 2019 European elections on 26 May and the entrance of the rescue boat ‘Sea Watch 3’ into Italian territorial waters without permission at end of June 2019. The former is a significant electoral event since the ‘Lega’ became the most voted Italian party with 34.1% of the vote share reversing the political result of the 2018 general elections in which ‘Movimento 5 Stelle’ emerged with a similar percentage of preferences. The latter is one of the most debated events about migration before the fall of the government: after two weeks sailing, on 29 June 2019, the rescue boat ‘Sea Watch 3’ entered Italian territorial waters and its captain Carola Rackete was subsequently arrested. Matteo Salvini as the Minister for Internal Affairs accused Rackete of trying to ram an Italian patrol boat. The Italian vessel had tried to intercept ‘Sea Watch 3’ before docking and a collision ensued. At the same time, Rackete was under investigation by Italian authorities for alleged criminal offenses concerning undocumented activities in ‘search and rescue’ operations. After a brief detention under house arrest, she was eventually released by an Italian court ruling that asserted that she had acted to protect the passengers safety.

3.2 Data sets description

The reliability of social media data for gaining insights into political processes is a widely debated topic. Due to the large availability of data extracted through its programming interfaces (APIs), Twitter has quickly become the main platform around which the investigation of social and political dynamics takes place. Consistently, in what follows, I will focus on the debates unfolding on this specific platform. Several researchers argue that Twitter does not offer an unbiased and representative picture for studying societal events such as electoral contests. These studies pointed out two major issues that make it impossible to use online data for offline predictions of political and public events: on the one hand, Twitter users tend not to be representative of the a country's population as a whole [84]; on the other hand, Twitter's Streaming and Search API mostly rely on the freely available 1% sampling technique which does not guarantee an unbiased data selection [85].

Nevertheless, extant studies show that Twitter is massively employed in correspondence with public events such as electoral debates [61, 79] and has a strong influence that enhances participation and inclusivity during demonstrations [70] or societal discussions [59]. Furthermore, Twitter is used by a vast majority of public figures (e.g. political leaders, journalists, official media accounts, etc.) who give visibility to their statements in this platform. In the Italian hybrid media system, Twitter is recognized to affect the public debate and to have an 'agenda setting' effect [75] due to its costant relationship with mainstream news media. For all these reasons, although Twitter users are known not to be representative of the Italian population [86], the following data sets are considered as providing a relevant starting point to inquiry the political discussions around the two selected case studies.

When studying collective political identities, amongst the various types of Twitter interaction, retweets are recognized as a particularly insightful relational mechanism. While mentions and replies sustain direct interaction and dialogue between users, retweets signal an explicit recognition (for better or for worse) and a deliberate broadcasting mechanism of the contents published by a specific user. As such, retweets can be conceived of as a powerful 'mode of repetition' [58] able to reinforce collective political identities [12, 60]. Albeit not necessarily implying an endorsement or an alignment between ideological positions, the overall effect of retweeting is to draw attention to the same piece of information. This, in turn, becomes relevant to sustain the formation of collective identities [60]. Moreover, the retweeting activity of online users has been recognized as proxying actual political alliances better than mentions and replies. For instance, in [12], authors conclude that the use of retweets was more relevant than that of mentions to uncover the bipartisan structure of online debates in the run up to the 2010 US midterm elections.

A brief overview of the main features of the two data sets is reported in Tab. 3.1. The 2018 Italian Election data set has been extrapolated by employing the Twitter Search API

	2018 Italian general elections	2019 Refugees crisis
Total number of tweets	1.199.001	4.924.432
Original tweets	331.721	854.882
Retweets	809.569	3.616.966
Replies	57.711	452.584
Users	123.210	306.894

Table 3.1: Brief overview of a set of quantities for both data sets. For each data set, the following quantities are reported: total number of tweets, number of original tweets, retweets and replies and total number of users.

with a set of terms linked to the discussion of the present case study. In particular, each tweet is collected on the basis of a set of anchor terms comprising highly-circulated hashtags and keywords related to Italian Elections that were prominent during the data acquisition period: *elezioni*, *elezioni2018*, *4marzo*, *4marzo2018* (literally, *elections*, *elections2018*, *4march*, *4march2018*). The anchor terms have been selected by the entire research team through a qualitative procedure. This operation consists of matching the information of Twitter trending topics with the most frequent hashtags within the tweets containing the anchor term *elezioni*. The data collection has been carried out across a time period of 51 days, from 28 January 2018 to 19 March 2018, which is an interval covering the overall electoral campaign (officially starting one month before the Election Day) and the two weeks after Election Day, i.e. March 4, 2018. The final data set on 2018 Italian Elections contains about 1.2 million tweets posted by 123.210 users uniquely identified via their user ID.

The 2019 Italian discussion about migration policies and the refugee crisis is mapped through tweets extracted via the Twitter Streaming API over the period from April 24, 2019 to November 24, 2019, requiring that each tweet contains at least one of the following anchor terms: *accoglienza* (*hospitality*), *apriteiporti* (*open the ports*), *chiudiamoiporti* (*close the ports*), *immigrazione* (*immigration*), *integrazione* (*integration*), *migranti* (*migrants*), *restiamoumani* (*let's stay human*), *rifugiati* (*refugees*), *sbarchi* (*landings*), *stopinvasione* (*stop the invasion*). Similarly to the case of the 2018 Italian Elections, these anchor terms have been selected by the entire research team through a qualitative procedure. This process is performed by comparing the information of Twitter trending topics with keywords used in literature to draw boundaries around migration debates [59, 87] as well as with hashtags referring to the main political slogans. The data acquisition procedure led to a data set of approximately 5 million tweets posted by 306.894 users uniquely identified via their user ID. It is worth noting that the present data set has been made public for reproducibility purposes in the Section *Datasets* of the TOFFEE (*TOol for Fighting FakEs*) project website [88].

Chapter 4

Methods for network filtering

In the last decade, network science [89, 90] has progressively emerged as a powerful framework to extract information from many different kinds of real-world networks, such as ecological networks [91], financial networks [92] and social networks [93]. Among its many applications, those concerning social systems increasingly gained popularity for studying dynamical processes, such as the spread of epidemics [94], and structural properties of complex interactions, as in social network analysis [24, 95]. Grounded in the theoretical framework outlined by these studies, this Chapter presents the approach adopted for building the communities of Twitter users and their corresponding semantic networks.

The Chapter is organized as follows: Section 4.1 provides a detailed overview of the implementation of the users bipartite networks and the user-hashtag bipartite networks, as presented respectively in Section 4.1.1 and 4.1.2. Section 4.2 and Section 4.3 offer a comprehensive description of the step-by-step procedures for the formation of the discursive communities and the semantic networks. Finally, Section 4.4 discusses the limitations of the current methodological approach.

4.1 Defining bipartite networks

Bipartite networks provide a useful representation of the relationships between two disjointed set of nodes, also called *layers*, where edges connect only nodes belonging to different sets. These kinds of networks are used in several research areas to gather information on mechanisms driving the organization of complex systems. For instance, the bipartite structure of the World Trade Web (WTW), representing the network of countries and products they export, has been analyzed in order to detect modules of similar industrial systems and recognize different strategies of specializations in the trading activities of countries [96–99].

4.1.1 Users bipartite networks

The present analysis focuses on all Twitter users who published tweets or retweeted messages containing *at least* one of the keywords in the lists reported in Section 3.2 and that are considered as meaningful anchor keywords for tracking the Twitter discussions under examination. These users are subsequently divided into two distinct groups: the former is composed by users who are *verified* by the platform, the latter gathers all the other *non-verified* accounts. The Twitter verified badge can be requested by any user to guarantee that the account is «authentic, notable, and active» [100]: therefore, verified accounts are usually associated with prominently recognized individuals, organizations or brands in line with notability criteria defined by Twitter itself. Despite pausing the profile verification process as of 2017 [101], Twitter's Head of Product declared that Twitter still grants verification on an *ad hoc* basis [102] confirming the enduring validity of the distinction between verified and non-verified users. The verified accounts in Twitter usually refer to politicians or parties active at the state or national level, companies and non-profit organizations, qualifying news organizations (as well as personal accounts of journalists), major entertainment companies, professional sports teams and players or other influential individuals and organizations. This information can be easily retrieved by employing the Twitter APIs.

Given these two groups of users, the users bipartite network is built as follows: a verified and a non-verified user are linked if one of the two accounts retweets the other *at least once* during the observation period. It is worth noting that in the selected Twitter data sets a retweet is most often executed by non-verified accounts in favor of verified users. A formal representation of the users bipartite network is given by the definition of an observation period $\mathcal{T}$ and the two layers of users denoted with $\top$ and $\bot$. While the observation period $\mathcal{T}$ is the final set of days when the Twitter data are retrieved, the two layers are defined as:

$$\begin{aligned} \top &= \{\alpha : \alpha \text{ is non-verified AND } \alpha \text{ retweeted at least one message of any } j \in \bot \text{ during } \mathcal{T}\} \\ \bot &= \{j : j \text{ is verified AND } j \text{ posted a message retweeted by any } \alpha \in \top \text{ during } \mathcal{T}\} \end{aligned}$$

In other words, $\top$ represents the list of unique non-verified users retweeting at least one tweet in the observation period $\mathcal{T}$ while $\bot$ represents the list of unique verified users tweeting (and being retweeted at least once) in the observation period $\mathcal{T}$. Given the set of edges $E \subseteq \top \times \bot$, the final set of retweets between a user $\alpha \in \top$ and a user $j \in \bot$ can be described as a bipartite network $\mathbf{M_u}$=($\top$,$\bot$,E). On the basis of this definition of a bipartite network, retweets between two verified accounts or two non-verified accounts are excluded. Each edge $(\alpha, j) \in E$ indicates that the non-verified user α has retweeted the verified user j's content at least once during the observation period $\mathcal{T}$. The bipartite network can be denoted as a biadjancency matrix $\mathbf{M_u}$= $\{m_{\alpha j} : \alpha \in \top \text{ and } j \in \bot\}$ where $m_{\alpha j}$ is equal to 1 if at least one retweet between nodes α and j exists and 0 otherwise. The final number of nodes and edges for each resulting bipartite network of the two data sets are specified in Table 4.1 while a

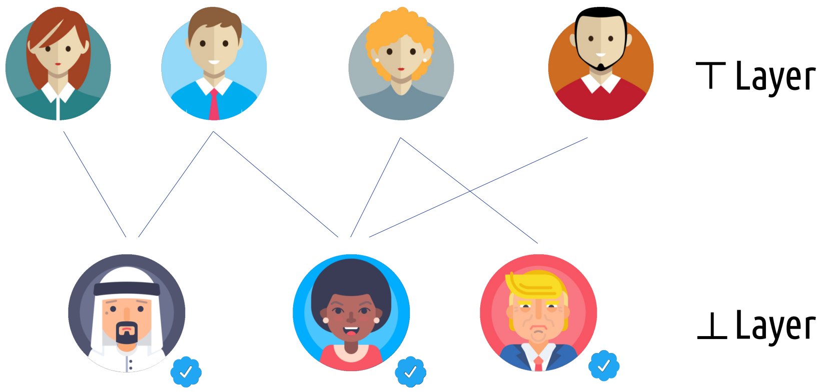


Figure 4.1: Diagram of the bipartite network of users $\mathbf{M_u}$. Each edge indicates that a non-verified user has retweeted a verified user's content at least once during the observation period $\mathcal{T}$. Given the Twitter verification policies, verified users are usually associated with prominently recognized individuals, organizations or brands in line with notability criteria.

visual sketch of the two layers of the bipartite network of users is reported in Figure 4.1.

Before moving on, two technical observations are necessary:

- Even though each tweet author is known, network edges of the present bipartite network are undirected. This network feature does not affect the final results since, as already reported, retweeted users are mostly verified ones. Furthermore, the initial bipartite network is only the initial step employed to assign a label to each verified user (Section 4.2) while each non-verified account will be then assigned an affiliation to a single community of verified users (Section 4.3). Thus, the effect of non-verified users retweeting activity is eventually recovered in the final discursive communities;

- The second remark concerns the centrality given in the following algorithm to the tweets posted by verified accounts as the initial source for the formation of consistent discursive communities: in fact, there are political or societal contexts, such as protests or mobilization, where the presence and/or the relevance of verified accounts could be limited in the overall Twitter discussion. Nonetheless, verified accounts are taken here as the starting point from which to build significant user communities since the users are digital elites who are perceived as more credible [8, 103]. This is consistent with my research interest in examining political discussions where political parties and leaders play a role as central actors who interact constantly with a variety of other Twitter accounts. Although the bipartite nature of the users networks causes the loss of information on the retweeting activity between verified users, it is worth noting that their mutual interactions are recovered in the final discursive communities.

Data set	$N_\perp$	$N_\top$	<k>	k_{min}	k_{max}	ρ
2018 Italian general elections	908	52622	4.3	1	5860	0.0024
2019 Refugees crisis	1144	115885	5.2	1	21545	0.0023

Table 4.1: Brief overview of a set of network quantities of the bipartite networks $\mathbf{M_u}$. For each data set, the following network quantities of the bipartite networks $\mathbf{M_u}$ are reported: number of nodes per layer, average degree, minimum and maximum value of the observed degrees, connectance of the network.

4.1.2 User-hashtag bipartite networks

Hashtags play a pivotal role in the Twitter environment where they function as *thematic tags* [104]. Consistently, the practice of hashtagging can be considered as an emergent organization whereby specific Twitter publics coalesce into a larger collective storytelling [55]. For these reasons, the communication within Twitter can be stylized in a bipartite structure of users who are connected with hashtags extracted from the text of their tweets. Hashtags are recognized in Twitter as they are preceded by the symbol #: as a consequence, tweets containing at least one hashtag can be easily retained for the construction of the bipartite network. Following this preliminary selection process, the final amount of tweets employed in the construction of the user-hashtag bipartite networks are ≃450 thousand tweets (≃ 37%) in the 2018 Italian general elections data set and ≃1.9 million tweets (≃ 38%) in the 2019 Italian migration discussion with respect to the initial data set. Even though their number is significantly lower than the initial amount of collected tweets, this sample is nevertheless considered as a valid proxy for the overall Twitter discussions as it focuses the analysis on topics, events and actors deliberately tagged in users' tweets.

Hashtags were subjected to a pre-processing procedure where any two hashtags have been merged if found to be 'similar' according to a specific measure. Finally, for each couple of similar hashtags, only the most frequent has been considered in the final list. Here, the similarity measure between hashtags has been computed through the *Levenshtein* or *edit distance* which is one of the most common sequence-based similarity measures [105]. This particular distance aims at spotting human editing errors, such as inserting extra characters or swapping any two characters. This pre-processing phase is needed to avoid duplication of strings occurence which can be responsible for creating two nodes of the same string in the final bipartite networks. The threshold for the maximum number of allowed differences between any two strings is set to 2, so that only occurrences of possible typing errors or different conjugations of verbs/substantives are merged. The following scenario is an example of what happens when this technique is applied: the hashtags *#migranti* (*migrants*) and *#migrante* (*migrant*) are merged together while *#migrante* and *#migrare* (*to migrate*) are considered as distinct occurences.

Given the list of users and the list of merged hashtags, the user-hashtag bipartite network is then built as follows: a user is linked to a hashtag if he/she tweets or retweets that hashtag *at least once* during the observation period. A mathematical representation of the user-hashtag bipartite network is given by the definition of an observation period $\mathcal{T}$ and the two layers of users and merged hashtags which are denoted with $\top$ and $\perp$, respectively. User-hashtag bipartite networks are built on a *daily* time scale in the case of the 2018 Italian general elections data set and on a *monthly* time scale in the case of the 2019 migration discussion data set. The observation period is thus respectively equal to $\mathcal{T} = 1$ day and $\mathcal{T} = 1$ month. The different levels of data aggregation have been chosen by considering the available amount of data and the time-scales of the selected online discussions. Moreover, these levels are selected by identifying time periods that suited their peculiar nature. A user-hashtag bipartite network is built for each main discursive community within the observation period $\mathcal{T}$. For each community, the lists of users and merged hashtags were then subdivided for each day or month of the observation period producing 51 different bipartite networks in the case of the 2018 Italian general elections data set and 7 different bipartite networks in the case of the 2019 migration discussion data set. In each user-hashtag bipartite network, the two layers are defined as follows:

$$\begin{aligned} \top &= \{u : u \text{ tweeted or retweeted a hashtag } h \in \perp \text{ during } \mathcal{T}\} \\ \perp &= \{h : h \text{ is tweeted or retweeted at least once by } u \in \top \text{ during } \mathcal{T}\} \end{aligned}$$

Summarizing, $\top$ represents the list of unique users u tweeting or retweeting a hashtag h at least once in the observation period $\mathcal{T}$ while $\perp$ represents the list of merged hashtags h tweeted or retweeted at least once in the observation period $\mathcal{T}$. Let $E \subseteq \top \times \perp$ be the set of edges. The Twitter conversation of each collected data set can be stylized as a bipartite network $\mathbf{M_h}$=($\top$,$\perp$,E). The bipartite network can be thus rewritten as a biadjacency matrix $\mathbf{M_h}$= $\{m_{uh} : u \in \top \text{ and } h \in \perp\}$ where m_{uh} is equal to 1 if node u has tweeted or retweeted at least once node h in the observation period $\mathcal{T}$ and 0 otherwise. Each edge $(u, h) \in E$ indicates that the user u tweeted or retweeted the hashtag h at least once during the observation period $\mathcal{T}$. The unweighted nature of the user-hashtag bipartite network is motivated by the fact that the number of times a hashtag h has been tweeted or retweeted by each user u is not as relevant as the co-occurrence of that specific hashtag with other hashtags during the observation period $\mathcal{T}$. In fact, since the present thesis is focused on the symbolic aspects of conversational dynamics produced by Twitter users, the inclusion of weight information is not only unnecessary for the purpose but yields also technical complexity to the overall inference procedure. A diagram of the two layers of the user-hashtag bipartite network is reported in Figure 4.2.

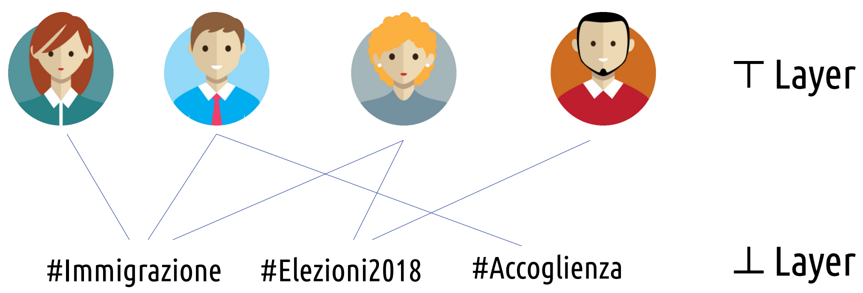


Figure 4.2: Diagram of the user-hashtag bipartite network $\mathbf{M_h}$. Each edge indicates that a hashtag h has been tweeted or retweeted at least once by user u during the observation period $\mathcal{T}$.

4.2 Projecting bipartite networks

The bipartite structure can be compressed into its *monopartite projection* containing only nodes of a specific layer and providing fine-grained inspections on nodes pertaining to that layer - whether users or hashtags. The resulting connection scheme depends on the edges surviving a proper projection algorithm. Thus, a methodological approach is needed to obtain a statistically validated monopartite projection while preserving the information encoded in the original bipartite structure.

The simplest projection is obtained by linking any two nodes of the same layer as long as they share at least one common neighbor. As observed in [98], this projection, called *naïve*, often results in an almost trivial topological structure. Since the present analysis is focused on tweets and retweets which are countable interactions, the corresponding monopartite network could be represented as a weighted network according to the number of common neighbors. Even though researchers have implemented algorithms to retain the significant weights, these algorithms are grounded on the arbitrariness of a chosen threshold [106] or on procedure based on null models that encode relatively little information of the original bipartite network [107].

4.2.1 ERG framework and projection algorithm

The approach adopted in this work is based on a general method in order to infer the monopartite projection by linking any two nodes in the same layer if their number of common neighbors is statistically significant. This class of algorithms is based on the computation of a similarity measure between two nodes of the same layer in terms of the shared portion of common neighbors [108, 109]. The mathematical framework beyond these algorithms is defined within the Exponential Random Graph (ERG) formalism [10, 97], an approach consisting in the constrained maximization of Shannon entropy. In mathematical notation, given the ensemble Ω of networks with the same number of nodes N of the real network $\mathbf{G}^*$ and $P(\mathbf{G})$ the probability of occurrence of a network $\mathbf{G} \in \Omega$, the Shannon entropy can be written as:

$$S = -\sum_{\mathbf{G}\in\Omega} P(\mathbf{G}) \ln P(\mathbf{G}) \tag{4.1}$$

This maximization procedure is subject to a normalization condition $\sum_{\mathbf{G}\in\Omega} P(\mathbf{G})=1$ and a collection of constraints $\mathbf{c}^*$ representing the aggregated macroscopic or mesoscopic properties of the system. The microcanonical ensemble imposes hard constraints assigning a uniform probability $P(\mathbf{G})$ to those network $\mathbf{G}$ that satisfy $\mathbf{c}(\mathbf{G})=\mathbf{c}^*$ and 0 otherwise. Since this approach is not feasible for analytical computation, the canonical ensemble is preferred. This ensemble is obtained by fixing the expected value of the constraints, that is:

$$\sum_{\mathbf{G}\in\Omega} P(\mathbf{G})\mathbf{c}(\mathbf{G}) = \mathbf{c}^* \tag{4.2}$$

The constrained entropy maximization assign an exponential probability to each network $\mathbf{G}$ equal to:

$$P(\mathbf{G}) = \frac{e^{-H(\mathbf{G},\boldsymbol{\theta})}}{\mathbf{Z}(\boldsymbol{\theta})} \tag{4.3}$$

where $H(\mathbf{G}, \boldsymbol{\theta})$ is the hamiltonian $H(\mathbf{G}, \boldsymbol{\theta})=\boldsymbol{\theta}\cdot\mathbf{c}(\mathbf{G})$ and $\mathbf{Z}(\boldsymbol{\theta})=\sum_{\mathbf{G}\in\Omega} e^{-H(\mathbf{G},\boldsymbol{\theta})}$ is the partition function. In order to determine the optimal set of Lagrange multipliers $\boldsymbol{\theta}^*$ for a given observed network $\mathbf{G}^*$, the maximization of the following likelihood functional is carried out:

$$\mathcal{L}(\boldsymbol{\theta}) = \ln P(\mathbf{G}^*|\boldsymbol{\theta}) \tag{4.4}$$

Within the ERG framework, a statistically validated monopartite projection can thus be obtained by employing a null model with properly defined constraints as a filter. As performed in previous studies employing the ERG framework [98], three null models have been considered in the present work: Bipartite Random Graph Model (BiRGM), Bipartite Partial Configuration Model (BiPCM) and Bipartite Configuration Model (BiCM). These null models (described in detail in Appendix B) respectively constrain the total number of edges E, the degrees of nodes belonging only to one layer and the degrees of nodes belonging to both layers $\mathbf{k}^*$.

In the case of user-hashtag bipartite networks, the final result is a semantic network where each couple of hashtags is linked if their number of common neighbors (i.e., the users tweeting or retweeting them) is statistically significant according to one of the aforementioned null models. After obtaining the three different monopartite projections, in the study of daily semantic networks reported in Chapter 6 the outcomes resulting from these models will be compared with the *naïve* projection. The use of four monopartite projections in the analysis of daily semantic networks lies within its scope: as in [98], the projections are compared in order to study the peculiar semantic behavior within discursive communities on a different filtering grained level. In fact, the null models retain a different amount of information inducing projections with a different level of detail: the BiRGM employs the same probability distribution to validate the similarity between any two nodes, thereby preferentially connecting high degree nodes; the BiPCM constraints the degrees of nodes belonging only to one layer thus representing the simplest benchmark which properly takes into account the heterogeneity of the features of the considered nodes; finally, the BiCM provides the strictest benchmark amongst the three aforementioned models constraining the degrees of nodes of both layers. In the case of users bipartite networks, the projection consists in a network where two verified accounts are connected if their number of common neighbors (i.e., the non-verified users retweeting their messages) is statistically significant according to the BiCM procedure. This procedure has been adopted also in the analysis of monthly semantic networks. The BiCM has been selected to detect only the most significant connections between verified users on which discursive communities are subsequently implemented. Due to the fact that the following algorithm is based solely on the retweeting action of non-verified users without any prior information on the system, it represents an unsupervised technique for detecting communities of verified users.

Within the ERG framework, the algorithm for obtaining monopartite projections from the original users and user-hashtag bipartite networks can be summarized by the following three steps.

Step 1: Measuring the degree of similarity between two nodes. Let $\mathbf{M}$ be the bipartite matrix whose dimensions are $N_\top \times N_\perp$ where $N_\top$ and $N_\perp$ represent the total number of nodes on the layer $\top$ and $\perp$, respectively. In order to quantify the similarity between two nodes α and β beloging to the same layer $\top$, the simplest method is to count the total number of common neighbors $V_{\alpha\beta}$ shared by these two nodes, also called *V-motifs*. In mathematical notation, the total number of V-motifs is computed by the following equation:

$$V_{\alpha\beta} = \sum_{j=1}^{N_\perp} m_{\alpha j} m_{\beta j} = \sum_{j=1}^{N_\perp} V_{\alpha\beta}^{j} \tag{4.5}$$

The term $V_{\alpha\beta}^{j}{=}m_{\alpha j}m_{\beta j}$ denotes the single V-motif defined by nodes α and β with node j belonging to the opposite layer. The value of $V_{\alpha\beta}^{j}$ is 1 if nodes α and β share the node j as a common neighbor and 0 otherwise. It is worth noting that the naïve projection of the bipartite network $\mathbf{G}^*$ corresponds to a monopartite network $\mathbf{A}$ whose entry is equal to $a_{\alpha\beta} \equiv \Theta[V_{\alpha\beta}]$. In other words, an edge between two nodes α and β is present in $\mathbf{A}$ in correspondence to any non-zero value of $V_{\alpha\beta}$. In Fig. 4.3, an illustration of the V-motif $V_{\alpha\beta}^{j}$ is shown for clarity purposes.

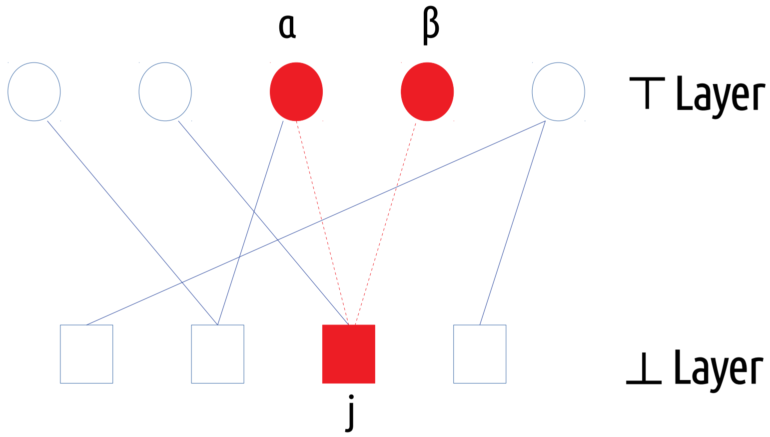


Figure 4.3: Visual representation of the V-motif $V^j_{\alpha\beta}$. The total number of V-motifs is employed to define the similarity between two nodes α and β. The statistical significance of this similarity is then quantified by adopting a null model with properly defined constraints.

Step 2: Quantifying the statistical significance of nodes similarity. The statistical significance of the similarity between two nodes α and β is quantified by adopting a null model with properly defined constraints. Each null model considered in the present work gives a connection probability $p_{\alpha j}$ for any couple of nodes of opposite layers. Since the edges in the ERG null-models adopted in this work are considered as indipendent variables, the presence of a V-motif $V^j_{\alpha\beta}$ can be described as the outcome of a Bernoulli trial obeying the following prescription:

$$\begin{aligned} f_{\mathrm{Ber}}(V^j_{\alpha\beta} = 0) &= 1 - p_{\alpha j} p_{\beta j} \\ f_{\mathrm{Ber}}(V^j_{\alpha\beta} = 1) &= p_{\alpha j} p_{\beta j} \end{aligned} \tag{4.6}$$

The independence of the edges implies that, once a couple of nodes α and β is chosen, the observed number of V-motif $V^*_{\alpha\beta}$ can be described as the sum of independent Bernoulli trials, each characterized by a distinct probability. The distribution describing the behavior of this random variable is called *Poisson-Binomial*, a generalization of the Binomial distribution when the single Bernoulli trials are characterized by different probabilities.

The computation of the statistical significance of the similarity measure between two nodes α and β consists in calculating the p-value of the aforementioned Poisson-Binomial distribution, i.e. the probability of observing a number of V-motifs greater than, or equal to, the observed value $V^*_{\alpha\beta}$:

$$\text{p-value}(V^*_{\alpha\beta}) = \sum_{V_{\alpha\beta} \geq V^*_{\alpha\beta}} f(V_{\alpha\beta}) = \sum_{V_{\alpha\beta}=V^*_{\alpha\beta}}^{N_\perp} \sum_{C_n} \left[\prod_{c_i \in C_n} p_{\alpha c_i} p_{\beta c_i} \prod_{c_i' \notin C_n} (1 - p_{\alpha c_i'} p_{\beta c_i'}) \right] \tag{4.7}$$

where C_n is the set indicating all the possible n-tuples of nodes. Eq. 4.7 can be interpreted as the *Survival Distribution Function* (SDF) of the Poisson-Binomial distribution corresponding to the observed occurence of V-motifs $V^*_{\alpha\beta}$ which can be easily calculated by adopting the approach proposed in [98, 110]. For each pair of nodes α and β, the procedure of computing the SDF is repeated until a $N_\top \times N_\top$ matrix of p-values is obtained.

Step 3: Validating the monopartite projection. Once the statistical significance of the similarity between each pair of nodes α and β is quantified by computing the corresponding p-value, only those pairs whose p-value is statistically significant must be retained. To this aim, a validation procedure for testing simultaneously multiple hypotheses is needed. The choice of this work has been directed towards the False Discovery Rate (FDR) procedure. As shown in previous works [98], in contrast to other validation procedures, the FDR prescription is a suitable strategy for this analysis since it allows for monitoring the expected number of incorrectly rejected null hypotheses, i.e. the incorrectly validated edges, regardless of the independence of the hypoteses tested. Indeed, it is worth noting that the measured p-values are not independent since each observed edge can be responsible for the similarity of multiple nodes. The FDR procedure traslates into sorting in increasing order the m p-values (where $m = N_\top(N_\top - 1)/2$ is the total number of p-values in the monopartite projection $\mathbf{A}$) p-value$_1 \leq \ldots \leq$p-value$_m$ and then identifying the largest integer $\hat{i}$ which satisfies the condition:

$$\text{p-value}_{\hat{i}} \leq \frac{\hat{i}t}{m} \tag{4.8}$$

where t represents the single-test significance level. In this work, the significance level for each single-test is equal to t=0.01. In other words, only the pair of nodes α and β satisfying the condition p-value$_{V^*_{\alpha\beta}} \leq$ p-value$_{\hat{i}}$ is considered as having a significantly similar number of common neighbors. Eventually, each couple of nodes whose corresponding p-value is validated by the FDR procedure is linked in the resulting monopartite projection $\mathbf{A}$.

4.2.2 Community detection in monopartite projections

Once the procedure to obtain statistically validated projections is applied, the nodes in the final networks can be clustered in groups of verified users/hashtags who are interacting more among themselves than with the rest of the network. Amongst the community detection algorithms [111], the chosen procedure to identify communities of verified users/hashtags is based on searching for the optimal partition according to a modularity function Q. The modularity optimization algorithms are one of the most used clustering techniques in empirical applications [111]. Within this class of algorithms, a reshuffled version of the Louvain algorithm [112] is employed in this work. In fact, the Louvain procedure has been observed to be one of the best performing community detection algorithms [113] and one of the most used clustering technique in several applications [111]. More specifically, as pointed out in [5], the reshuffled version adopted in this work has another advantage, namely its capacity to overcome the original order dependence of the algorithm presented in [111]. Further details on this detection algorithm are described in Appendix C.

4.3 Label propagation procedure

The identification of users' discursive communities requires a further step. This additional step is represented by the assignment of non-verified users to the communities of verified users which are obtained through the procedure described in the previous Section. To this aim, the full retweeting network comprising all Twitter users involved in the discussions throughout the entire observation period $\mathcal{T}$ is considered. The information about the community affiliation of verified users is preserved in this network via a *community label*. A measure of the distribution of the community label of its neighbors is then assigned to each non-verified user. This measure is called the *polarization index* and determines the balance of the interactions of each non-verified user in the retweeting network, i.e. how the portion of interactions of non-verified users is distributed towards each community of verified users.

Let C_c (where c represents the community label) be the set of users belonging to community c with whom the non-verified user α has interacted, and let N_α be the set of neighbors belonging to one of the community c who had interactions with node α during the observation period $\mathcal{T}$. The *polarization index* for each non-verified user α is defined as:

$$\rho_\alpha = \max_c \{I_{\alpha c}\} \tag{4.9}$$

where:

$$I_{\alpha c} = \frac{|C_c \cap N_\alpha|}{|N_\alpha|}. \tag{4.10}$$

As shown in [5], the polarization index reveals the level of unbalance of the distribution of retweets made by a single user, thus providing a clear indication of the target community of each non-verified user. *Polarized users* are defined as those non-verified users with a value of the polarization index greater than, or equal to, 0.9. All polarized users are included within the corresponding communities of verified users identified via the Louvain procedure. This step represents a first stage for building insightful discursive communities: in fact, the percentage of polarized users is only ≈8% and ≈15% of the total number of users in the retweeting network of the 2018 Italian general elections and of the 2019 Italian migration discussion data sets, respectively.

The affiliation of the remaining non-verified users is inferred in the retweeting network by *propagating* the initial community labels assigned to verified users. A label propagation algorithm, as proposed in [114], has been run on the retweeting network: this algorithm implements the idea that each node in the retweeting network joins the same community the majority of its neighbors belongs to. Labels assigned to verified and polarized users propagate through the network until densely connected groups of nodes reach consensus. In case no majority is found, the algorithm randomly removes an edge and re-evaluates the labels. Due to the randomness introduced by this latter step, the label propagation algorithm is repeated 1.000 times. In fact, although this algorithm can modify the label previously assigned to a non-verified user, it has been observed to remain stable after a large number of runs [5]. At the end of this process, each non-verified user is assigned the community label held most frequently across all iterations. After this last step, more than 90% of the total number of Twitter users included in the initial retweeting networks is assigned to one of the main discursive communities in both data sets.

4.4 Final remarks and limitations

The methodological approach adopted to infer discursive communities and semantic networks, already applied in other works [4–7], is innovative with respect to the state-of-the-art of socio-semantic analyses in a twofold way. First, it does not require any prior information either based on users features or sharing patterns (i.e. hashtags or media contents), unlike other studies that involve a manual labelling of at least one of these items [39, 61]. Second, the political valence of online actions, such as retweeting, is behavior-driven rather than ideologically determined as it is based on the direct action performed by non-verified users to re-broadcast verified users' messages via retweets. As such, after properly defining a list of anchor keywords, the present algorithm is applicable to *any* Twitter discussion. Although the selected case studies concern two Italian Twittersphere discussions that took place in close periods, the present methodological approach can be adapted to study other Twitter conversations. In fact, discursive communities can be identified regardless of country, topic, or language of that specific discussion and without the need for any previous knowledge about the political actors or relations between them.

The only limitation in the formation of discursive communities along these lines is a technological constraint which is, however, inextricably tied to social media analysis: the identities of single users on social media are intrinsically nested and multidimensional [67]. In this sense, there are no exact procedures to investigate users' political affiliation, without complete information (i.e. all the Twitter timeline and/or through a survey of his/her political orientation towards a specific topic). Twitter data completely lack information about the actual political affiliation of users and this can be a severe limitation in the prediction of relevant political outcomes such as voter turnout, election results, or the formation of political alliances between parties. Nevertheless, a plethora of studies based on Twitter freely available data has shown that it is indeed possible to infer the political orientation of users from tweets and to analyze electoral debates and societal discussions [5–7, 12, 69, 115] shedding light on the political implications of non-traditional political acts such as public expressions on social media. Another remark regarding the discursive communities is that it is widely recognized that Twitter users are not representative of broader populations (in this case, of all Italian citizens) and that Twitter Search API does not guarantee the representativeness of the data themselves [85]. Without claiming to provide a method to reach representativeness, the present methodological approach affords nonetheless an efficient unsupervised technique based on a solid theoretical framework that is able to infer Twitter users' political orientation starting only from interactional information provided by the freely available API's data.

With respect to the study of semantic networks, the following analysis remains a partial investigation of the symbolic universe that is produced and circulated online in conjunction with relevant political events and dynamics. The semantic structures investigated in this work are formed in the space created by a single platform and pivot around the use of a specific feature for marking contents - i.e. the hashtag. In fact, ad-hoc publics that assemble around topics are not exhausted by communities that form on particular social media platforms, let alone around specific hashtags [116]. Moreover, looking at conversations that form around specific hashtags fails to include those contributions that, albeit pertinent, are delivered without including any specific content marker [117]. Nonetheless, without any claim of exhaustivity, the present mapping of two Italian Twittersphere discussions provides a useful entry point for consideration of the online formation of political collective identities. Users employing specific keywords and hashtags in their tweets did in fact contribute to framing the public debates along certain lines and they did so upon a platform that was not only used by a large segment of Italian population [86] but also plays a pivotal political communication role, exerting a regular effect of agenda setting on the national mainstream media [15, 75, 118].

Chapter 5

Networked partisanship: Mapping political discursive communities

The present Chapter focuses on the analysis of *networked partisanship* conceived as online discursive processes through which attention is directed towards specific political actors. Given the multidimensional nature of polarization dynamics, networked partisanship is investigated starting from the identification of *discursive communities* that assemble around the use of hashtags related to migration and general elections discussions. Discursive communities are here induced by the common retweeting activity of Twitter users. Their topological study is linked to main events *on the ground* according to a longitudinal perspective of hybrid interactional dynamics between online and offline spaces. In this Chapter, discursive communities are obtained by applying the procedure described in Chapter 4 to the data sets described in Chapter 3. After a description of the structural features of each discursive community, their users are classified according to their retweeting behavior and on the basis of the semantic polarization occurring inside each discursive community.

The following Sections report the main results on the in-depth analysis of the discursive communities extrapolated in correspondence to the two considered Twitter discussions. In Section 5.1, their internal structure will be described along with the retweeting and mentioning activities in order to gain insights into the collective identities of discursive communities. Section 5.2 analyses the users in terms of a characterization which focus on their role in information spreading. The categorization in user types already employed in *González-Bailón et al.* [69] on the *Indignados* movement in Spain will be here used to expand the operational definition of discursive communities. Finally, Section 5.3 describes the *semantic polarization* occuring within each discursive community through a measure of the overrepresentation of specific hashtags. This last evidence further enhances the aspects related to topics and keywords of the discursive communities in a perspective of semantic uniqueness. Most of the results presented in this Chapter have already been published [4, 6].

5.1 Composition and activity of the discursive communities

As mentioned before, networked partisanship is explored starting from the identification of *discursive communities* which are *groups of Twitter users who share similar retweeting patterns* where the concept of similarity is based on the statistical significance of the number of retweets with respect to a benchmark model. In this sense, discursive communities are here considered as coherent groups of Twitter users (both verified and non-verified) assembled around the use of specific keywords and induced by the retweeting activity of non-verified users.

5.1.1 Verified users affiliation

For both Twitter data sets, the identified communities closely resemble the Italian political coalitions. This is the first evidence of their consistency in terms of political and discursive affinity reflecting the political fragmentation of the Italian political context. The complete lists of verified users within discursive communities are reported in Appendix A for both data sets. In the data set on the 2018 Italian general elections, five main communities are identified as follows:

- **Center-right and far-right (CDX)**: a community of users composed of accounts of allied right-wing political parties (e.g. *@forza_italia*, *@LegaSalvini*), center and right-wing politicians (e.g. *@renatobrunetta*, *@matteosalvinimi*), their institutional representative groups (e.g. *@GruppoFICamera*), and users interacting with all of them;
- **CSX**: a rather heterogeneous community of users composed of accounts of political parties forming the center-left alliance (e.g. *@pdnetwork*, *Radicali*), their politicians (e.g. *riccardomagi*, *@matteorenzi*), some journalists (e.g. *@vittoriozucconi*, *@jacopo_iacoboni*), and users engaging with them;
- **'Movimento 5 Stelle' (M5S)**: a community composed of accounts of politicians belonging to the Italian populist party 'Movimento 5 Stelle' (e.g. *@DaniloToninelli*, *@luigidimaio*), institutional accounts of the party (e.g. *@M5S_Camera*, *@M5S_Senato*) and users engaging with all of them.
- **'Liberi e Uguali' (LEU)**: a community of users composed of politicians belonging to 'Liberi e Uguali' (e.g. *@PietroGrasso*, *@lauraboldrini*), institutional accounts of the party (e.g. *@liberi_uguali*, *@SI_sinistra*) and users engaging with all of them;
- **Media organizations (MEDIA)**: a community not strictly party-related and composed of a rather heterogeneous set of media organizations, mostly affiliated with the Vatican (e.g. *@Avvenire_NEI*, *@Rvaticanaitalia*), television shows (*@chetempochefa*),

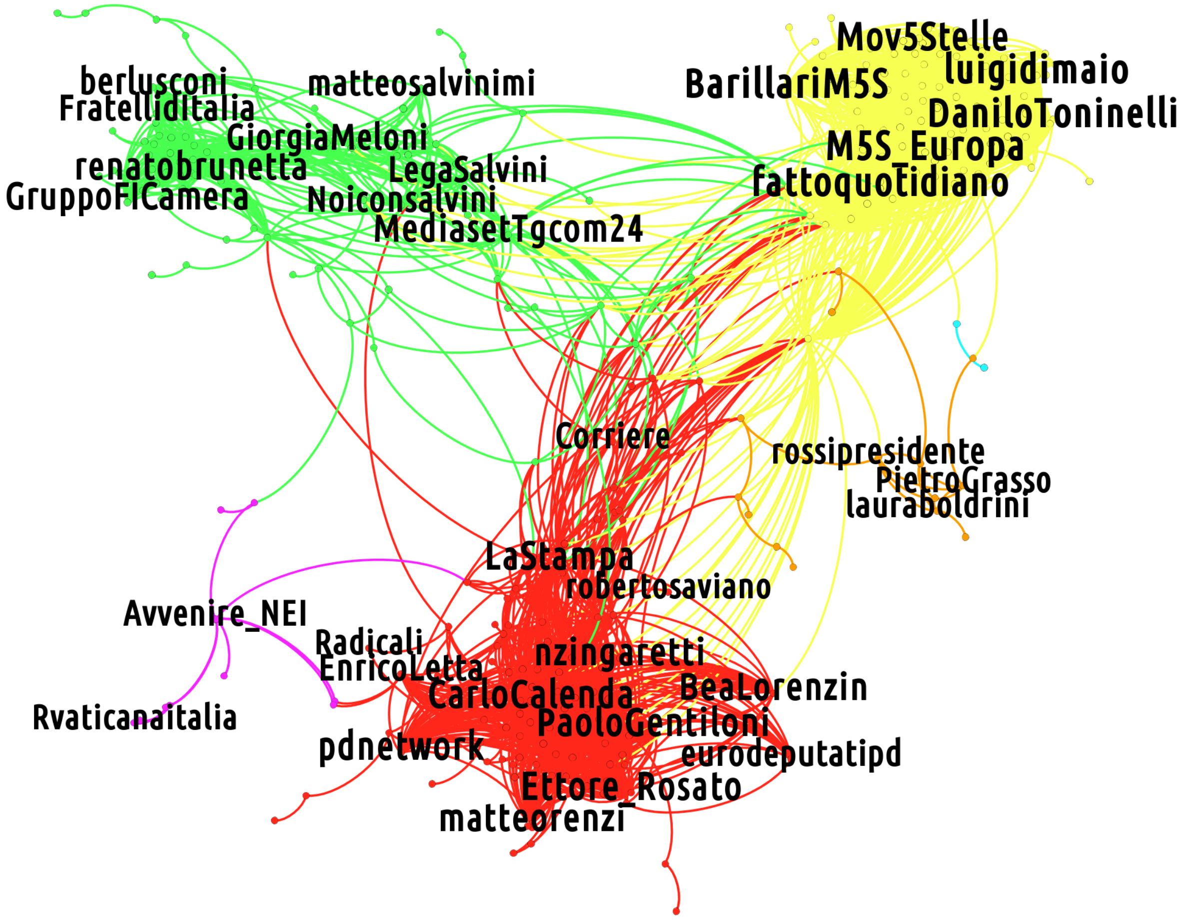


Figure 5.1: Representation of the network of verified user communities in the 2018 data set. The final communities pivot around verified accounts of the main Italian political parties/coalitions and politicians as well as around accounts of media, intergovernmental and non-governmental organizations. Both data sets show three main discursive communities colored respectively in red, green and yellow (respectively, the center-left wing, the right-wing and the 'Movimento 5 Stelle' communities) reflecting the tripolar competition of the institutional political parties within the Italian political context.

journalists (e.g. *@BentivogliMarco*, *@francescoseghez*) and users engaging with all of them.

Similarly, in the data set on 2019 Italian migration discussion, five discursive communities are labeled as follows:

- **Far-right (DX)**: a community revolving around official accounts of right-wing political parties such as 'Fratelli d'Italia' and the 'Lega' (i.e. *@FratellidItalia*, *@LegaSalvini*), their leaders (i.e. *@GiorgiaMeloni*, *@matteosalvinimi*), politicians and journalists working for right-wing national newspapers (e.g. *@NicolaPorro*);

- **Center-left (CSX)**: a community built around official accounts of the political parties composing the center-left alliance such as the 'Partito Democratico' and 'Italia Viva' (i.e. *@pdnetwork*, *@ItaliaViva*), their leaders (i.e. *@nzingaretti*, *@matteorenzi*), writers and journalists working for left-wing national magazines and newspapers (e.g. *@eziomauro*, *@robertosaviano*);

- **'Movimento 5 Stelle' (M5S)**: a community pivoting around accounts related to the 'Movimento Cinque Stelle' and including its leaders (e.g. *@luigidimaio*, *@Roberto_Fico*) and institutional accounts (e.g. *@M5S_Camera*, *@M5S_Senato*) but also the national newspaper 'Il Fatto Quotidiano' and the journalists working for it (e.g. *@fattoquotidiano*, *@petergomezblog*). Notably, the Twitter account of the Italian Prime Minister Giuseppe Conte (proposed by the 'Movimento Cinque Stelle' to overcome the impasse in forming a new government after the 2018 national elections) is included in this community;

- **'Forza Italia' (FI)**: a smaller community revolving around the official accounts of prominent members of the FI party, the party founded by Silvio Berlusconi (e.g. *@berlusconi*, *@gabrigiammarco*, *@forzaitalia*);

- **Media, international governmental and non-governmental organizations (MINGOs)**: a fifth community not strictly party-related and emerging around a variety of verified accounts connected to three main sets of actors:

    1. accounts of weekly national magazines such as 'L'Espresso' (*@espressoline*), online newspapers like the Italian 'Huffington Post' or 'Il Post' (*@huffpostitalia*, *@ilpost*), television shows of investigative journalistic and documentary nature (*@reportrai3*), journalists covering foreign affairs and migrations (e.g. *@martaserafini*, *@mannocchia*);

    2. accounts of non-governmental organizations aimed at defending human rights (e.g. *@amnestyitalia*) or specialized on migration and international cooperation issues (e.g. *@emergency_ong*, *@ActionAidItalia*), accounts of prominent activists in this domain (such as Regina Catrambon, the initiator of a search-and-rescue NGO called 'Migrant Offshore Aid Station', or Carola Rackete, the captain of Sea-Watch 3, the ship entering the Italian port of Lampedusa in spite of the opposition of the back-then Minister of the Internal Affairs Matteo Salvini);

    3. accounts of international governmental organizations such as the Italian branches of UNICEF and the International Organization for Migrations, but also the United Nations Refugee Agency.

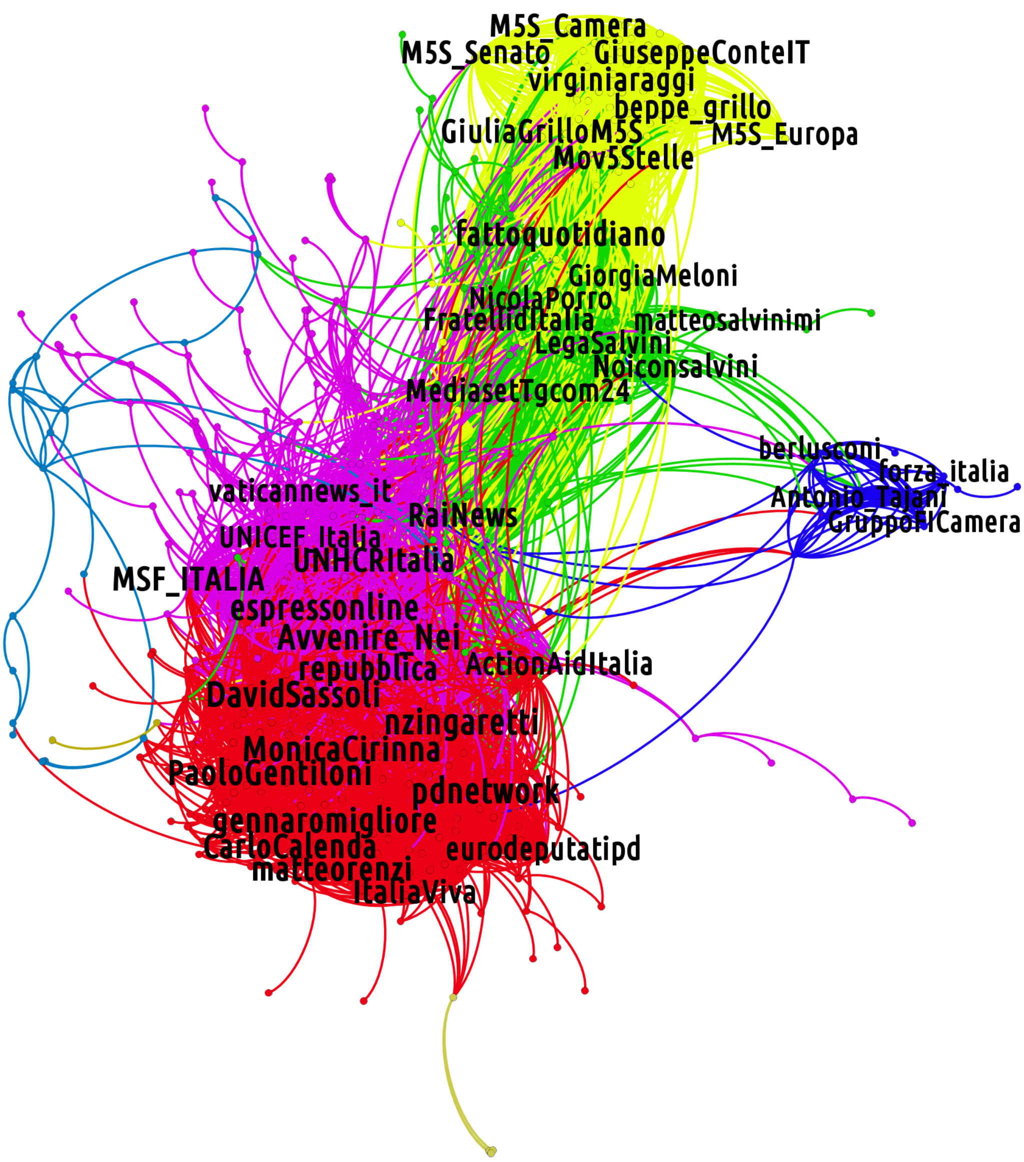


Figure 5.2: Representation of the network of verified user communities in the 2019 data set. The final communities pivot around verified accounts of the main Italian political parties/coalitions and politicians as well as around accounts of media, intergovernmental and non-governmental organizations. Both data sets show three main discursive communities colored respectively in red, green and yellow (respectively, the center-left wing, the right-wing and the 'Movimento 5 Stelle' communities) reflecting the tripolar competition of the institutional political parties within the Italian political context.

Figs. 5.1 and 5.2 display the communities of verified users found in the two data sets. Notably, both networks reveal the presence of three discursive communities retracing the tripolar structure of the Italian institutional political scenario. These communities are populated by verified accounts belonging to politicians or parties of the center-left wing, the right-wing and the 'Movimento 5 Stelle' (colored in red, green and yellow, respectively). Moreover, in both data sets, two additional major communities of verified users characterize the final composition of the largest connected component of users. In the data set of the 2018 Italian general elections, a community of verified accounts of politicians belonging to LEU and another community of accounts of news media organization and journalists are present (respectively, colored in orange and in purple in Fig. 5.1). In the data set of the 2019 Italian migration discussion, a community of verified accounts belonging to ‘Forza Italia’ and another group of accounts of news media organizations, journalists and NGOs are present (respectively, colored in blue and in purple in Fig. 5.2).

As shown by the qualitative description of these communities, the adopted procedure allows for the detection of political coalitions through the lens of conversational dynamics. Regardless of the specific discussion topic, this has an effect also on the internal structure of discursive communities. For instance, the verified accounts belonging to ‘Forza Italia’ have been detected as part of the center and far-right political coalition in the run up to the 2018 general elections. Interestingly, applying the Louvain algorithm within the CDX discursive community leads to the recovery of these accounts as a separate sub-community whose composition is similar to the FI community detected in Fig. 5.2. Retrospectively, the FI community can be interpreted as an early signal of a more profound political fragmentation inside the initial CDX discursive community. In fact, within this community, the far-right component, consisting of accounts belonging to ‘Fratelli d’Italia’ and the ‘Lega’ (i.e., the main bulk of the accounts within the DX discursive community), has taken on a more prominent role in the public discourse inside Twitter, as already pointed out in Section 3.1.2.

Looking at the internal composition of discursive communities provides a better description of both processes of inter-group distancing and intra-groups consolidation regardless of formal political alliances. In this sense, tracing discursive communities represents a mechanism for detaching networked partisanship from formal political alliances. This is particularly evident in the separation within the 2019 data set between the community of the ‘Movimento 5 Stelle’ and that of the ‘Lega’ also during periods in which the two parties jointly shared seats within the first Conte government: as shown in Fig. 5.3, the two communities do merge only under exceptional circumstances that affect more directly the governmental alliance (i.e., during the Sea-Watch 3 crisis) but, in fact, their discursive coalition is exceptional and just temporary. Analogously, verified users belonging to ‘Fratelli d’Italia’ and the ‘Lega’ are indistinguishable as they are part of the same discursive community.

However, networked partisanship cannot be thought of in isolation from political dynamics on the ground, as it is well demonstrated by the fracture within the left-wing community. While the LEU community disappears in the 2019 data set, two accounts belonging to former members of ‘Liberi e Uguali’ (i.e. *@lauraboldrini* and *@rossipresidente* who are respectively owned by the center-left politicians Laura Boldrini and Enrico Rossi) move within the CSX community. In fact, after an electoral competition in which ‘Liberi e Uguali’ and the ‘Partito

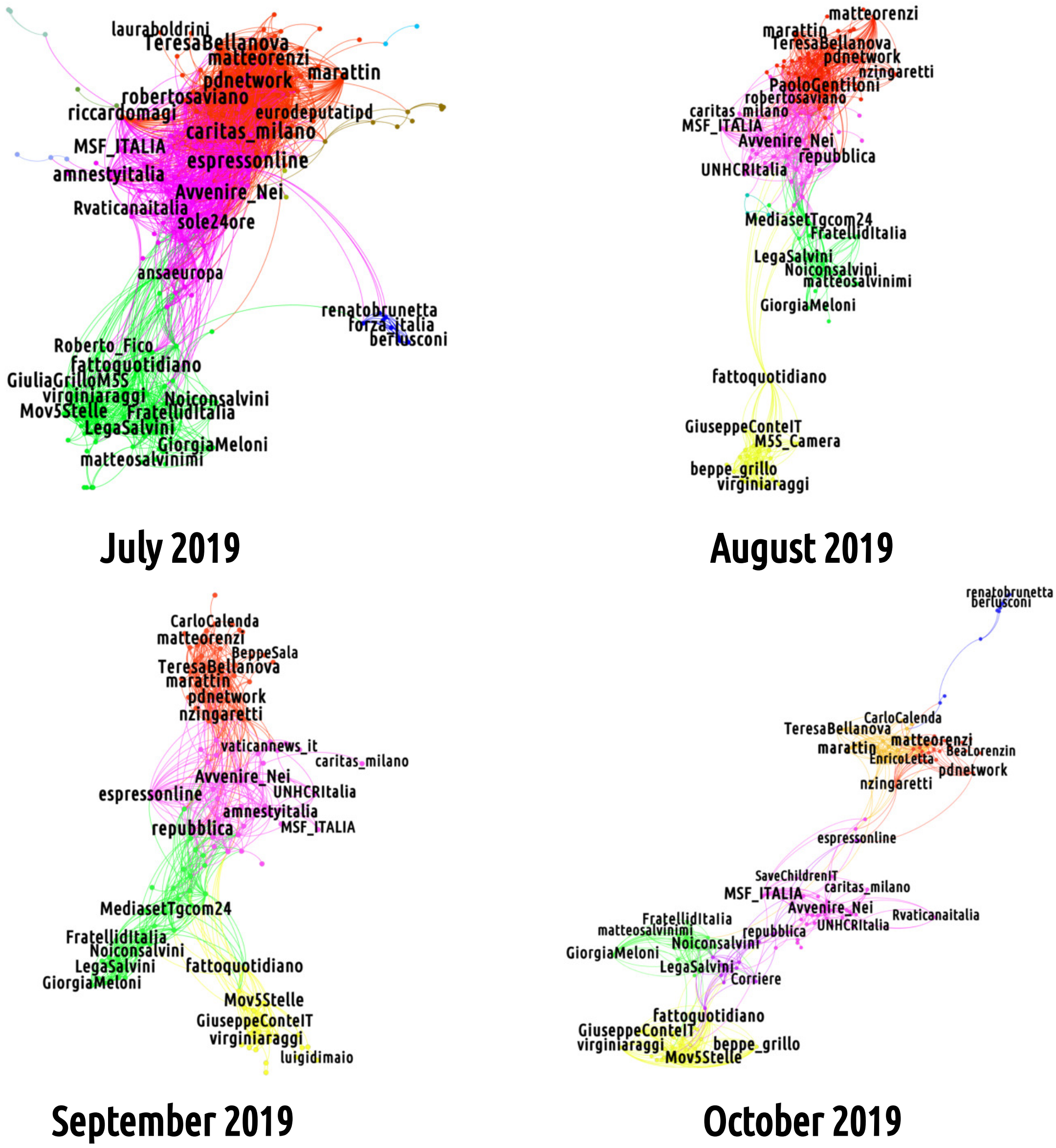


Figure 5.3: Representation of the evolution from July to October 2019 of verified user communities within monthly networks in the 2019 data set. The discursive communities evolution reflects a more profound political dynamics occuring within the institutional parties: while M5S and DX communities appear as merged only during the 'Sea Watch 3' crisis in July, a substantial portion of politicians of the 'Partito Democratico' (united in the red cluster in September) splits in October 2019 into an additional sub-community. This new community is induced by the Twitter activity of the members of a new center-left political formation (i.e. the 'Italia Viva' party, in orange).

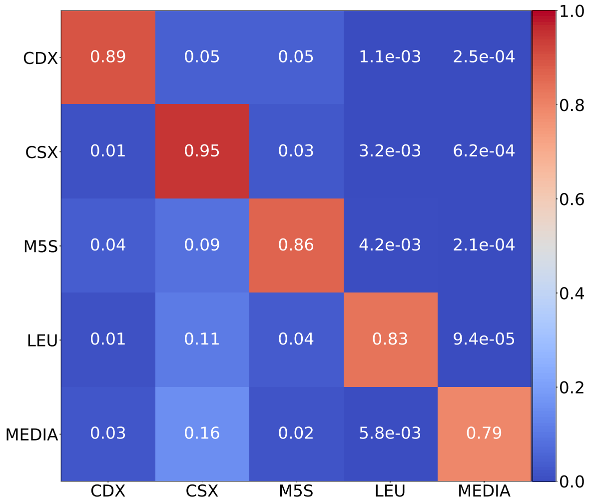
Democratico' run separately, in 2019 both politicians formalized their return to the 'Partito Democratico'. This is a clear signal that, in both cases, political orientation of party candidates, politicians and public figures can be traced back through the membership to the corresponding discursive communities.

The evolution of networked partisanship over time can be studied at a finer grained degree by selecting a temporal window and partitioning the data set accordingly. To this aim, the tweets of the 2019 Italian migration discussion data set are aggregated on a monthly basis. Thus, a set of seven monthly bipartite networks are built by considering the retweeting activity of non-verified users across this limited amount of time. The evolution of the monthly networks of verified user communities between July and October is shown in Fig. 5.3. A closer inspection of monthly attention flows suggests that the evolution of the discursive alliances mostly within (but also across) discursive communities match that of the 'offline' political dynamics. Two examples well illustrate this point. As already noticed, albeit sharing seats in the Italian government, the 'Movimento 5 Stelle' and the 'Lega' are supported online by two systematically separated discursive communities even when the analysis is performed on a monthly basis. While this misalignment reflects the existence of a fracture within the governmental coalition, networked partisanship dynamics evolve fluidly and, in some occasion, witness the effort of overcoming disagreements internal to the governmental coalition. In fact, in correspondence to July 2019, as shown in Fig. 5.3, the M5S and DX communities merge after the entrance, without permission, of the rescue boat Sea-Watch 3 into the Italian territorial waters and the arrest of its captain, Carola Rackete (see Chapter 3). This event, occurred at the end of June 2019, inflamed the political debate and forced governmental parties to cooperate on (at least formally) the Sea-Watch 3 crisis by making a common effort to retweet the messages from the accounts of the members of the government. This, in turn, causes the transient formation of a single discursive community which gathers accounts supporting the government line.

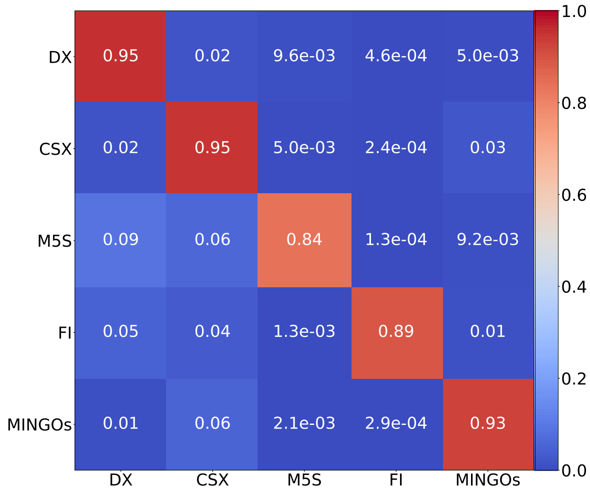
Similarly, the 2019 Italian government crisis between August and September 2019 triggers a fracture in the CSX community which reflects the internal break of the 'Partito Democratico'. In particular, some users change their retweeting behavior, revealing the birth of a new discursive community (in orange in Fig. 5.3) pivoting around accounts connected with 'Italia Viva', the party founded by former Prime Minister Matteo Renzi in overt opposition to the 'Partito Democratico'.

5.1.2 Structural network properties

The polarization index $\langle \rho_\alpha \rangle$ defined in Eq. 4.9 has been employed in order to measure the embeddedness of both verified and non-verified users within each discursive community. From this measure, the mean portion of interactions within and outside each community is extracted to compare the number of edges between (i.e., inter-communities) and within communities (i.e., intra-communities). Results referred to both data sets are visualized as heat maps in Fig. 5.4 where the matrix of the mean polarization indices for all discursive communities is reported. Each mean polarization index is obtained by averaging the polarization index value over all the users belonging to the community on the x-axis. Within each row of the heat map, this parameter is computed by considering only the set of neighbors belonging to the community on the y-axis. Fig. 5.4 demonstrates that the fraction



(a) 2018 Italian general elections



(b) 2019 Italian migration discussion

Figure 5.4: Heat map of the mean polarization index $\langle \rho_\alpha \rangle$ values. Both data sets show that the fraction of interactions within each discursive community is significantly higher than those outside.

of intra-communities edges is larger than the fraction of inter-communities edges for all the discursive communities. This result highlights the tendencies of these communities to be highly polarized as an extremely high portion of the neighbors of their members tends to belong to the same community of the members themselves, as already observed in [5, 7].

In Tab. 5.1, some topological properties of the major discursive communities are shown after having retrieved the corresponding sub-networks in the full retweeting network. Each discursive community is characterized by structural peculiarities, revealing different levels of activity. In terms of their number of users, the CSX community is the most populated in both data sets while the least populated are the MEDIA community in the 2018 data set and the FI community in the 2019 data set. An interesting phenomenon is the variation of the percentage of users in each community across the two case studies. While the portion of CSX users remains nearly stable, the size of the CDX discursive community increases significantly in the 2019 data set, even though the former sub-community of accounts belonging to FI is no longer part of it. The remaining communities include a relatively low percentage of users except for the MINGOs community that represents the third greatest community in the 2019 data set. In fact, the users in the M5S discursive community are reduced between the two data sets since in the 2019 data set their number is significantly lower than the CSX and the CDX communities. In this context, a different quantity of the users within discursive communities and their variation throughout the two observation periods can be considered as a direct signal of different degrees of attention and mobilization inside each community.

Regarding the other statistical properties of the sub-networks, it is clearly visible how, in the 2018 data set, the CSX and the CDX communities display opposite behaviors: even if

2018 Italian general Elections							
Discursive community	N_u (%)	N_e (%)	$\langle k \rangle$	$\langle k \rangle_N$	$\langle \rho_\alpha \rangle$	μ_r	μ_m
CSX	83230 (64%)	302315 (44.7%)	7.3	$6.5 \cdot 10^{-4}$	0.95	0.88	0.69
CDX	22811 (17.5%)	167962 (24.8%)	14.7	$3 \cdot 10^{-3}$	0.89	0.84	0.58
M5S	22269 (17.1%)	199156 (29.5%)	17.8	$2 \cdot 10^{-3}$	0.86	0.84	0.35
LEU	1527 (1.2%)	6119 (0.9%)	8	$1 \cdot 10^{-2}$	0.83	0.72	0.57
MEDIA	201 (0.2%)	260 (0.1%)	5.2	$3 \cdot 10^{-2}$	0.79	0.68	0.21
2019 Italian migration discussion							
CSX	112459 (52.4%)	1080271 (32.4%)	19.2	$9 \cdot 10^{-4}$	0.95	0.9	0.47
DX	70061 (32.7%)	2075013 (62.3%)	59.2	$4 \cdot 10^{-4}$	0.95	0.93	0.52
M5S	6313 (2.9%)	107100 (3.2%)	33.9	$7 \cdot 10^{-3}$	0.84	0.73	0.39
FI	598 (0.3%)	3133 (0.1%)	10.5	$1 \cdot 10^{-2}$	0.89	0.77	0.34
MINGOs	24994 (11.7%)	65388 (2%)	5.2	$6 \cdot 10^{-4}$	0.93	0.78	0.47

Table 5.1: Structural characteristics of the five discursive communities in both data sets. In both data sets, discursive communities show distinct structural characteristics (number/percentage of users N_u, number/percentage of edges N_e, mean degree $\langle k \rangle$ and normalized mean degree $\langle k \rangle_N$) but exhibit a rather similar communicative behavior in terms of the polarization index ρ_α and the self-reference indeces for retweeting (μ_r) and mentioning (μ_m).

the number of users N_u in the CSX communities is greater, the average degree of the users within CDX is nearly double compared to that of the CSX community. In other words, a CDX user is, on average, twice as connected with the rest of his/her community as a CSX user is. The same behavior is observable in the 2019 data set where the average degree of DX users is nearly 3 times that in the CSX community. In both data sets, M5S accounts are also particularly active, even though in the 2019 data set the most active users are those in the DX community [119].

For the sake of completeness, Tab. 5.1 reports the information regarding the normalized mean degree, i.e.

$$\langle k \rangle_N = \left\langle \frac{k - \min\{k\}}{\max\{k\} - \min\{k\}} \right\rangle = \frac{\langle k \rangle - \min\{k\}}{\max\{k\} - \min\{k\}} \tag{5.1}$$

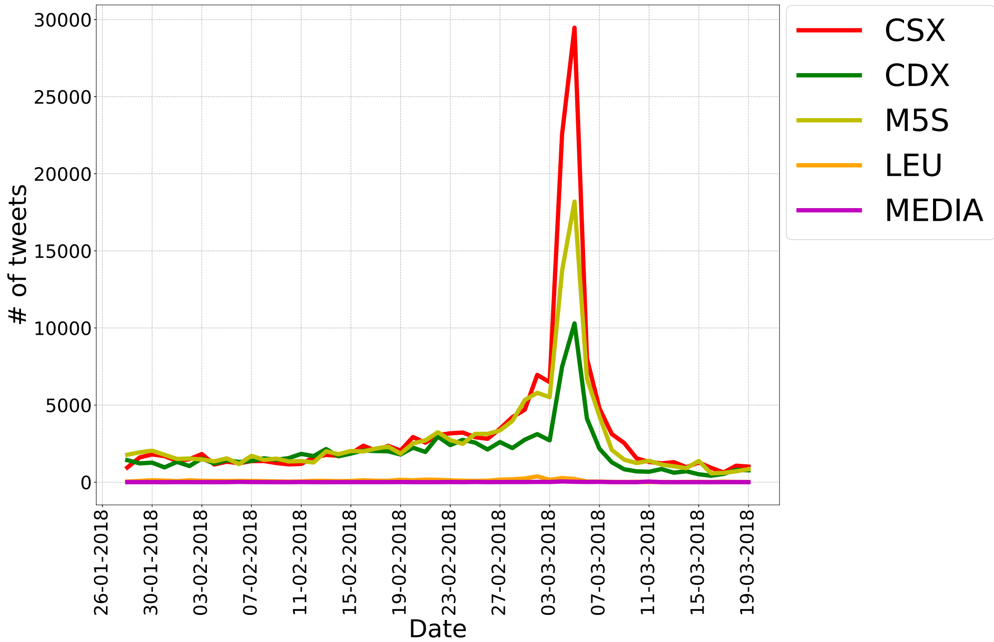
Since taking the average of normalized degrees is equivalent to normalizing the mean degree itself, the behavior of $\langle k \rangle_N$ also provides information about the range of variation of degrees. As Table 5.1 reveals, the degrees of DX and CDX users are, overall, more similar (i.e. their range of variation is smaller) than the user degrees of the other communities.

Looking more closely at the type of interactions sustained within the discursive communities, a rather regular pattern seems to emerge: all communities endorse the same communicative behavior as their users tend to employ retweets to broadcast opinions and contents generated by their own community members while mentions are used to establish an indirect contact with users in other communities. This element confirms the results of previous studies based on Twitter data[61, 120] and clearly emerges by looking at the self-reference indeces μ_r and μ_m. These two indeces are calculated as the ratio of the number of retweets (μ_r) and mentions (μ_m) of the users belonging to a given community over the total number of retweets, or mentions, performed within that same community. In both data sets, values of these two parameters are included within two different ranges: while μ_r assumes values between $\approx$0.7 and $\approx$0.9, μ_m has its minimum and maximum equal to 0.21 and 0.69 in correspondence to the 2018 data set respectively in the MEDIA and the CSX communities. This phenomenon can be explained as follows: the self-reference index μ_r shows that retweets tend to be employed to re-broadcast contents produced 'internally', thus endorsing the formation of partisan collective identities. At the same time, the values of μ_m reveal that, albeit polarized, these communities are characterized by non-trivial levels of inter-activity. Against this common background, however, there is room for topological variation. A deeper cleavage seems to separate the two largest communities in both data sets from the rest of the discussion: in fact, the CSX and CDX/DX communities show the highest polarization $\langle\rho_\alpha\rangle$ and self-reference communicative behaviors μ_r and μ_m. Conversely, the other three communities show a higher tendency to broadcast internally also contents produced elsewhere.

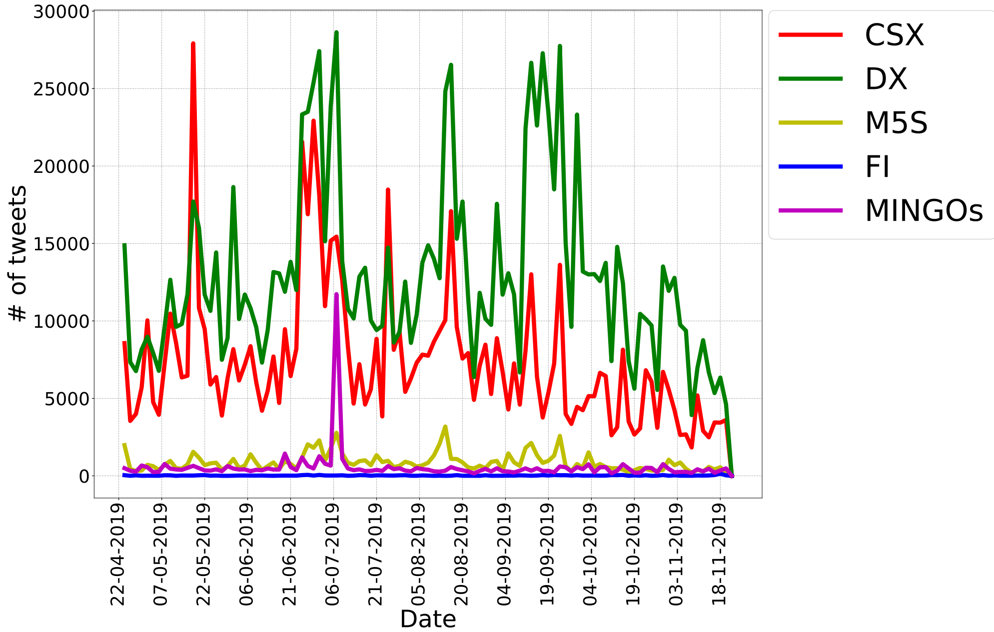
5.1.3 Activity patterns

A first analysis of the activity within the discursive communities consists of observing the evolution of the volume of tweets across the two data acquisition periods. As shown in Fig. 5.5, this evolution presents totally different behavior in the two data sets. In the data set about the 2018 general elections, the number of tweets is characterized by a similar trend for all the five discursive communities, namely a steady value of the volume of tweets followed by a steep rise a few days before Election Day. A peak in the tweeting activity is then registered in correspondence to the day after Election Day, i.e. March 5^{th}, immediately followed by a rapid decrease of the number of tweets. This common feature signals a nearly identical approach towards the tweeting activity across the entire observation period: the interest in the electoral process is indeed greater as Election day is approaching and it abruptly drops immediately after. The main difference between discursive communities is represented by the size of the peak mainly due to a disproportionate number of users who are part of the CSX discursive community and a significantly lower amount of accounts in the LEU and MEDIA communities.

Conversely, in the data set about the 2019 Italian migration discussion, levels of activity within the five communities are marked by weekly oscillations and are mainly driven by events taking place on the same time scale and driving the Twitter discussion. In fact, peaks of tweeting activity are observed in correspondence to the 2019 European elections at the end of May 2019, the Sea-Watch 3 episode from the end of June 2019 until the beginning of July 2019 and the Italian government crisis between the end of August and the beginning of September 2019. Unlike the first data set, the DX discursive community shows a



(a) 2018 Italian general elections



(b) 2019 Italian migration discussion

Figure 5.5: Volume of tweets published by users of each main discursive community. While in the 2018 data set the volume follows the same trend for all five discursive communities, in the 2019 data set the evolution of the number of tweets present different patterns highlighting the diverse activity behavior of each discursive community.

systematically greater amount of tweets, except for an isolated peak of the CSX discursive community. This result is consistent with the outcome of another study on the behavior of Twitter user communities in correspondence to the societal discussion about the COVID-19 pandemic [119] and with a greater attention towards this topic within the DX discursive community and, more in general, in far-right parties [121], as described in Section 3.1.2.

In order to gain more insights into the activity patterns within each discursive community, the diverse behavior of these communities can be quantified to trace the attention flows within and between communities. To do that, mentions and retweets are detected for all tweets posted by their members and the most mentioned and retweeted accounts are identified. The ranking of the five most retweeted and mentioned verified user reveals interesting differences across groups giving a hint of the partisan alignment of discursive communities. In fact, the list of most retweeted users reflects coherent behavior related to the peculiar political affiliation of each discursive community, further confirming the value of this specific interaction feature for the formation of collective identities. Moreover, rankings can be compared across the two data acquisition periods.

Results in Tab. 5.2 reveal two different modes of building collective political identities. The first interactional form consists of an 'institutional' pattern displayed within communities where retweeted accounts mainly belong to either parties or political leaders. This is the case of the M5S communtity which in both data sets shows a clear preference in retweeting accounts of politicians belonging to the 'Movimento 5 Stelle' (i.e. *@ManlioDS*, *@luigidimaio*, *@carlosibilia* and *@virginiaraggi* belonging respectively to Manlio Di Stefano, Luigi Di Maio, Carlo Sibilia and Virginia Raggi) and to journalists of the newspaper 'Il Fatto Quotidiano' (i.e. *@fattoquotidiano* and *@petergomezblog*) recognized as a media organization close to this party. In general, all the communities belonging to the 2018 data set, except for the MEDIA community, display an institutional pattern strongly directing their retweeting activity towards politicians or institutional party accounts belonging to that community. In fact, while the CDX community which is mainly interested in retweeting center- and far-right politicians or institutional party accounts (e.g. *@matteosalvinimi*, *@renatobrunetta*, *@borghi_claudio* belonging respectively to Matteo Salvini, Renato Brunetta and Claudio Borghi), the CSX community preferentially retweets center-left politicians or institutional party accounts (i.e. *@matteorenzi* and *@pdnetwork* belonging respectively to Matteo Renzi and the institutional party account of the 'Partito Democratico').

The second mode of interaction consists of a 'NGOs and media-oriented' pattern observed, for instance, in the 2019 data set both within the CSX community and, even to a larger extent, within the MINGOs community. This interactional mode occurs in those communities where most retweeted accounts belong either to media or non-governmental organizations. For instance, the most retweeted account in the CSX community refers to the online newspaper 'Linkiesta' (e.g. *@Linkiesta*) while the second one to the 'Caritas Italiana' (e.g. *@caritas_milano*), a Catholic NGO based in Milan; other accounts belong to public figures very active on social media, such as Roberto Saviano (e.g. *@robertosaviano*), a journalist who is known for his reports about mafia crimes in Southern Italy. Instead, in the MINGOs community, the presence of Catholic organizations reveals an attention towards the account of the Pope (e.g. *@Pontifex*) whose pleas for solidarity resonated loudly during the observation period, and Catholic media outlets, such as the newspaper 'L'Avvenire' (e.g.

@Avvenire_Nei).

In contrast to retweets, mentions are also transversally used to interact with members of other communities. Results in Tab. 5.2 better specify the substance of cross-community interactions and confirm that discursive dynamics accompany and, to some extent, overlap with offline political alliances. This evidence is less pronounced in the discursive communities of the 2018 data set where the electoral debate taking place in the Twitter conversations results in a mentioning induced ranking list of politician or institutional party-related accounts similar to the retweeting induced ranking. Whereas, in the 2019 data set, the DX and the M5S communities reciprocally open up to each other as they jointly sit in the government chaired by Giuseppe Conte (i.e. *GiuseppeConteIT*), who was a 'Movimento 5 Stelle' candidate in the run up to the 2018 general elections and appears only in the 2019 data set after the formation of the government with the 'Lega'. All the other communities find in the account belonging to Matteo Salvini a common target for their communications. The former Minister of the Interior was strongly positioned against migration and hostile to search-and-rescue missions as well as to sheltering operations This common trend towards directly addressing his account results in a strong personalization of the overall debate on migration, which ends up pivoting around the role and the responsibility of this specific political actor. A relevant finding shown in Tab. 5.2 is the substantial difference in the ranking lists of the most retweeted and the most mentioned accounts. As already noticed, this can be explained by the different meaning given by users to these two Twitter actions [12, 60]: while mentions, as replies, represent a conversational interaction and actual recognition (either positive or negative) between users, retweets express a deliberate broadcasting mechanism often employed to spread other users' tweets. In this sense, retweets display a better capacity in detecting actual political affiliation, as highlighted also by the distribution of the mean polarization index $\langle\rho_\alpha\rangle$ reported in Fig. 5.4.

2018 Italian general Elections									
CDX		**M5S**		**CSX**		**LEU**		**MEDIA**	
Retweets	*Mentions*	*Retweets*	*Mentions*	*Retweets*	*Mentions*	*Retweets*	*Mentions*	*Retweets*	*Mentions*
@borghi_claudio	@forza_italia	@fattoquotidiano	@matteorenzi*	@repubblica	@pdnetwork	@PietroGrasso	@PietroGrasso	@chetempochefa	@fabfazio
@matteosalvinimi	@berlusconi	@petergomezblog	@Mov5Stelle	@matteorenzi	@repubblica	@repubblica*	@civati	@civcatt	@roncellamare*
@renatobrunetta	@matteosalvinimi	@you_trend	@fattoquotidiano	@Corriere	@matteorenzi	@civati	@repubblica*	@francescoseghez	@RaiUno*
@SkyTG24	@AlbertoBagnai	@ManlioDS	@pdnetwork*	@pdnetwork	@LaStampa	@lauraboldrini	@lauraboldrini	@Avvenire_NEI	@lucianinalitti*
@LeaveEUOfficial	@borghi_claudio	@Agenzia_Ansa	@luigidimaio	@Agenzia_Ansa*	@luigidimaio*	@fabiochiusi	@robersperanza	@ReginaCatrambon	@Altroconsumo*
2019 Italian migration discussion									
DX		**M5S**		**CSX**		**FI**		**MINGOs**	
Retweets	*Mentions*	*Retweets*	*Mentions*	*Retweets*	*Mentions*	*Retweets*	*Mentions*	*Retweets*	*Mentions*
@matteosalvinimi	@matteosalvinimi	@fattoquotidiano	@matteosalvinimi*	@caritas_milano	@matteosalvinimi*	@renatobrunetta	@forza_italia	@Pontifex	@repubblica*
@GiorgiaMeloni	@GiuseppeConteIT*	@virginiaraggi	@GiuseppeConteIT	@Linkiesta	@repubblica	@msgelmini	@matteosalvinimi*	@repubblica*	@matteosalvinimi*
@LegaSalvini	@repubblica*	@GiuseppeConteIT	@luigidimaio	@robertosaviano	@GiorgiaMeloni*	@berlusconi	@berlusconi	@UNHCRItalia	@LaStampa*
@Capezzone	@GiorgiaMeloni	@Mov5Stelle	@virginiaraggi	@repubblica	@pdnetwork	@GiorgioBergesio	@simonebaldelli	@Pontifex_it	@Avvenire_Nei
@NicolaPorro	@luigidimaio*	@carlosibilia	@Mov5Stelle	@CarloCalenda	@CarloCalenda	@mara_carfagna*	@renatobrunetta	@Avvenire_Nei	@RaiNews

Table 5.2: Ranking of the five most retweeted and mentioned verified accounts for each discursive community in both data sets. In both data sets, asymmetries in the ranking lists can be explained by the different role of relational mechanisms within Twitter: while mentions represent a recognition (either positive or negative) in a conversational interaction between users, retweets express a deliberate broadcasting mechanism. In each ranking list, account names marked by an asterisk are external to that discursive community.

2018 Italian general Elections				
CDX	**M5S**	**CSX**	**LEU**	**MEDIA**
affariitaliani.it (772)	ilfattoquotidiano.it (2373)	repubblica.it (3174)	torinoggi.it (97)	avvenire.it (12)
google.it (770)	ansa.it (1158)	corriere.it (1878)	repubblica.it (67)	rainews.it (7)
corriere.it (682)	repubblica.it (1135)	lastampa.it (1160)	lastampa.it (52)	laciviltacattolica.it (4)
ansa.it (567)	corriere.it (862)	informazione.it (1042)	ilfattoquotidiano.it (45)	oggicronaca.it (4)
youtube.com (546)	youtube.com (601)	youtube.com (1002)	corriere.it (38)	huffingtonpost.it (2)
2019 Italian migration discussion				
DX	**M5S**	**CSX**	**FI**	**MINGOs**
ilgiornale.it (70814)	ilfattoquotidiano.it (10892)	repubblica.it (41418)	freenewsonline.it (252)	vatican.va (16047)
imolaoggi.it (39895)	lanotiziagiornale.it (3379)	avvenire.it (23032)	gregoriofontana.com (54)	repubblica.it (2180)
laverita.info (34316)	ilblogdellestelle.it (1748)	linkiesta.it (16121)	ilgiornale.it (37)	rainews.it (1620)
ansa.it (33056)	affaritaliani.it (1352)	lastampa.it (15726)	repubblica.it (29)	avvenire.it (1358)
liberoquotidiano.it (27525)	ansa.it (1341)	corriere.it (10408)	milanopost.info (27)	unhcr.org (1357)

Table 5.3: Ranking of the five most shared domain names of news media contents for each discursive community. The difference in ranking lists confirm the users segregation in communities with similar media contents consumption patterns. For each domain name, the total number of shared links is shown in parentheses.

Another interesting insight about the Twitter activity of discursive communities can be seen by digging into the web domains most often shared via retweets. As confirmed in previous studies [7, 23], Twitter, like other social media platforms, is a digital space where algorithmic mechanisms and homophilic tendencies spur users with a similar content consumption pattern to form communities. This result can be witnessed in Tab. 5.3 which provides an overview of the five most shared domain names of news media in the five discursive communities for both data sets. As shown in these ranking lists, most of the shared domain names within each discursive community tend to differ. In particular, the analysis of the ranking lists of the 2019 data set reveals that the most shared domain names belong to newspaper websites with known political orientation. For instance, the domains *ilgiornale.it* and *imolaoggi.it* in the DX community belong to two right-wing newspapers, while *repubblica.it* and *lastampa.it* in the CSX community refer to news media websites with center-left positions. Finally, *ilfattoquotidiano.it* and *ilblogdellestelle.it* are the domain names respectively of the ‘Il Fatto Quotidiano’, whose Twitter account is present within the M5S discursive community, and of the official blog of the ‘Movimento 5 Stelle’.

A closer look at the most shared single web-link is another indicator of the pro- or anti-migration positions of each discursive community in the 2019 data set: while the most prominent news inside the CSX community concerns a UN sentence against Salvini’s security decree published by ‘La Stampa’[1], in the CDX community the most relevant news regards the alleged absence of a sanitary crisis on the rescue boat Open Arms, the reason for which it landed in Lampedusa[2]. While in the M5S community, the most shared news is

[1]The web-link of the present news whose title is ‘«Viola le norme e promuove la xenofobia», dura condanna dell’Onu al decreto Salvini’ (‘*«Italy violates the rules and promotes xenophobia», harsh UN sentence towards Salvini’s security decree*’) published on ‘La Stampa’ on 19 May 2019 is present here: *www.lastampa.it/news/decreto-salvini*

[2]The web-link of the present news whose title is ‘Il medico di Lampedusa nei guai perché i migranti non sono malati’ (‘*The doctor of Lampedusa in trouble because migrants are not sick*’) published on ‘Il Giornale’

an interview with the Mayor of Lampedusa who criticized Salvini's activity as the Minister of Internal Affairs and published a few days before the no-confidence motion presented by the 'Lega' against the Prime Minister[3]. Closely related with the peculiar composition of the other two communities, the most relevant news in the FI community is the news about an arrested group of smugglers in Salerno[4] while in the MINGOs community it is the Homily in the Mass for migrants given by Pope Francis on 8 July 2019[5]. Similarly, the five discursive communities present in the 2018 data set tend to differ when the single most shared web-links are compared.

The outcome of the overall analysis confirms the significance of the final composition of the discursive communities. Given the previous results, discursive communities can be therefore defined as coherent partisan groups of Twitter users providing an innovative starting point to study online collective identities.

5.2 Analysis of users' role and influence

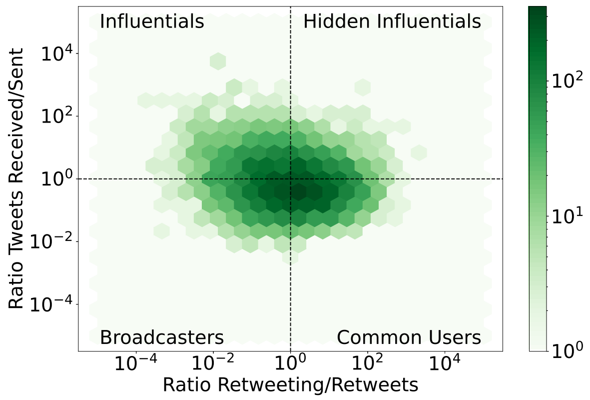
Users belonging to partisan discursive communities can be studied in terms of their influence and visibility within the information diffusion environment on Twitter. These features can be assessed by employing a quantitative approach which measures their actual relevance within discussions in which they participate. In [69], Twitter users are divided into four different categories according to the values of two quantities: the ratio of the number of 'received messages' (i.e., the tweets where a user is mentioned) over that of 'sent messages' (i.e., the tweets published by each single user); and the ratio between the friends and the number of followers a single user has. The former is a measure of users' influence within the discussion: the more a user is mentioned, the more he/she is considered a target in the domain-specific communication flow of the audience. The latter is a measure of visibility since the number of followers of a single user represents an estimate of the potential global audience that can be reached. To measure more effectively the actual audience of each user, in this work an adjustment of this latter quantity is implemented: instead of using the ratio 'friends over followers', the visibility of users is expressed with the ratio of the number of retweets performed by each user (i.e., his/her retweeting) over the number of retweets his/her tweets collect. While the ratio 'friends over followers' has a constant value for all the

on 17 August 2019 is present here: *www.ilgiornale.it/news/medico-lampedusa*

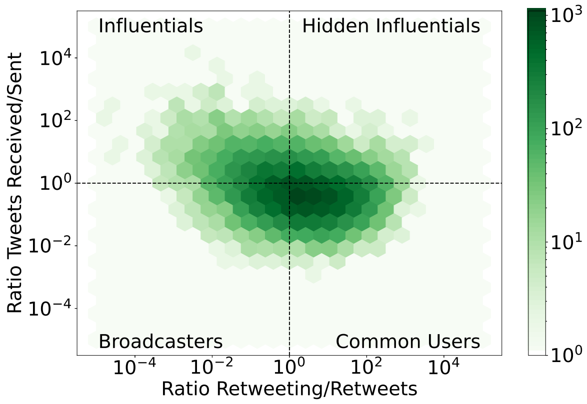
[3]The web-link of the present news whose title is 'Migranti, sindaco Lampedusa: «Salvini? Con lui gli sbarchi sono aumentati. Gli ho chiesto un incontro, non mi ha mai risposto»' ('*Migrants, Mayor of Lampedusa: «Salvini? The landings have increased with him. I asked him for a meeting, he never answered me.»*') published on 'Il Fatto Quotidiano' on 2 August 2019 is present here: *www.ilfattoquotidiano.it/news/lampedusa-salvini*

[4]The web-link of the present news whose title is 'Salerno, contrabbandieri sigarette col reddito cittadinanza: fermati' ('*Salerno, Cigarette smugglers with citizens' income: stopped.*') published on 'TGcom24' on 21 October 2019 is present here: *www.tgcom24.mediaset.it/news/salerno-contrabbandieri-sigarette*

[5]The web-link of the present news whose title is 'Holy mass for migrants - Homily of the Holiness Pope Francis' published on vatica.va on 8 July 2019 is present here: *www.vatican.va/homilies/papa-francesco-omelia-migranti*



(a) 2018 Italian general elections



(b) 2019 Italian migration discussion

Figure 5.6: Scatterplot of users categorization for the five discursive communities in both data sets. According to a categorization method of active users able to capture their visibility and influence, these users are divided into four distinct types depending on the quadrant of the X-Y plane where they are located: common users (in the bottom right quadrant), broadcasters (in the bottom left quadrant), influentials (in the top left quadrant) and hidden influentials (in the top right quadrant).

conversations occured over the same time period, this ratio allows for directly connecting the actual audience of each user to the specific conversation. In this way, the actual relevance of each user is computed as a combination of visibility and influence in a X-Y Cartesian plane.

Here, it is worth noticing that an additional benefit guaranteed by the use of this new variable is the selection of a specific subset of users: in fact, only *active users* who published one or more tweets, performed one or more retweets and received at least one retweet and a mention are considered in this analysis. As expected, in both data sets, the highest number of active users belongs to the CSX discursive community as it is the largest. Interestingly, while in the 2018 data set the CDX community shows a comparable number of users to that of the M5S community, in the 2019 data set the number of users in the DX community is similar to those in the CSX community.

According to the previously defined categorization method and to the same labels presented in [69], active users are divided into four distinct types depending on the quadrant of the X-Y plane where they are located:

- **Common users:** active users who are located in the bottom right quadrant because of their high number of retweets and low rate of received mentions. Their contribution in the information flow is the diffusion of their own and other authors' tweets;

- **Broadcasters:** active users who are located in the bottom left quadrant as they receive more retweets than those they perform without being mentioned a relevant number of times in other authors' tweets. In other words, their tweets have a large audience but they do not succeed in being considered as relevant actors in the discussion;

	Influentials	Hidden Influentials	Broadcasters	Common Users	
CDX	410 (23.8%)	355 (20.7%)	359 (20.8%)	597 (34.7%)	**1721**
CSX	**889** (23.8%)	**521** (14.0%)	**999** (26.7%)	**1324** (35.5%)	**3733**
M5S	233 (16.0%)	188 (12.9%)	383 (26.4%)	650 (44.7%)	**1454**
LEU	62 (53.4%)	19 (16.4%)	14 (12.1%)	21 (18.1%)	**116**
MEDIA	4 (57.1%)	0 (0.0%)	3 (42.9%)	0 (0.0%)	**7**
	1598	**1083**	**1758**	**2592**	**7031**

Table 5.4: Number and percentage of users according to the four-types categorization for the five discursive communities in the 2018 data set. Bold numbers are those indicating the largest amount of each user type within a single community. The total number of users for each community and user type are also displayed in bold. For each user type, the fraction of users shown in parentheses is computed over the total number of active users in each discursive community.

- **Influentials:** active users who are in the top left quadrant which is the area of most relevant accounts. Influentials tend to be mentioned and retweeted more than they do the same with other users within the discussion;
- **Hidden influentials:** active users who are located in the top right quadrant since they are mentioned more often as influentials but they publish tweets which collect a number of retweets similar to that of common users.

In Fig. 5.6, the separation into groups of active users according to the aforementioned categorization is shown through a logarithmic scatterplot. As expected, active verified users are mainly influentials: indeed, ≈69% and ≈78% of active verified users (respectively, in the 2018 Italian general elections data set and in the 2019 Italian migration discussion data set) fall within this category. As expected, only ≈3% of verified users of both data sets are identified as common users. Given the overall prominence within Twitter of verified users (who represent accounts of public figures, organization or companies), this result is not surprising. The interesting evidence is the fact that influentials are composed of a vast majority of non-verified users representing ≈85-90% of the whole number of users within this category for both data sets. The centrality of non-verified users sheds light on their prominent role within the Twitter communication process. This evidence can be also influenced by the reduced number of verified users due to the temporary suspension of the profile verification process since 2017 [101]. Nevertheless, the present finding clearly shows the need to include them within discursive communities.

In Tab. 5.4 and 5.5 a brief overview of the number of active users in each category subdivided into the five discursive communities is shown for both data sets. In almost all communities, common users exceed users belonging to other categories with a slight difference across the data sets: while in the 2018 data set their overall number is approximately comparable with that of broadcasters and influentials, in the 2019 data set common users are three times more than influentials. A possible explanation lies in the different data acquisition period covered by the two data sets: in fact, since the time period of the 2018 data set is less than two

	Influentials	Hidden Influentials	Broadcasters	Common Users	
DX	999 (10.8%)	1125 (12.1%)	**1886** (20.3%)	**5262** (56.8%)	**9272**
CSX	**1760** (18.6%)	**1437** (15.2%)	1781 (18.1%)	4489 (47.4%)	**9467**
M5S	245 (23.4%)	208 (19.9%)	192 (18.3%)	403 (38.4%)	**1048**
FI	16 (36.4%)	8 (18.2%)	11 (25.0%)	9 (20.4%)	**44**
MINGOs	250 (40.1%)	93 (14.9%)	157 (25.2%)	123 (19.7%)	**623**
	3270	**2871**	**4027**	**10286**	**20454**

Table 5.5: Number and percentage of user according to the four-types categorization for the five discursive communities in the 2019 data set. Bold numbers are those indicating the largest amount of each user type within a single community. The total number of users for each community and user type are also displayed in bold. For each user type, the fraction of users shown in parentheses is computed over the total number of active users in each discursive community.

months, it is more likely to gain influence and visibility with a few tweets over such a short time period. Moreover, in both data sets, only the communities with the lowest number of users (i.e., LEU and MEDIA in the 2018 data set and FI and MINGOs in the 2019 data set) display a peculiar behavior with a higher proportion of influentials with respect to the other categories. This could possibly be explained by the need for smaller communities to build their framing around a higher number of influentials who are responsible for most of the communication process.

Concerning the relative proportion of user categories, the three major communities in the 2018 data set show distinctive behaviors: while the CDX and the CSX communities present a similar percentage of influentials and common users, the fraction of common users within the M5S community is significantly higher than that of the remaining categories. Moreover, within the CDX community, the fraction of hidden influentials is similar to that of broadcasters. These findings highlight the specificities of the communication process occuring within each discursive community: in the M5S community common users appear to play a major role, while in the CDX and the CSX communities influentials and broadcasters are most relevant. Similarly, the 2019 data set offers interesting insights: while the CSX community host about twice the number of influentials compared to the DX community (representing nearly 60% of the overall amount of influentials), this latter is composed of nearly 80% of common users and broadcasters. This result can be another indicator of the different conversational nature of these two communities: users belonging to the DX community are more able to trigger chains of information flows through a retweeting mechanism. This kind of communication process is preferred to the mentioning activity which is instead more common within the CSX community.

5.3 Semantic polarization

The use of a common vocabulary and a shared narrative is crucial to the formation of coherent groups of Twitter users. Previous results on lexical convergence in online user communities

Discursive community	N_t (%)	N_h	max(n_h)	<n_h>	median(n_h)
CSX	34.0%	14073	987	4.6	2
CDX	37.3%	9293	1182	8.5	3
M5S	45.2%	8576	779	11.1	3
LEU	40.2%	997	163	6.2	3
MEDIA	49.9%	195	22	3.0	2

Table 5.6: Hashtag use for each discursive community in the 2018 data set. The percentage of tweets including at least one hashtag (N_t), the total number of single hashtags (N_h) and a few statistics of the distribution of the number of hashtags per user (n_h) are here reported.

show that people who interact more are more likely to share a common set of keywords [122]. Lexical convergence is mainly due to the emergence of a single and shared collective identity which can also pass through the use of negative narratives to represent other groups or individuals from which the users differentiate their own viewpoint [123].

Activity of the various discursive communities in terms of mentioning, retweeting and media content sharing is only one aspect of the more complex formation of coherent groups of users. This process can also be understood by looking at the use of keywords within each discursive community. In order to address this aspect, the term *semantic polarization* is employed to refer explicitly to a deliberate use of language depending on the political affiliation. The present analysis focuses on hashtags as they are *thematic tags* playing a central role within Twitter [104]. Hashtags are not only an important reference of electoral or political slogans but also an indicator of the main topics, actors or contexts within each discussion. Moreover, hashtagging is part of the lexical construction of narratives within homophilic groups which are connected through the exchange of stories and counter-narratives [55, 56]. In order to analyze the use of hashtags in each discursive community, the list of hashtags published or retweeted by each user is compiled through the same procedure employed for the construction of the user-hashtag bipartite networks described in 4.1.2.

To inspect the different use of hashtags within discursive communities, the total number of single hashtags N_h is computed along with few statistics of the distribution of the number of hashtags per user n_h. In Tab. 5.6 and 5.7, an overview of these quantities are reported for each discursive community. In both data sets, the CSX discursive community shows the highest value of N_h. Surprisingly, the M5S community has an approximately similar number of hashtags in both data sets even though an increasing number of hashtags was expected to be directly correlated with an increasing time period of the data acquisition window.

Regarding the distribution of n_h, common characteristics are found in both data sets. In particular, two features clearly signal that n_h is a scale-free distribution [124]: the order of magnitude of the difference between the median and the maximum values and the difference

Discursive community	N_t (%)	N_h	max(n_h)	<n_h>	median(n_h)
CSX	44.7%	31019	1255	12.5	4
DX	35.3%	28422	1836	27.2	4
M5S	46.8%	6700	832	26.9	7
FI	70.9%	657	289	10.7	4
MINGOs	46.9%	7878	555	7.4	3

Table 5.7: Hashtag use for each discursive community in the 2019 data set. The percentage of tweets including at least one hashtag (N_t), the total number of single hashtags (N_h) and a few statistics of the distribution of the number of hashtags per user (n_h) are here reported.

in the mean and the median values. This asymmetry is mainly due to an abundance of users who published or retweeted a low number of hashtags while few accounts are mainly responsible in terms of the overall production of hashtags in each discursive community. The distribution n_h is observed for all the discursive communities to roughly follow the '80-20 rule' of the Pareto distributions giving additional important evidence of its long-tail behavior. The maximum count of hashtags used by a single user is observed in the CDX and DX discursive communities while the M5S community is characterized in both data sets by a high mean value of n_h. This can be interpreted as different behavior of these discursive communities: while in the communication process within the CDX and the DX communities the massive use of hashtags is limited to a few users, the M5S community is characterized by a more widespread use of hashtags for each single user. It is worth noticing that in the ranking list of the first ten users who employ the highest number of hashtags per discursive community only three verified users are present and they belong to the CSX (*@Agenzia_Dire*) and CDX (*@forza_italia* and *@GruppoFICamera*) discursive communities in the 2018 data set. In light of the previous results on the central role of verified users in the communication process, this evidence can be interpreted as the fact that each user, in particular those who are mostly influentials, focuses on a relatively low number of hashtags.

To further investigate the use of hashtags in each discursive community, the TF-IDF value is computed for each hashtag. This parameter provides an estimate of the hashtag frequency combined with its uniqueness across all the discursive communities. The TF-IDF measure connects these two frequency terms by discounting the absolute frequency of a commonly used hashtag by the number of discursive communities where this hashtag appears:

$$\text{TF-IDF}(\text{comm}, \#) = \log(1 + f(\#, \text{comm})) \cdot \log\left(\frac{N_c}{1 + n_c} + 1\right) \tag{5.2}$$

where $f(\#, \text{comm})$ is the raw count of a single hashtag in each discursive community, N_c is the total number of discursive communities and n_c is the number of actual communities where the single hashtag appears. The TF-IDF measure can be described as follows: while the first term in Eq. 5.2 represents a logarithmic scaled term-frequency accounting for the order of magnitude of the hashtags' use, the second is a commonly used inverse document

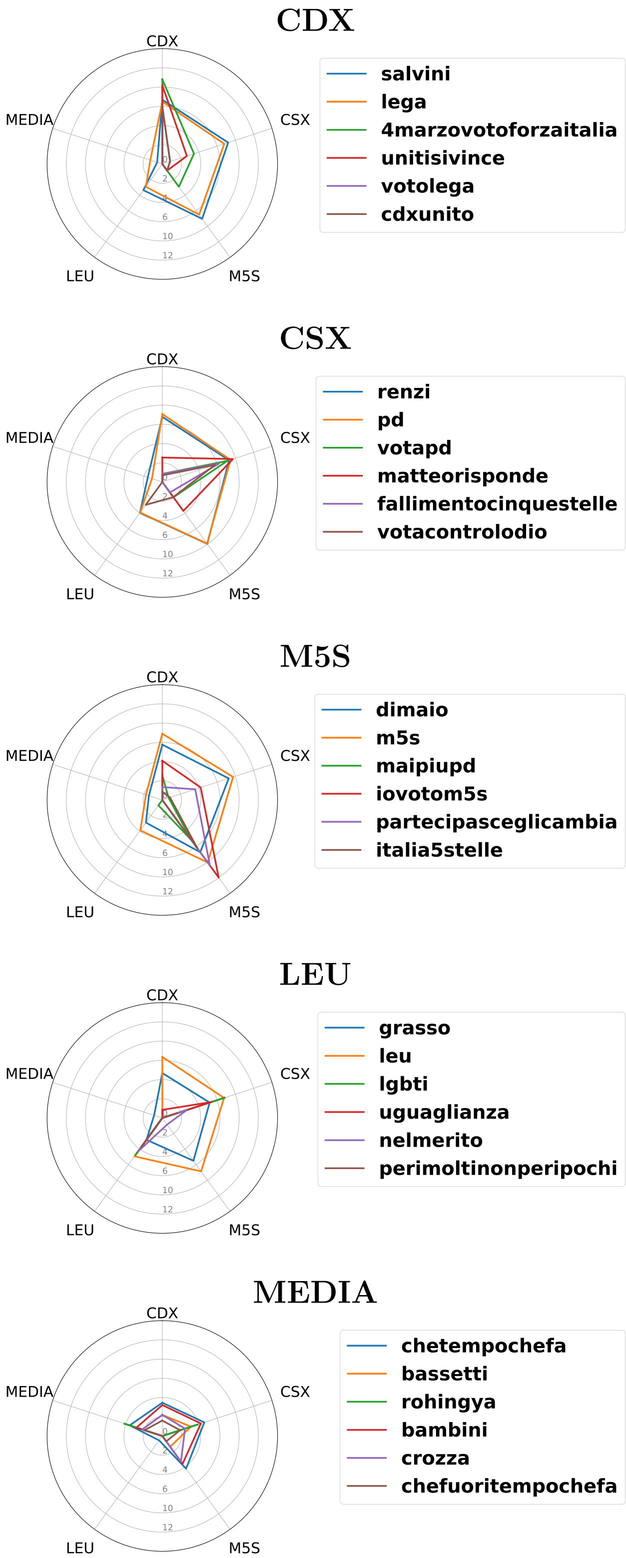


Figure 5.7: Radar plot of a selected list of TF-IDF values for each discursive community in the 2018 Italian general elections data set. Each radar plot displays TF-IDF values of a selected list of commonly-used hashtags in the discursive community reported in the plot title. A semantic polarization clearly emerges when hashtags concerning electoral slogans or political messages are considered.

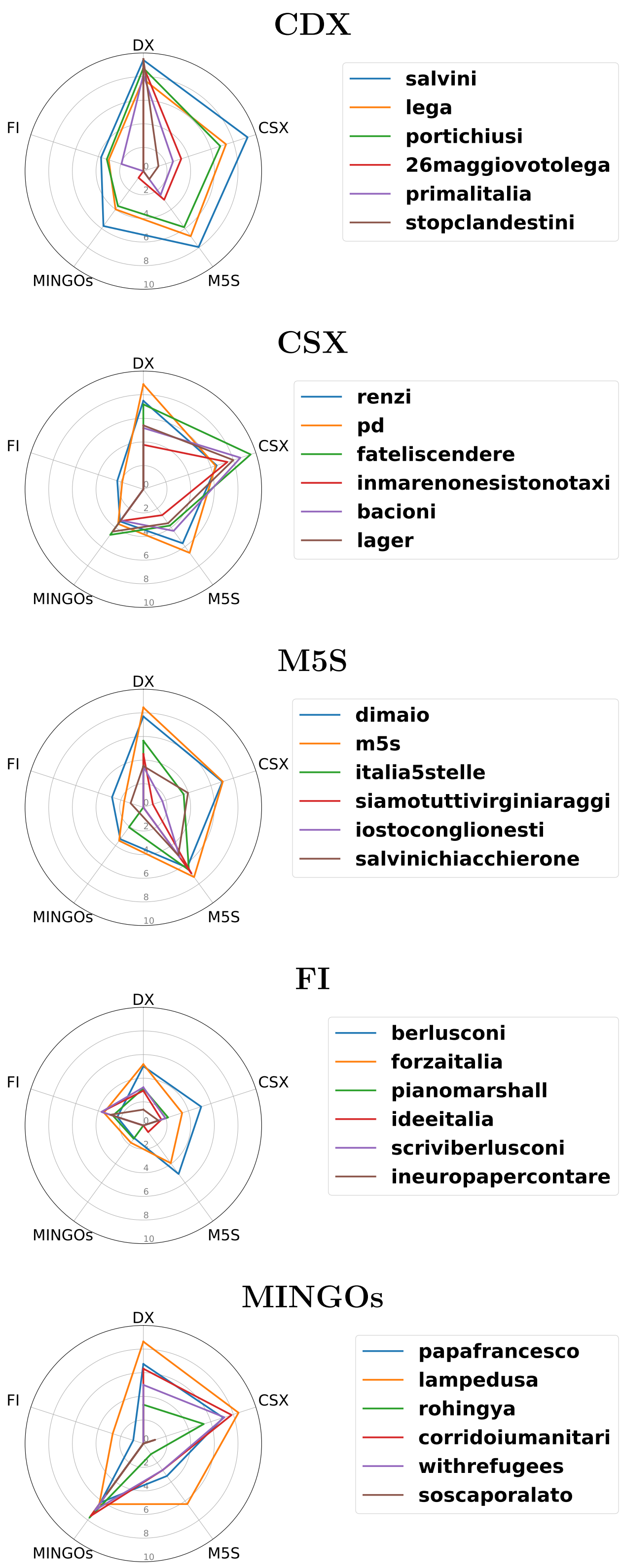


Figure 5.8: Radar plot of a selected list of TF-IDF values for each discursive community in the 2019 Italian migration discussion data set. Each radar plot displays TF-IDF values of a selected list of commonly-used hashtags within the discursive community reported in the plot title. A semantic polarization clearly emerges when hashtags concerning electoral slogans or political messages are considered.

frequency term. This last term is able to take a different account of hashtags indistinctly used in all the discursive communities and of those which are mainly employed by users of a single community. In Fig. 5.7 and 5.8, a comparison between TF-IDF values of a selected set of commonly used hashtags in each discursive community is shown through a series of radar plots.

In both data sets, the first two hashtags reported in the legend of each radar plot refers to political leaders, such as *#salvini*, *#renzi* and *#dimaio* (referring respectively to Matteo Salvini, Matteo Renzi and Luigi di Maio), and parties, such as *#lega*, *#pd* and *#m5s*. Results show that TF-IDF value of these hashtags is approximately similar in all discursive communities. The significant amount of common hashtags shared by all discursive communities, nearly equal to half of the lowest count of the N_H value for both data sets, is driven by a semantic convergence probably caused by the adoption of a similar vocabulary rather than by an opinion convergence or the mixing between discursive communities. This evidence clearly shows a tendency to use a shared set of hashtags referring to common actors, contexts or events in the discussion. Meanwhile, a different set of hashtags concerning electoral slogans or political messages specifically connected to each discursive community is mainly employed in that community. By looking at the last four hashtags reported in the legend of each radar plot, an asymmetric behavior is observed. In fact, these specific hashtags present TF-IDF values clearly peaked towards the specific community shown in the plot title. For instance, in the 2018 data set, the hashtags *#4marzovotoforzaitalia*, *#votapd*, *#iovotom5s* and *#perimoltinonperipochi* (respectively, *I vote 'Forza Italia' on 4 March*, *vote PD*, *I vote M5S* and *For the many not the few*) are clearly electoral slogans referring to a specific political coalitions. As expected, they reach their peak, respectively, in correspondence to the CDX, CSX, M5S and LEU discursive communities. Similarly, the hashtags *#portichiusi*, *#fateliscendere*, *#italia5stelle* and *#corridoiumanitari* (respectively, *closed ports*, *let them land*, *Italy 5 stars* and *humanitarian corridors*) used in the 2019 data set have similar asymmetries peaking towards the corresponding discursive community where these topic-related hashtags are more frequently used.

Chapter 6

Networked framing: Structure and dynamics of semantic networks

The study of networked framing of the debate about the 2018 Italian general elections and the 2019 migration discussion is carried out by looking at the topological features of the semantic networks induced by each discursive community. To this end, user-hashtag bipartite networks are built and projected on the layer of hashtags, for each community, via the procedure described in Chapter 4. Consistently with previous studies, the semantic analysis is centered on hashtags since they «create visibility for a message [. . .] not only marking context but also changing and adding content to the tweet» [76]. More specifically, hashtags are considered key devices for enacting networked framing practices within networked publics [55, 56, 58, 59] and, following Recuero et al. [76], their association is recognized within tweets as a strategy for conveying specific narratives but also for mobilizing specific audiences.

The results of the current Chapter have been published in two different papers: Section 6.1 shows an overview of the analysis presented in [4] and described as «pioneering» in [125] while Section 6.2 overlaps with those presented in [6]. The next Sections are organized as follows: in Section 6.1, the Twitter activity about the 2018 Italian general elections is revealed by the analysis of the daily semantic networks. It is worth noticing that, given the limited number of users within the LEU and MEDIA communities in the 2018 data set, the semantic analysis is performed only for the three main communities (i.e., CDX, CSX and M5S) following the discursive communities composition and the analysis flow reported in [4]. While Sections 6.1.1, 6.1.2 and 6.1.3 focus on the study of the non-filtered projections, Section 6.1.4 presents the results of the filtering network techniques on the daily semantic networks. Finally, in Section 6.2, the properties of the monthly semantic networks induced by the 2019 migration discussion are investigated. More specifically, unlike the analysis of the daily semantic networks, Section 6.2 will be only focused on the analysis of the BiCM projections as this null model provides a sufficiently non-sparse structure.

6.1 Daily semantic networks of the 2018 Italian elections

A closer inspection of semantic networks allows this analysis to be specifically focused on the networked framing practices unfolding within discursive communities. A first step in this direction can be made by exploring the number of *nodes*, which is the proxy for the number of topics discussed by users, and their *mean degree* (i.e. the mean number of neighbors per node), which is the proxy for the average prominence of topics that characterize the discussion.

Results obtained in this step are shown in Fig. 6.1. The number of nodes in all discursive communities increases up to the day after the general elections, and then decreases. This indicates that the number of topics debated by users increases as Election Day approaches. The M5S seems to be the most active community with the largest number of debated topics throughout the 2018 data set observation period. The trend characterizing the M5S community is closely followed by that of the right-wing alliance up to the end of February, when an inversion takes place and a rise in the number of topics debated by the supporters of the center-left alliance becomes clearly visible.

Overall, the trend of the mean degree is, much less regular: it is in fact characterized by several 'bumps of activity' throughout the entire period. Peaks in the daily use of hashtags correspond to so-called *mediated events*, i.e., events of social relevance broadcast by legacy media and, in particular, by national television channels. The importance of media events is suggested by the prominence of hashtags referring to political talk shows broadcast by Italian television, such as *#dallavostraparte* and *#tagadala7*. To confirm this intuition, tweets contributing to activity bumps have been explored through a qualitative check jointly performed by the entire research team. Indeed, contents tweeted by users and mainstream media have been found to sistematically match, particularly, in the case of TV shows featuring prominent politicians. Afterwards, a cross-check with printed and online contents has been carried out to verify the actual presence of political leaders and figures during talk shows mentioned by users in their tweets. Results suggest that users tend to become particularly active during or in close correspondence to political debates signalling and/or commenting the presence of certain candidates in TV shows. Such a behavior is particularly evident for the CDX community, whose mean degree is characterized by a larger number of peaks. More specifically, peaks are observed in correspondence of the following TV shows:

- **09 February**: interview of Silvio Berlusconi at TG La7 (hashtags: *#silvioberlusconi*, *#tgla7*);

- **11 February**: Silvio Berlusconi and Matteo Salvini are interviewed at 'Mezz'ora in più' (hashtags: *#il4marzovotaefaivotareforzaitalia*, *#mezzorainpiù*);
- **13 February**: Nicola Porro, an Italian journalist, announces via a Facebook video, the topics that will be discussed on his TV show 'Matrix', broadcast by 'Canale 5', a TV channel owned by the Berlusconi family (hashtags: *#nicolaporro*, *#matrix*);
- **18 February**: interview of Silvio Berlusconi in the TV show 'Che tempo che fa' (hashtags: *#chetempochefa*, *#silvioberlusconi*);
- **19 February**: interview of Silvio Berlusconi in the TV show 'Dalla vostra parte' (hashtags: *#dallavostraparte*, *#silvioberlusconi*);
- **22 February**: Matteo Salvini and Anna Maria Bernini, a member of 'Forza Italia', are hosted in the TV show 'Quinta colonna' broadcast on 'Rete 4', another TV channel owned by the Berlusconi family (hashtags: *#forzaitaliaberlusconipresidente*, *#quintacolonna*);
- **26 February**: Guido Crosetto and Maurizio Gasparri (both from the right-wing alliance) are hosted in the TV show 'L'aria che tira' (hashtag: *#lariachetirala7*);
- **16 March**: interview of Michaela Biancofiore, a member of 'Forza Italia', in the TV show 'Tagadà' (hashtags: *#tagada*, *#tagadala7*).

Beside confirming that Twitter discussions can be *influenced* by external events, these results point out that Twitter discussions can be also *triggered* by external events. This is especially true for the CDX community whose Twitter discussions are mainly driven by the aforementioned mediated events [126]. This could be an effect of the interplay between the behavior of CDX users who still conceive of television as the main source of information when it comes to political processes and the massive presence of center- and far-right politicians in TV shows. In particular, the latter might be part of a different communication strategy in their hybrid campaigns [78].

6.1.1 Analyzing hashtags persistence and centrality

A second step towards a better understanding of the networked framing practices within discursive communities consists of quantifying the interest towards a topic throughout the entire data acquisition period. To this end, *hashtag persistence*, H_t (i.e., the percentage of days a hashtag is present in the 2018 data set) has been analyzed on non-filtered projections. As reported in Tab. 6.1, results shows that the most persistent hashtags are those referring to the name of political parties (i.e. *#lega*, *#m5s*, *#pd*) and political leaders (i.e. *#berlusconi*, *#dimaio*, *#renzi*, *#salvini*). Moreover, most persistent hashtags in almost all discursive communities refer mostly to political actors and figures of the opposite coalition. When it

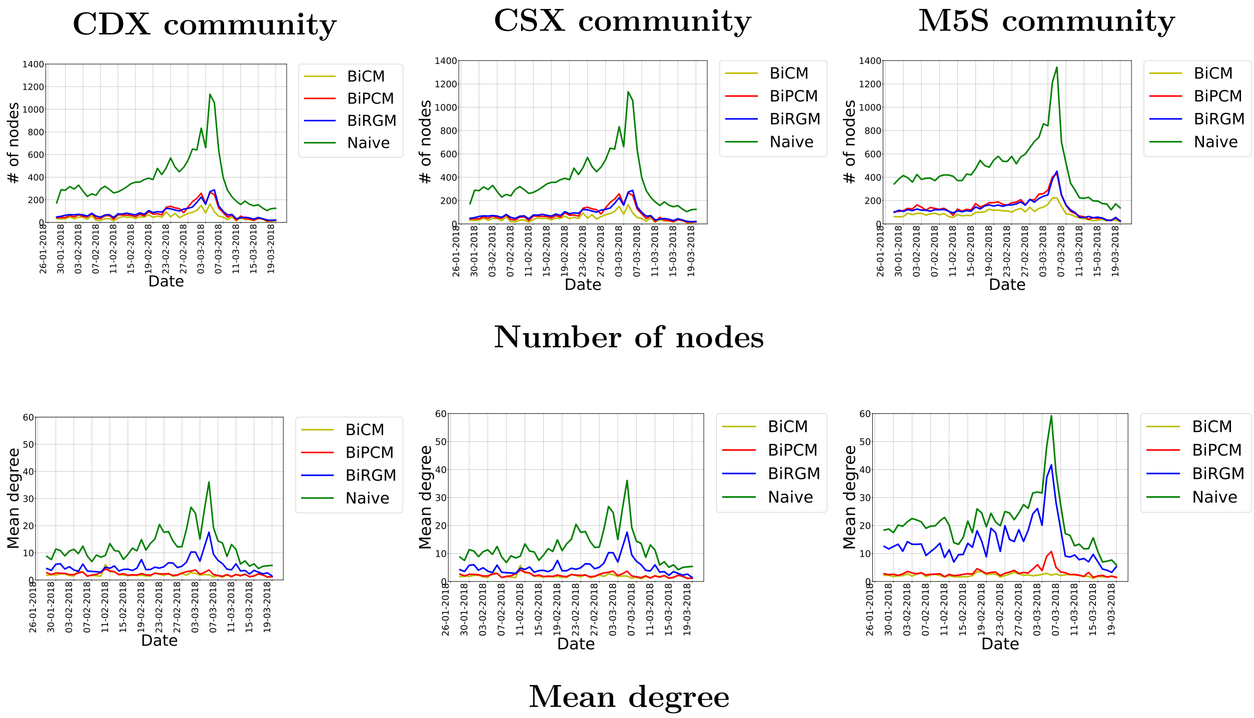


Figure 6.1: Temporal evolution of the number of nodes (top panels) and of the mean degree (bottom panels) of community-specific semantic networks. Peaks in the daily use of hashtags correspond to *mediated events*, i.e. events of social relevance broadcast by mass media. This behavior is particularly evident for the CDX community whose activity increases in correspondence with TV shows where politicians from the right-wing alliance are hosted.

comes to substantive electoral themes, instead, the three communities seem to hold a common interest for work-related matters but also concentrate on particular interests: migration flows for the M5S, taxation for the CDX and the role of Europe for the CSX. This finding highlights the twofold nature of the online discussions which center on single personalities/political entities but also, albeit to a lesser extent, on themes of public interest.

In order to identify topics that, regardless of their prominence and persistence, are more pivotal to the unfolding of the discussion, *hashtag betweenness centrality* is computed following this formula:

$$b_\gamma = \sum_{\beta(\neq\alpha)} \sum_{\alpha} \frac{\sigma_\gamma^{\alpha\beta}}{\sigma^{\alpha\beta}} \tag{6.1}$$

where $\sigma_\gamma^{\alpha\beta}$ is the number of the shortest paths between hashtags α and β passing through hashtag γ and $\sigma^{\alpha\beta}$ is the total number of shortest paths between hashtags α and β. In this sense, hashtag betweenness centrality provides an entry point to identify strategic topics that 'coordinate' the discussion. Interestingly, the set of the most strategic hashtags (i.e. *#pd*, *#m5s*, *#renzi*, *#salvini*, *#berlusconi*, *#italia*, *#dimaio*, *#lega*, *#centrodestra*) is basically the same for all communities. This result reveals how the overall discussion is strongly influenced by the main players of the 2018 Italian Elections who embody crucial

H_t	**M5S**	**CDX**	**CSX**
100%	dimaio, lega, renzi, berlusconi, m5s, pd, italia	salvini, m5s, centrodestra, pd, lega	renzi, salvini, dimaio, m5s, pd
98%	forzaitalia, salvini	berlusconi, italia, renzi	
96%	roma, ottoemezzo	forzaitalia	berlusconi, italia, lega
94%	centrodestra, ricercapubblica		russia
92%	boschi, politica	dimaio	europa, politica, roma
90%	fi, governo	fi, governo	
88%	casapound	roma	
86%	meloni		
84%	fakenews, lavoro, liberieuguali	casapound, politica	forzaitalia, lavoro, usa
82%	8800precari, gentiloni, migranti, senato, voto	governo, lombardia	centrodestra, leu, liberieuguali
80%	bonino, campagnaelettorale, casini, leu, rosatellum	cdx, flattax, sinistra	milano, partitodemocratico, ue
78%	avanti, iovotom5s, movimento5stelle, precari, sinistra	lavoro, ue	campagnaelettorale, fakenews, governo

Table 6.1: Hashtag persistence for each discursive community across the entire temporal period covered by the 2018 data set (namely, 51 days in total), on non-filtered projections. The first column shows the percentage of days each hashtag is present in the set of tweets of each community. Notice that the hashtags that are always present are those carrying the name of political parties and political leaders, while other supposedly relevant themes for the political debate are absent from at least some of the discursive communities. These findings suggest that online political debate is largely focused on single personalities/political entities (as particularly evident upon inspecting the CSX hashtags) and only to a much smaller extent on themes of public interest.

concepts addressed by the political debates of all discursive communities. Nonetheless, partisan specificities of each community are maintained when it comes to economic and societal issues.

As discussions develop around 'communities of hashtags', increasingly complex semantic structures are considered. To this end, the presence and the persistence of *triadic closures* (i.e., triangles of connected hashtags) is analyzed. Triadic closures can be seen as the 'seeds' from which more complex discussions grow - exactly as triangular motifs represent the simplest (yet informative) example of communities [127]. As it has been noticed, this kind of structure provides deeper insights into users' tweeting behavior, by revealing which concepts appear *simultaneously* in a discussion and measuring how often they do so [128]. This analysis offers important insights into the behavior of the three communities. As shown in Tab. 6.2, while both the CDX and the CSX communities are characterized by triads of concepts

T_t	**M5S**	**CDX**	**CSX**
31%	(ricercapubblica, 8800precari, campagnaelettorale)		
27%			(salvini, pd, m5s)
24%	(pd, italia, m5s)		(pd, lega, m5s); (pd, dimaio, m5s)
21%	(cnr, campagnaelettorale, ricercapubblica); (precari, campagnaelettorale, ricercapubblica); (politica, pd, m5s); (dimaio, pd, m5s); (lega, pd, m5s)		(m5s, dimaio, salvini); (liberieuguali, pd,m5s); (m5s, berlusconi, pd); (usa, europa, russia); (savona, accettolasfida, poterealpopolo)
20%	(berlusconi, pd, m5s); (ottoemezzo, pd, m5s); (salvini, pd, m5s); (berlusconi, politica, m5s); (centrodestra, pd, m5s); (italia, stopinvasione, italiani); (italia, stopislam, italiani); (campagnaelettorale, piemonte, forzaitalia); (m5s, pd, m5salgoverno)	(salvini, pd, m5s)	(pd, m5s, renzi); (pd, italia, m5s); (salvini, lega, m5s); (forzaitalia, pd, m5s); (fattinonparole, partitodemocratico, avanti); (berlusconi, salvini, pd); (salvini, m5s, berlusconi)
		Dates	
	02 Mar 2018	20 Feb 2018	02 Mar 2018
	20 Feb 2018	27 Feb 2018	06 Mar 2018
	21 Feb 2018	02 Mar 2018	23 Feb 2018
	07 Mar 2018	01 Mar 2018	04 Mar 2018
	16 Feb 2018	22 Feb 2018	21 Feb 2018

Table 6.2: Persistence of triadic closures for each discursive community across the entire temporal period covered by the 2018 data set (namely, 51 days in total) on non-filtered projections. This analysis offers important insights into the behavior of the three communities: while the CSX and CDX communities are characterized by triads exclusively about political leaders, parties and electoral slogans, the triads observed within the M5S community focus more on concepts related to themes of public interest. Notice that the largest T_t value (i.e., the largest percentage of days a specific triadic closure is present) is sensibly lower than the number of days covered by the current data set. Dates refer to the days with the largest number of triadic closures.

exclusively about political leaders, parties and electoral slogans, the triads observed within the M5S community reveal a greater concern for themes of public interest (e.g. the issues of precarious labor, migrants landing, public research).

Interestingly, it is worth noticing on specific days an abundance of triadic closures is registered. For instance, on the first day of electoral silence (i.e., 2 March 2018), users are particularly active in building narratives around electoral slogans, while themes of public interest constitute the topic of tweets at the end of the electoral campaign (i.e., the last days of February). Finally, the abundance of hashtag triads tends to increase in correspondence with mediated events, as observed for the mean degree: this is the case for the days 27 February 2018 for the M5S community (when Luigi Di Maio was interviewed at the political talk show 'diMartedì'), 20 February 2018 for the CDX community (when Silvio Berlusconi was interviewed in a talk show called #Italia18 organized by the Italian newspaper *Corriere della Sera*) and 23 February 2018 for the CSX community (when Laura Boldrini was interviewed on the radio show 'Circo Massimo').

6.1.2 ANND and clustering coefficient

A closer inspection of correlations between hashtag degrees allows a more detailed description of the ways prominent topics are connected to others, shaping the networked framing practices occurring within each discursive community. To this end, the *average nearest-neighbors degree* (ANND) is defined, for a generic hashtag α, as the arithmetic mean of the degrees of the neighbors of a node, i.e.

$$\kappa_\alpha^{nn} = \frac{\sum_{\beta(\neq\alpha)} a_{\alpha\beta}\kappa_\beta}{\kappa_\alpha}, \quad \forall\, \alpha \tag{6.2}$$

with κ_α indicating the degree of hashtag α in the considered monopartite projection. The degree-degree correlation structure of a network can be easily inspected by plotting the κ_α^{nn} values versus the κ_α values. A *decreasing trend* leads to the conclusion that correlations between degrees are *negative* - that is, nodes with a small degree are 'preferentially' connected to nodes with high degree and vice versa. Conversely, an *increasing trend* signals that correlations between nodes are *positive* - that is, nodes with a low (or high) degree are 'preferentially' connected to nodes with a small (or large) degree. Thus, decreasing and increasing trends offer an entry point to explore whether discussions in the three communities tend to anchor onto some key themes that work as conversational drivers. The decreasing behavior of the ANND throughout the present data set confirms the presence of negative degree-degree correlations. In other words, the considered networks are *disassortative* (i.e., less prominent hashtags are connected with more prominent hashtags and vice versa). Examples of the aforementioned trends are reported in fig. 6.2. The days considered here, i.e. 19 February 2018 and 5 March 2018, have been chosen to highlight an interesting feature of the semantic networks: as clearly visible by inspecting the behavior of the CDX and the CSX communities, groups of nodes with larger values of the ANND appear. As will become evident in what follows, these hashtags constitute the *core* of the Twitter discussion in the corresponding community and are characterized by a daily time-scale dynamics: in fact, they appear in correspondence to a specific event (i.e., in the case of the CDX community, the interview of Silvio Berlusconi in a TV show while, in the case of the CSX community, Laura Boldrini's Twitter campaign) and disappear the day after.

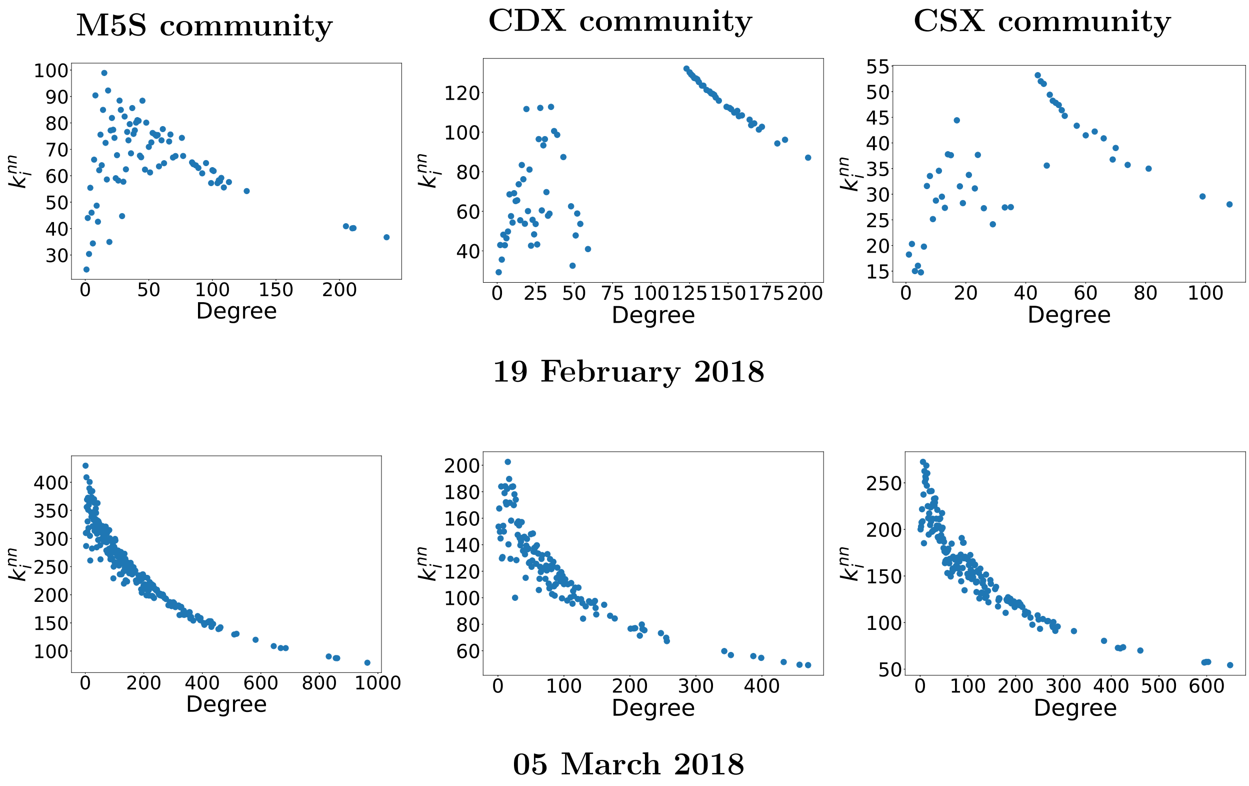


Figure 6.2: Analysis of the degree-degree correlations for two specific days, i.e. 19 February and 05 March 2018, on the non-filtered projections. As the trend of κ_α^{nn} reveals, daily semantic networks are disassortative for all communities, i.e. nodes with small degree are preferentially connected to nodes with high degree and vice versa. As this analysis also reveals, by inspecting the behavior of the CDX and the CSX communities, groups of nodes with a larger value of the ANND appear: these clusters of hashtags constitute the core of the Twitter discussion in the corresponding community, appearing in correspondence to specific events and disappearing the day after.

As an additional analysis, the *clustering coefficient* is also considered and defined as follows

$$c_\alpha = \frac{\sum_{\gamma(\neq\alpha,\beta)}\sum_{\beta(\neq\alpha)} a_{\alpha\beta}a_{\beta\gamma}a_{\gamma\alpha}}{\kappa_\alpha(\kappa_\alpha - 1)}, \quad \forall\, \alpha \tag{6.3}$$

According to its definition, the clustering coefficient quantifies the percentage of neighbors of a given node α that are also neighbors of each other (i.e. the percentage of triangles, having α as a vertex, that are actually present). As shown in Fig. 6.3, decreasing trends are observed: poorly-connected hashtags are strongly interconnected and vice versa, thus suggesting the presence of several interconnected 'small' discussions that are connected to a set of central topics. A network with these features is also said to be *hierarchical*. Furthermore, it is also evident that the hashtags with a higher value of ANND are also those with a higher value of the clustering coefficient - confirming the 'coreness' of this group of topics. Taken altogether, these results suggest that all discursive communities revolve around a handful of a few conversational drivers upon which practices of networked framing are structured: overshadowed by the predominance of these issues, a set of niche discussions tend nonetheless

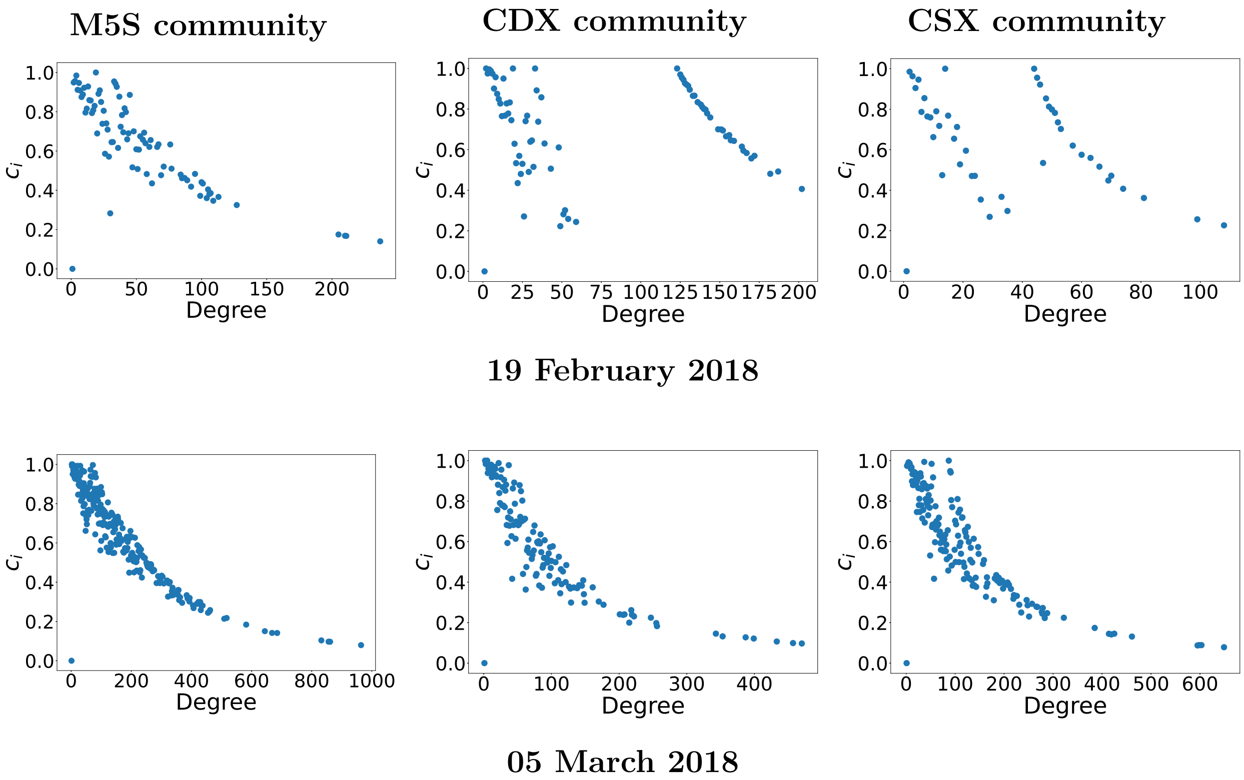


Figure 6.3: Analysis of the networks hierarchical structure for two specific days, i.e. 19 February 2018 and 05 March 2018, on the non-filtered projections. Plotting the clustering coefficient c_α values versus the degree k_α values for the three communities reveals that daily semantic networks are hierarchical, i.e. poorly-connected hashtags are strongly interconnected and vice versa. Furthermore, this plot also shows that the nodes with a larger value of the ANND are those nodes characterized by a larger value of the clustering coefficient.

to emerge, pointing out a variety of interests even within a single discursive community.

6.1.3 k-core decomposition

In order to gain insights into the mesoscopic organization of semantic networks (i.e. into a less trivial dimension of networked framing), the so-called *k-core decomposition* has been employed. Shifting perspective onto the mesoscale structure of semantic network helps to better clarify the power of conversational drivers already identified within discursive communities. The k-core decomposition assigns a 'coreness' score to each node of the network which remains naturally divided into shells whose node coreness is equal to k if the node is present in the k-core of network but not in the $(k+1)$-core. In other words, the k-core decomposition can be described as a sort of pruning process, where nodes that have degree less than k are removed, in order to identify the largest subgraph of a network whose nodes have *at least* k neighbors. This kind of analysis, widely adopted to find the structural properties of networks across a broad range of disciplines including ecology, economics and social sciences [129], eventually partitions a network into shells as the threshold value k varies. In what follows, the semantic analysis is centered on 19 February 2018, but similar considerations

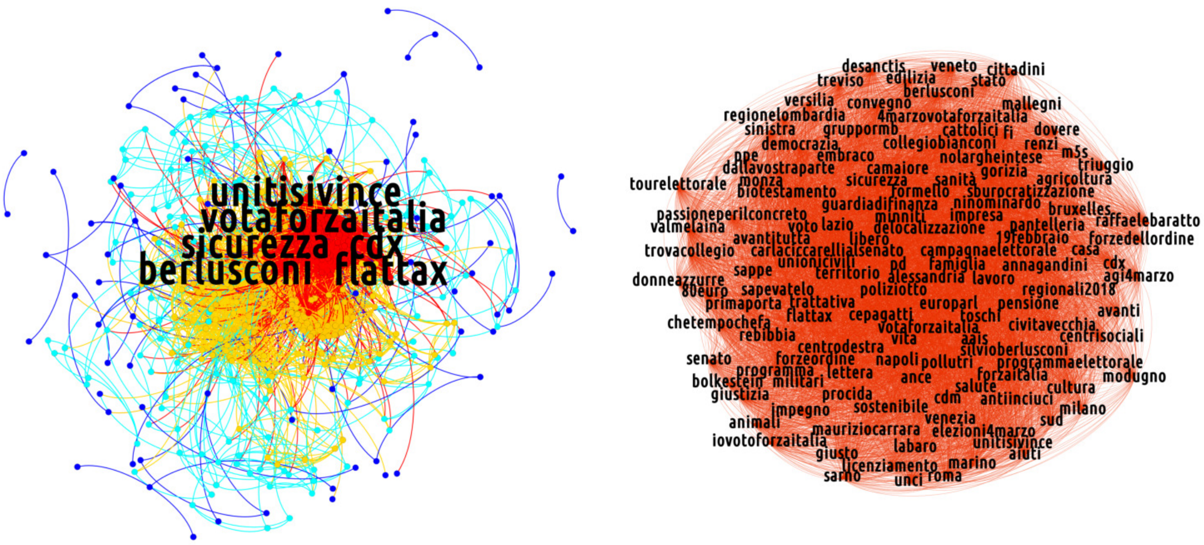


Figure 6.4: k-core decomposition of the semantic network for the non-filtered projection of the CDX discursive community on 19 February 2018. In the left plot, five k-shells for each semantic network are represented with different colors. In the right plot, an expanded view of the innermost k-shell - basically overlapping with the properly defined core individuated by the bimodular surprise - is represented. The compact bulk is triggered by the interview of Silvio Berlusconi in the TV show 'Dalla vostra parte'.

can be made for other daily semantic networks.

Figs. 6.4, 6.5, 6.6 show the k-shell decomposition for semantic networks of each discursive community for the day 19 February 2018: five k-shells, corresponding to four quantiles of the degree distribution, confirm the presence of a core of highly debated hashtags (i.e., the red nodes present in the most prominent and intertwined k-shell) surrounded by a peripheral region, colored in orange, light blue and dark blue.

To inspect the presence of an additional substructure nested in the discussion core, the Louvain algorithm is run on the innermost k-shell of the semantic networks of each discursive community. Their shell structure is indeed rich, as particularly evident when considering the CSX and the M5S discursive communities shown in Figs. 6.5 and 6.6: indeed, several communities appear within their core, seemingly indicating that the discussions in which members of these communities are more engaged self-organize around sub-topics. As a second observation, the communities partitioning the core, when present, are 'held together' by the nodes with the highest values of betweenness centrality: as these nodes are the hashtags related to the name of political parties/leaders, they can be imagined to act as 'bridges' connecting different discussions. Generally speaking, this indicates that the concept of 'most influential nodes' can be applied also within the core of the interactional hashtag networks - a result that complements the presence of influential spreaders individuated within the user networks [130].

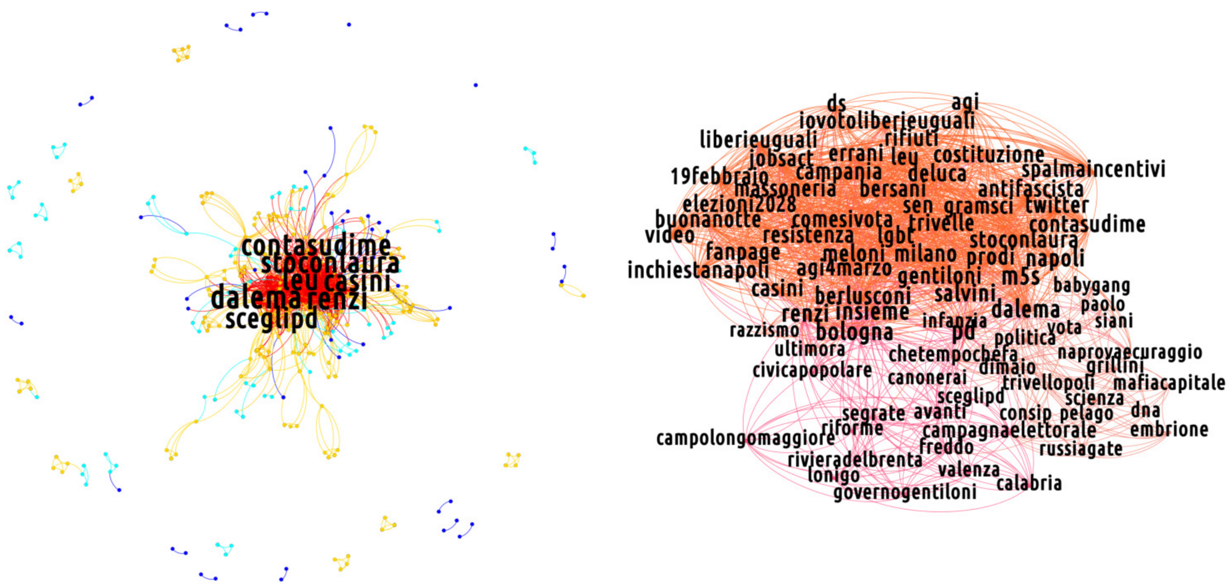


Figure 6.5: k-core decomposition of the semantic network for the non-filtered projection of the CSX discursive community on 19 February 2018. In the left plot, five k-shells for each semantic network are represented with different colors. In the right plot, an expanded view of the innermost k-shell - basically overlapping with the properly defined core individuated by the bimodular surprise - is represented. By running the Louvain algorithm, the presence of communities emerges as a consequence of diverse factors, such as the Twitter campaign born in support of the center-left candidate Laura Boldrini, the visit of Matteo Renzi in Bologna, and the presence of Massimo D'Alema in the radio show 'Circo Massimo'.

As for the CSX community, hashtag sub-communities emerge in connection with the Twitter campaign in support of the center-left candidate Laura Boldrini (revealed by the presence of hashtags such as *#stoconlaura* and *#contasudime*), the visit of Matteo Renzi to Bologna (revealed by the presence of hashtags such as *#bologna*, *#renzi*, *#errani*, and *#casini*, hashtags that refer to Vasco Errani, and Pier Ferdinando Casini, center-left wing candidates for the Senate in Emilia-Romagna), and the presence of Massimo D'Alema, another leader of the center-left alliance, on the radio show 'Circo Massimo' (revealed by the presence of hashtags such as *#dalema*). On the other hand, the presence of multiple debates within the core of the M5S semantic network is related to events like the electoral tour of Alessandro Di Battista, a prominent figure of the party, who presented the M5S electoral program in southern Italy in the region of Basilicata (revealed by the presence of hashtags such as *#dibattista*, *#ilfuturoinprogramma*, *#programmaindiretta*, and *#basilicata*), the presence of a journalist of 'Il Fatto Quotidiano' invited to the TV show 'Otto e mezzo' (revealed by the presence of hashtags such as *#ilfattoquotidiano* and *#ottoemezzo*) and the presence of politicians supporting other coalitions on several TV shows such as 'Porta a Porta', 'Mezz'ora in più' and 'Dalla vostra parte'. However, these observations do not hold true when the CDX induced semantic network is considered. Its innermost shell is, in fact, a compact group of topics that cannot be further partitioned.

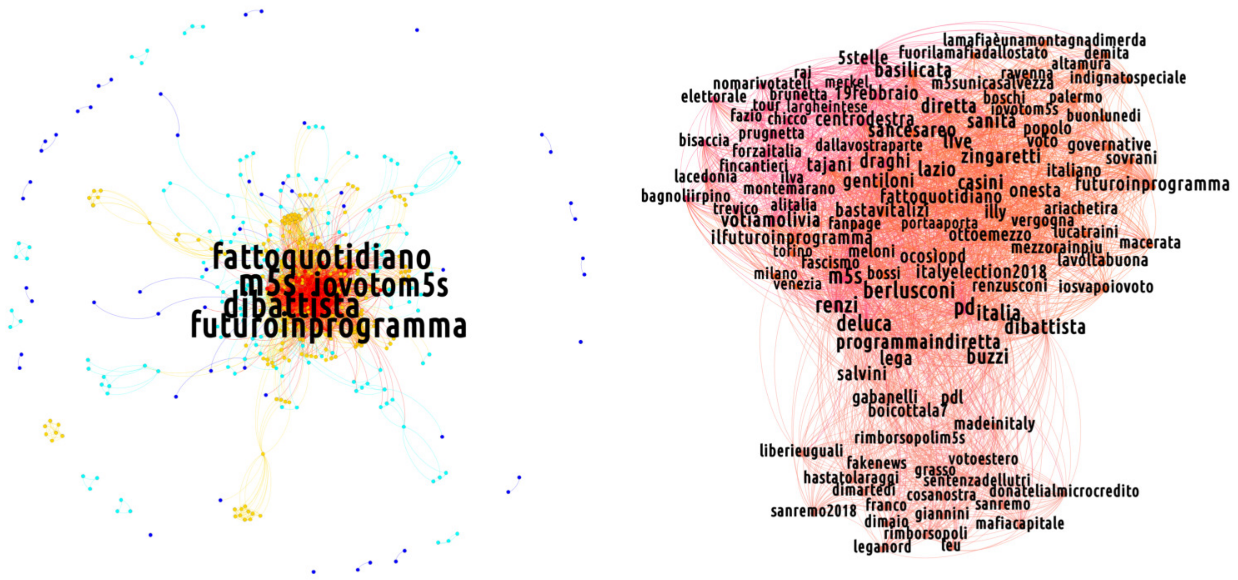


Figure 6.6: k-core decomposition of the semantic network for the non-filtered projection of the M5S discursive community on 19 February 2018. In the left plot, five k-shells for each semantic network are represented with different colors. In the right plot, an expanded view of the innermost k-shell - basically overlapping with the properly defined core individuated by the bimodular surprise - is represented. By running the Louvain algorithm, the presence of communities emerges as a consequence of events as the electoral tour of Alessandro Di Battista and the presence of politicians on TV shows such as 'Porta a Porta', 'Mezz'ora in più' and 'Dalla vostra parte'.

In order to complement this analysis, a core-periphery detection algorithm is implemented. This method, proposed in [131], searches for the network core-periphery partition minimizing a quantity called *bimodular surprise*, described more in detail in Appendix C. The mesoscale structure characterizing all discursive communities consists of a bunch of well-connected nodes linked to a group of low-degree, loosely interconnected nodes, as shown in Figs. 6.4, 6.5, 6.6. This structure is known as *core-periphery* and is present in many social, economic and financial systems [132]. This is particularly relevant as the present core-periphery detection algorithm can be thus applied to detect the presence of statistically-significant core-periphery structures in all kind of network structures, as proposed in [133]. Remarkably, nodes belonging to the innermost shell overlap with the core computed with the multinomial version of the *surprise*, as proven by computing the *Jaccard index* over the two sets of nodes. It is worth noticing that the Jaccard index is a measure of similarity between two sets of elements and is defined as the size of the intersection divided by the size of the union of the two sets: $J(A,B) = \frac{|A \cap B|}{|A \cup B|}$. As a final comment, the evolution of the number of nodes belonging to the core and to the periphery is explicitly shown for each discursive community. As Fig. 6.7 shows, the core size is nearly constant throughout the considered observation period while the periphery size rises in correspondence of the Election Day, showing a peak in correspondence of the day after the elections (i.e. 5 March 2018). This behavior, common

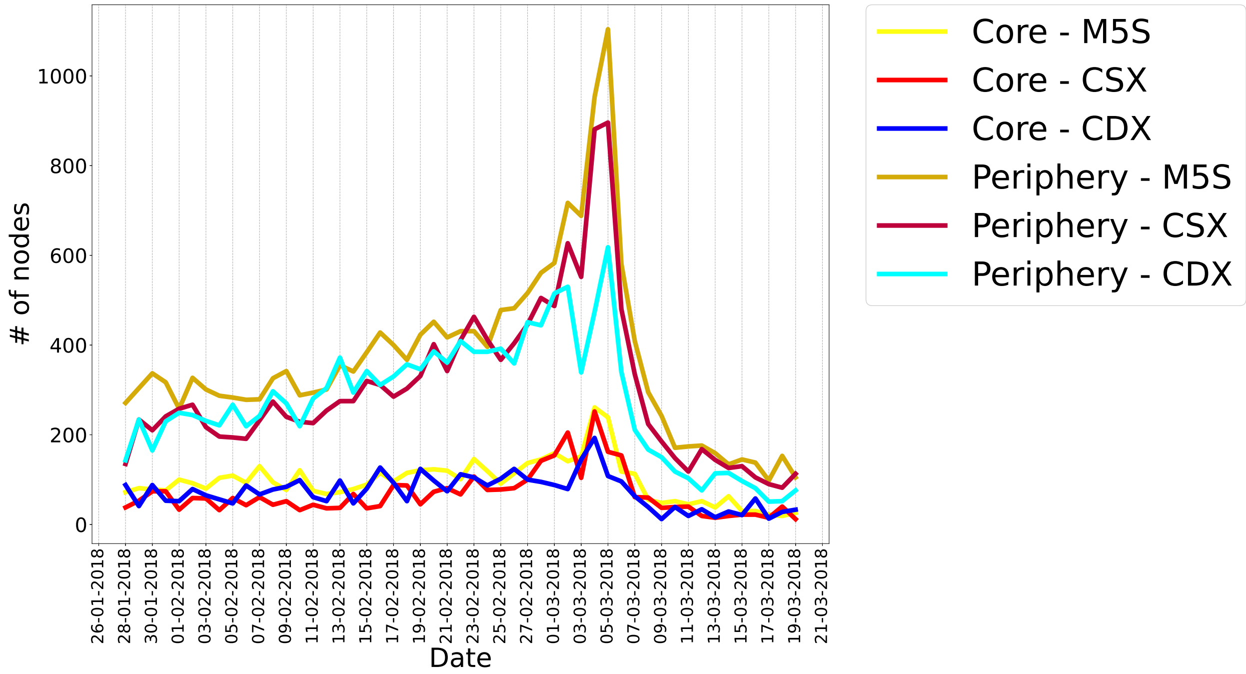


Figure 6.7: Evolution of the number of nodes belonging to the core and to the periphery of each discursive community. The core size is nearly constant throughout the observation period while the periphery size rises as the Election Day approaches (the peak appears in correspondence to the day after, i.e. the 5 March 2018). This behavior, common to all communities, reveals that, as the Election Day approaches, the number of peripheral topics animating the discussion increases within the framing dynamics.

to all communities, indicates that, as the Election Day approaches, the number of peripheral topics discussed and, hence, concurring to framing dynamics does increase.

6.1.4 Filtering the projection

In the present work, the final analysis on daily semantic networks related to the 2018 Italian elections is the analysis of the structural features of filtered projections. Before presenting these results, it is worth recalling briefly how the filtering procedure works. Filtering allows the detection of statistically significant co-occurences of hashtags measured on the initial bipartite network. More in detail, for each couple of hashtags, the number of users who employ that specific couple is counted. Then, an ensemble of networks preserving on average a specific information of the initial user-hashtag bipartite network is considered as a benchmark. The information preserved in the ensemble are embedded in the constraints describing different null models, as shown in Appendix B: the BiRGM preserves the total number of edges, the BiPCM preserves the degree sequence of the hashtag layer, and the BiCM preserves the degree sequence of both layers. The stricter the constraints (i.e. the properties preserved by the ensemble), the more detailed the description of the ensemble, as compared to the real network, and fewer edges are validated. In this sense, the number of validated co-occurences are those non compatible with the expected structural features of the different null models. In other words, the validation procedure retains the non-trivial co-occurences of hashtags present in the initial bipartite structure, namely those that are not explained by the constraints of the null model used for filtering. Among the aforementioned null models, the BiRGM is expected to retain the highest number of edges between hashtags while the BiCM is the most restrictive among the filtering null models.

Summarizing, the purpose of these projections is to retain an edge between two hashtags if their number of co-occurences is significantly larger than expected by the strucutural network properties constrained under any of the chosen null models. A greater number of hashtags in the validated projection does not imply a greater activity, but simply a more focused activity of the users within a discursive community. That is, more validated nodes mean that, a larger number of hashtags is jointly used in order to discuss a specific topic. Against this background, filtered projections will be employed to study the different conversational lines shaping networked framing pratices of the three discursive communities.

As mentioned above and as Figs. 6.8, 6.9, 6.10 show, the overall effect of adopting a filtering procedure - regardless of its peculiarities - is that of reducing the total volume of nodes of the semantic networks. In this Section, while the filtering procedure can be applied to each monopartite projection, I will discuss the mesoscale structure of the filtered projections looking, as previously done, at the specific day that shows the richest structure (i.e., 19 February 2018). This particular choice can be thought of as a solid starting point for extracting insightful information when recognizable mediated events are present within the filtered projections. Moreover, in this specific date, almost all hashtags representing topics of interest of the 2018 Italian electoral campaign persist. The same conclusion holds true for the number of triadic closures observed in correspondence to mediated events: their number is significantly larger with respect to a network model constraining the total number of edges only. This result is in line with what has been found for other socio-economic systems (e.g. the World Trade Web [96]) whose abundance of triadic closures is not reproduced by the BiRGM.

The CDX discursive community. Fig. 6.8 depicts the semantic network of the center-right alliance on 19 February 2018. In the BiCM projection, few edges survive. In this situation, it becomes almost inappropriate to talk about communities of hashtags, since it is possible to find only edges connecting two otherwise isolated nodes, or small cliques and chains. Nevertheless, even these few hashtags carry important information regarding the keywords used during the electoral campaign. In the CDX community, this is the case of the cluster including *#stopislam*, *#stopinvasione* (*stop the invasion*), *#cdm* (i.e., the acronym for the Italian Council of Ministers) and *#forzeordine* (*law enforcement agencies*), asking for stronger countermeasures to the migration flows from Northern Africa, perceived as a danger for the security and for Italian cultural identity. On a similar topic, there is a clique composed by *#roma*, *#labaro* and *#primaporta*: the hashtags refer to neighborhoods in Rome, in which, during the days of the data collection, some thefts in apartments were reported. These hashtags were used to criticize the city administration of Rome, run by Virginia Raggi of the 'Movimento 5 Stelle'. Moreover, a pair of nodes which represents insulting nicknames for the rivals are connected between themselves. Those hashtags, *#pdioti* (i.e., an hashtag mixing the acronym of the *'Partito Democratico'* and the word *idiots*) and *#m5stellisti*, are present in a popular message displaying both hashtags. There is also a clique formed by *#casapound* (i.e., a neo-fascist party), *#rai2* (i.e., the second channel of the national TV public service) and *#19febbraio*. This clique is the result of a viral tweet intended to advertize the presence of the leader of Casa Pound in a public debate held on Rai2 on 19 February. Finally, the last clusters present in the BiCM induced projection are more institutional: the first contains *#torniamoagovernare* (*let's go back to governing*), *#elezioniregionali2018* (*2018 regional administrative election*) and *#salvini*, while the other one is composed by *#flattax*, *#programma* (*program*) and *#veneto*. The latter set of hashtags is related to an event where the flat taxation government, as part of the electoral program, is presented.

The BiPCM projection displays a structure in which the various sub-groups described above are reinforced (for instance, the set consisting of *#flattax*, *#programma* and *#veneto* is closed in a clique) and introduces new topics such as *#calenda* (i.e., the Minister of the industrial development at the time of the electoral campaign) *#ilva* and *#alitalia*, respectively the greatest European steel factory which had severe problems of environmental, health and economic sustainability, and the Italian national airline, which has been at default risk in recent years. These hashtags are intended to criticize the action of the government in charge at that time. Interestingly, another cluster consisting of the hashtags *#sullepalle*, *#fazio* and *#salvini* is detected by the validated BiPCM projection. These hashtags need a bit of context: during the electoral campaign, the journalist Fabio Fazio invited politicians from all political coalitions to his TV program 'Che tempo che fa' broadcast on the national television service, to promote their campaign. Fazio has been accused by all political forces of being too condescending with their opponents. Salvini refused Fazio's invitation, publicly with insulting language his aversion for the journalist. These hashtags, together with others related to right-wing campaign topics such as *#vita* and *#famiglia* (*life* and *family*, both related to the Italian anti-abortion movement) are associated with the communication strategy of the most radical part of the center-right alliance.

In the BiRGM validated projection, the clusters found in the previous stricter projections are merged together to form a network organized along two poles: the first is more 'institutional' with keywords related to the electoral campaign of 'Forza Italia', including hashtags such as *#campagnaelettorale* (*electoral campaign*), *#unitisivince* (*united we will win*), *#votaforzaitalia* (*vote Forza Italia*); the second is linked to the other two right-wing parties with both the names of their leaders (*#salvini* and *#meloni*), but also including their opponents, as *#pd*, *#renzi*, *#pdioti* and *#m5stellisti*. Interestingly, both poles are organized in a core-periphery fashion: the two cores are connected by the hashtag *#centrodestra* (*center-right*),

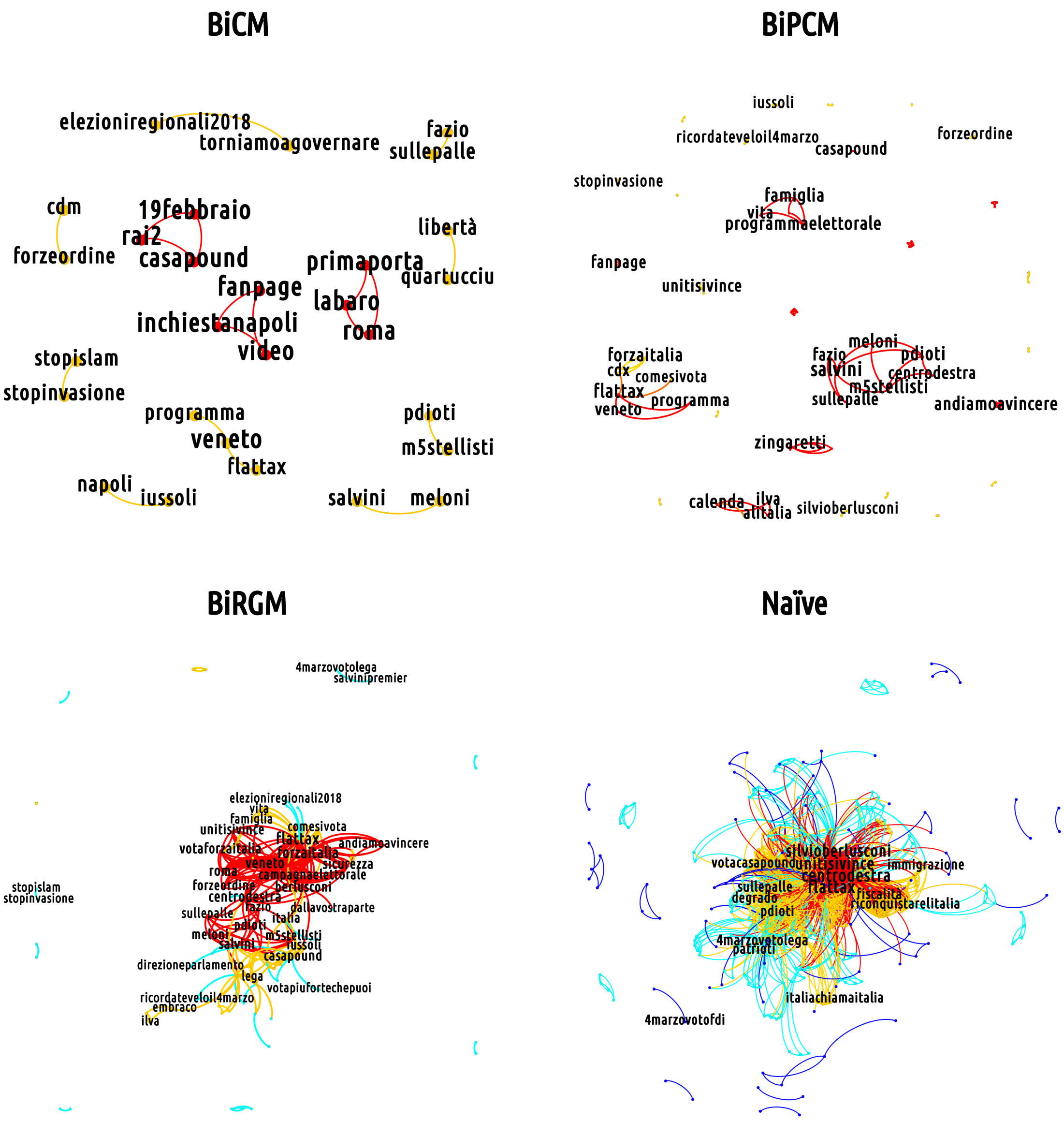


Figure 6.8: Filtered and non-filtered monopartite projections of the CDX induced semantic network on 19 February 2018. Mesoscale structure of the non-filtered projection of the semantic network corresponding to the CDX discursive community on 19 February 2018 and of the projection of the same network filtered according to the BiRGM, the BiCM and the BiPCM, respectively. These projections are presented from top left to bottom left in the clockwise direction. In the BiCM projection, only *few* hashtags survive, i.e. *#iussoli*, *#sicurezza*, *#stopinvasione*, *#stopislam*.

the peripheries by *#casapound* (i.e., the aforementioned neo-fascist party).

Summarizing, in the CDX community, a clear thematic distance is present between the far-right leaders Matteo Salvini and Giorgia Meloni and the center-right politicians (i.e., Silvio Berlusconi and members of his party ‘Forza Italia’) in terms of topics and electoral slogans promoted by these two poles. While the former insists on security issues related to migration flows from Northern Africa, the latter tends to promote a united center-right alliance. The qualitative description of the groups of hashtags within the filtered projections shows an evident semantic diversification with completely different keywords used in the tweets: the far-right uses more aggressive statements and vituperative language, while the attitude of the center-right is more reassuring and institutional.

The M5S discursive community. Overall, the communication strategy of the M5S is peculiar since its users tend to discuss a higher number of topics with respect to the other discursive communities. The evidence of a larger amount of validated hashtags can be explained by a higher number of retweeted messages that are able to generate peculiar hashtag co-occurences. In this sense, Twitter users in the M5S community appear to be more coordinated and thus manage to give their hashtags more visibility.

Interestingly, the core portion of the semantic network corresponding to the M5S discursive community survives the BiCM filtering, signalling the presence of a non-trivial group of keywords constituting the bulk of communication in that community. Considering the tweets and retweets published on 19 February, the M5S validated semantic networks of Fig. 6.9 display a rich structure, even in the BiCM projection, due to the tweeting behavior described above. In particular, several clusters can be found, including the name of the opponents (*#renzi*, *#salvini*, *#gentiloni*, *#pd*) or a few nicknames assigned to them (i.e., *#prugnetta*, *little plum*, for Renato Brunetta, a member of 'Forza Italia', and *#renzusconi*, a mix between Matteo Renzi and Silvio Berlusconi, intending that there is little difference between the two of them), and other slogans taunting rivals (*#votiamolivia*, *let's vote them away*; *#nomarivotateli*, *no, but vote them again*, ironically targeting the former voters of 'Partito Democratico'; *#ocosìopd*, *this way or 'Partito Democratico"s way*, advertising that the only political alternative to 'Partito Democratico' is the 'Movimento 5 Stelle'). A few clusters represent political events in the electoral campaign. For instance, a cluster following the electoral campaign tour of Alessandro Di Battista, a key member of the 'Movimento 5 Stelle', appears in this projection. Even a clique advertising a live streaming on Facebook can be observed, discussing the management of the public health system in the Lazio region governed by the 'Partito Democratico', with the hashtags *#lazio*, *#sanità* (*public health system*), *#sancesareo* (i.e., the town were the live streaming was set), *#zingaretti* (i.e., Nicola Zingaretti who is the president of the Lazio region).

The topic of bad governance of political opponents represents a big part of the semantic network of the M5S community: in addition to the cluster mentioned above, another cluster focuses on the news about a journalist who was attacked during a campaign event held by the center-left coalition in Naples (*#fanpage* which is the online newspaper for which the journalist worked; *#inchiestanapoli*, *Naples investigation*; *#video*). Moreover, the hashtags *#donatelialmicrocredito* (*give them to the microcredit*) and *#rimborsopoli* (*refund scandal*) refer to the scandal of a criminal organisation in Rome bribing members of established

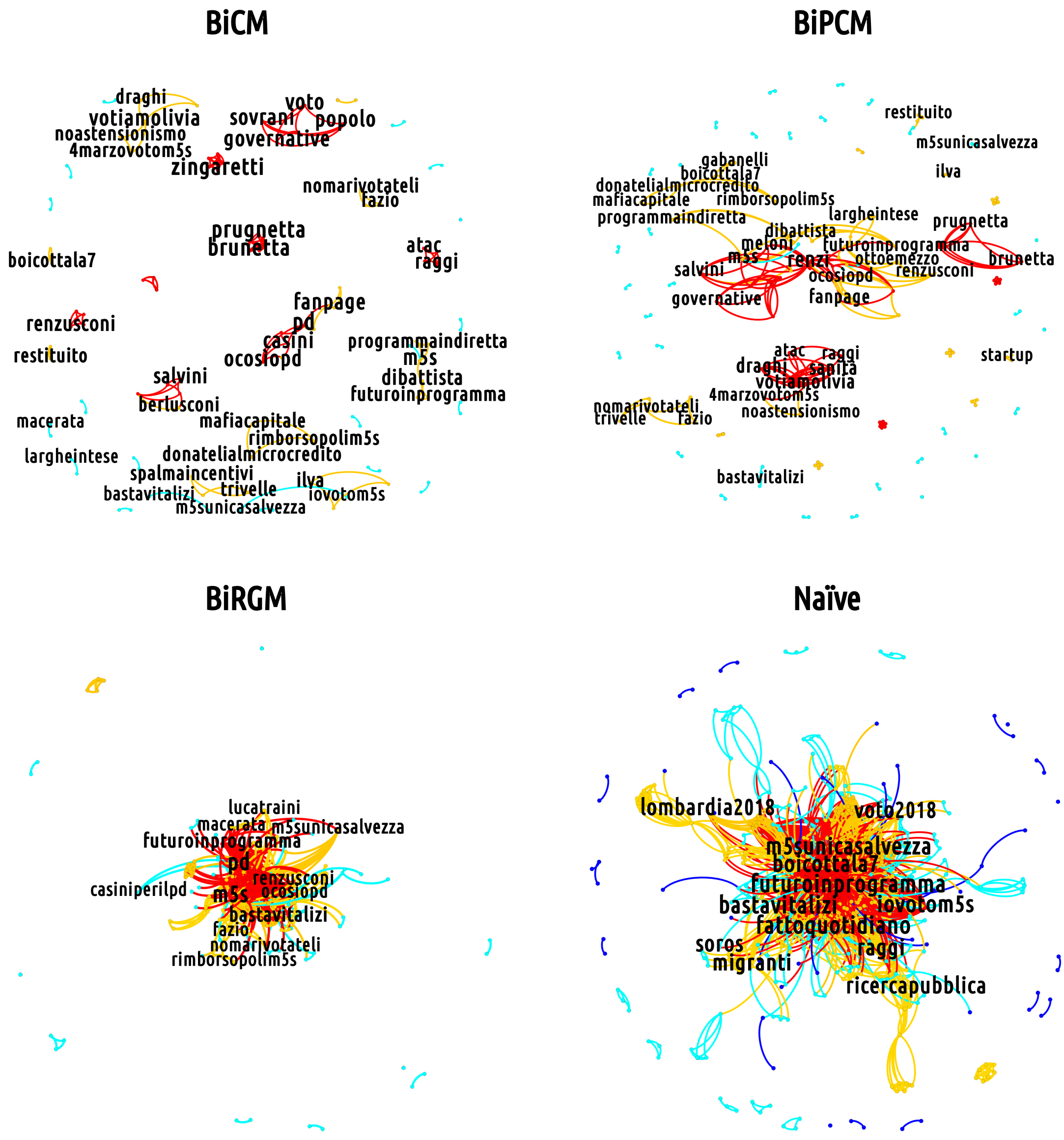


Figure 6.9: Filtered and non-filtered monopartite projections of the M5S induced semantic network on 19 February 2018. Mesoscale structure of the non-filtered projection of the semantic network corresponding to the M5S discursive community on 19 February 2018 and of the projection of the same network filtered according to the BiRGM, the BiCM and the BiPCM, respectively. These projections are presented from top left to bottom left in the clockwise direction. The core portion of the semantic network survives the BiCM projection, indicating that basically all hashtags representing topics of interest of the 2018 Italian electoral campaign persist.

political parties. The 'Movimento 5 Stelle' expelled its representatives involved in this investigation and proposed that other parties do the same and transfer the amount of the bribe to the microcredit. There are also clusters targetting harsh political debates, such as the case of *#dibiase*, referring to Letizia Di Biase who is the wife of the Italian Minister of Cultural Heritage and Activities. After being elected member of the council in the city of Rome, she did not resign when elected as member of the council of the region of Lazio. Di Biase was also criticised for criticizing the mayor of Rome Virginia Raggi of the 'Movimento 5 Stelle', for filing for bankruptcy for the city agency of public transportation while salvaging the regional one operated by the regional administration of the 'Partito Democratico'. Finally, there are traces of the debate between the virologist Roberto Burioni and the Head of the Italian Order of Biologists, Vincenzo D'Anna, concerning the presence of anti-vaccine groups and 'Movimento 5 Stelle' supporters during a national conference of the order of biologists. Other hashtags refer to Giorgia Meloni, leader of 'Fratelli d'Italia', and the charges moved against her for her sympathy for neo-fascist parties and ideology. The BiPCM projection increases the connections among the topics and a few hashtags appear to be related to names of places covered by the campaign tour of Carlo Sibilia, another member of the 'Movimento 5 Stelle'. On the other hand, the BiRGM projection displays a core-periphery structure with a strong structural proximity to the naïve projection, as pointed out by their similar trend in the temporal evolution of the mean degree in Fig. 6.1.

Summarizing, the M5S filtered semantic networks are especially rich in structure with a great number of hashtags employed in this community. Most of them refer to political opponents with nicknames and ironic slogans. A great part of the filtered semantic network highlights the deceitfulness of the political opponents of the 'Movimento 5 Stelle' while other clusters display harsh discussions, such as that against the use of vaccines or against an alleged *quid pro quo* between members of the 'Partito Democratico' and businessmen.

The CSX discursive community. In the case of the CSX community, as shown in Fig. 6.10, the number of hashtags used is relatively small. Compared to other discursive communities, the semantic network of the center-left group validated by the BiCM focuses more on political topics, as proven by the pair *#diritti* (*rights*) and *#arcobaleno* (*rainbow*) which refer to LGBT civil rights, but also by the coupling of *#bimbi* (*children*) and *#rohingya*, both related to the topic of the Rohingya exodus in Myanmar and the condition of children. Other clusters are related to instructions for youngsters voting for the first time (*#primovoto*, *first vote*; *#comesivota*, *how to vote*; *#pernonsbagliare*, *how to avoid mistakes*) and calling for fact checking during the electoral campaign, with hashtags *#factchecking* and *#checkpolitiche2018*. Interestingly, a clique is formed by hashtags of a single popular tweet *#trivellopoli*, *#mafiacapitale* and *#consip*, i.e. three scandals in which the 'Partito Democratico' was involved. This tweet suggests that those scandals suspiciously appeared during the electoral campaign in order to damage the name of the 'Partito Democratico' thereby limiting its performance at the general elections. Another conversational line unfolds along the candidacy of Paolo Siani, a physician particularly active in providing support, in collaboration with local NGOs, to children of the poor neighborhoods of Naples at risk of being recruited for organized criminal activities. More broadly, the public presentation of the 'Partito Democratico' candidates team constitutes a topic widely debated within this discursive community as shown by the two hashtags *#renzi* and *#gentiloni*, respectively the National Secretary of the 'Partito Democratico' and the candidate for the position of Prime Minister,

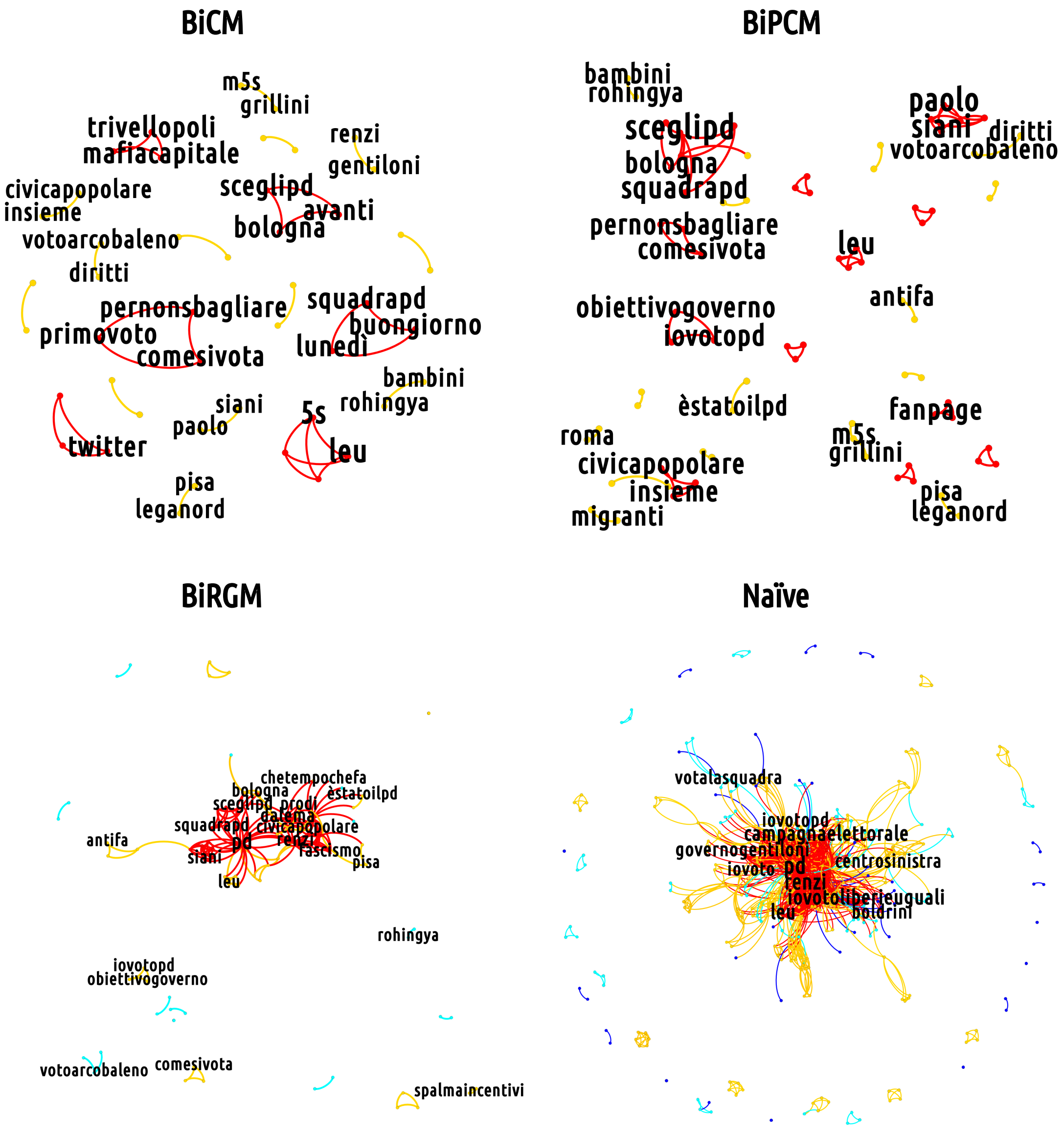


Figure 6.10: Filtered and non-filtered monopartite projections of the CSX induced semantic network on 19 February 2018. Mesoscale structure of (from bottom-right, clockwise) the non-filtered projection of the semantic network corresponding to the CSX discursive community on 19 February 2018 and of the projection of the same network filtered according to the BiRGM, the BiCM and the BiPCM, respectively. These projections are presented from top left to bottom left in the clockwise direction. The core portion of this network partially survives the BiCM projection while it is present in the BiPCM and the BiRGM induced-projections representing a structure between the M5S (Fig. 6.9) and CDX induced (Fig. 6.8) semantic networks.

in connection with other hashtags. For instance, the clique *#bologna*, *#avanti* (*let's move forward*) and *#sceglipd* (*choose 'Partito Democratico'*) refers to an event led by Paolo Gentiloni and Matteo Renzi in Bologna while the clique *#lunedì* (*Monday*), *#buongiorno* (*good morning*) and *#squadrapd* (*'Partito Democratico' team*) appeared in a message promoting a massive electoral campaign.

In the BiPCM validated projection more connections appear, enriching topics which had already emerged, as in the case of the candidacy of Paolo Siani mentioned above: *#baby-gang*, *#napoli* (*Naples*), #infanzia (*childhood*) merge with the previous hashtags *#paolo* and *#siani*. A new cluster containing the name of the opponents (*#dimaio*, *#salvini*, *#meloni*, *#fascismo*) is also present. In the BiRGM validated projection, the aforementioned structures gain new edges and new nodes and a richer structure becomes evident. In particular, three main communities of hashtags appear: the first (in orange in Fig. 6.3) pivoting around political adversaries, as revealed by the hashtags *#salvini*, *#meloni*, *#dimaio*, *#grillini*, *#berlusconi*; the second advertising political subjects and events of the electoral campaign, as revealed by the hashtags *#sceglipd* (*choose 'Partito Democratico'*), *#squadrapd* (*'Partito Democratico' team*), and*#diritti* (*rights*); and the third concerning the candidacy of Paolo Siani. A peripheral clique can be found with the hashtags *#antifa*, *#liberieuguali*, and *#venezia* (*Venice*) advertising the electoral event in Venice of 'Liberi e Uguali'.

Summarizing, filtered semantic networks of the CSX community are less rich than those of the M5S community, but more than those of the CDX community. Their major feature is to present mostly the events of the electoral campaign of the center-left parties, their candidates at a national and regional level and the weaknesses of the political opponents.

Final remarks on the filtering procedure. In all naïve projections a rich structure is observed, with a particularly evident core-periphery organisation. This structure is progressively disintegrated through different filtering techniques according to the strictness of the benchmark used. While this disintegration affects semantic structures generated by all discursive communities, each community displays a different 'resilience' to the filtering procedure. In the present analysis, filtering the projections by adopting an increasingly restrictive benchmark has the effect of 'sparsifying' the projection while letting the less trivial structures emerge [97, 98]. This technique allows the detection not only of *extremely popular hashtags* but also of those hashtags connected only to a single message retweeted by a significant number of users. The former, in fact, are single hashtags tweeted by a huge amount of users and therefore having a large number of co-occurences, a feature that is compatible with at least one of the null models considered here and thus possibly filtered out by this procedure. The latter, instead, are more likely to be groups of hashtags whose non-trivial co-occurrence will survive the filtering procedure.

As an example, a set of hashtags used together only in one tweet can be considered: if this particular tweet is retweeted a significantly high number of times, the number of co-occurences of these hashtags appearing *once* is given by the number of times the original message has been retweeted plus the contribution of the original message. In this case, as the null model distribute the co-occurence probability among all the other hashtags in the semantic network, the probability for their co-occurence to be validated thus becomes larger. Differently, a popular hashtag published by a large number of users has a higher probability to co-occur with other hashtags and, at which point the null model could explain its co-occurrence with these hashtags as induced by the network properties. In other words, the filtering algorithm presented in this work is more likely to discard co-occurences with popular hashtags than with these hashtags appearing together in a single tweet shared a significant number of times. In fact, as the retweeting mechanism has been shown to enhance collective identities [60], the co-occurence of hashtags in the latter case is considered as more informative in the development of proper *narratives* within the networked framing dynamics.

Given this filtering mechanism, the different network structures carry information about the electoral strategy followed by political parties or coalitions that are present within the discursive communities. The validation procedure allows the semantic analysis to disclose the least trivial connections, thus uncovering otherwise invisible conversational lines shaping the networked framing occuring within discursive communities. Differences between discursive communities emerge from the analysis of the filtered semantic networks, in particular when their nodes' mean degree is analyzed. A remarkable example in this respect is the behavior of the semantic networks of the M5S discursive community whose mean degree is affected to a much lesser extent by the BiRGM induced filtering than those of the CDX and the CSX communities. This, in turn, implies that the M5S community presents the least trivial semantic structure: in fact, the M5S bipartite user-hashtag network requires less trivial information that needs to be explained than that of the CDX and the CSX configurations. Similarly, it is possible to state that the information encoded into the BiRGM (which is the simplest filter) recognizes the initial structure of the M5S bipartite user-hashtag network as significant. As for the persistence of topics, the ranking observed on the non-filtered projection basically coincides with the ranking observed on the filtered projections. Regarding topics centrality, instead, results show that the filtering procedure with increasingly restrictive benchmarks leads to the 'emergence' of previously unscreened hashtags (e.g. *#sicurezza*, *#fallimentocinquestelle* and *#precariato*, respectively for the semantic networks induced by the CDX, CSX and M5S discursive communities). Betweenness centrality is, in fact, a highly non-trivial feature that, generally speaking, is reproduced neither by the information encoded into the total number of edges nor by that encoded into the degree sequence: in fact, as Figs. 6.8, 6.9, 6.10 clearly show, only the hashtags belonging to the innermost shells survive the present filtering procedure.

It is worth noticing that the peculiarities of the filtered semantic networks are present on other days which are not explicitly commented on, as the present analysis is focused on particularly illustrative examples occurring on specific days. For instance, on 11 February 2018, two different poles of the debates are still present in the CDX community, one promoted by the supporters of the center-right party 'Forza Italia' and the other promoted by the supporters of far-right wing parties. As observed for 19 February, the two poles use different vocabularies and focus, respectively, on reforming taxation and labor and on the migration issues. In all three networks there are also mentions of the demonstration involving nearly thirty-thousand people against neo-fascism held in Macerata on 10 February [115], albeit with different levels of attention.

6.2 Monthly semantic networks of the 2019 Italian migration discussion

In this Section, the topological features of the semantic networks for each discursive community for the 2019 Italian migration discussion will be investigated on a monthly time-scale. The aim of the following semantic analysis is twofold: on the one hand, observing the evolution of the conversational dynamics when considering a Twitter discussion that centers on a broader societal discussion; on the other hand, determining if the results found in daily semantic networks pertaining to national elections can be reproduced in a similar fashion by modifying the data aggregation time-scale. This procedure is thus proposed as suitable to monitor the evolution of Twitter discussions that are broader in scope as they are not centered around a specific event such as elections.

6.2.1 Analyzing topics centrality and mesoscale structures

Unlike the 2018 data set, the monthly semantic networks are limited to a small set of seven networks. In this context, hashtag persistence H_t as defined in the previous Section works less effectively as a measure for quantifying the prominence of hashtags in the networked framing given that each semantic network represents a picture with less detailed information. In order to identify the most relevant topics and slogans, the hashtag betweenness centrality is thus computed. In fact, this measure can be considered a proxy for the 'conductivity' [134] of those concepts functioning as bridges between communities of hashtags.

The comparison between most conductive hashtags for each discursive community confirms the results that have emerged in the study of communities present in the 2018 data set. First, as Table 6.3 shows, the hashtag *#salvini* is steadily among the first three most conductive hashtags in both the CSX and DX communities across the entire observation period - therefore, also after Salvini's exclusion from office. Thus, the two communities engage in discussions that personalize the debate on migration, making the controversial figure of Matteo Salvini central from a semantic perspective as well. While, at first sight, this result may suggest a semantic alignment between the two communities, a closer look at other conductive hashtags helps to specify that the common reference to Salvini provides the baseline for sustaining polarized positions with respect to migration issues. This is well exemplified by hashtags recalling political slogans: while in the DX community the hashtag *#portichiusi* (*closed ports*) remains present until the falling of the 'Lega'-'Movimento 5 Stelle' government, the CSX community repeatedly encourages undertaking collective actions via the hashtag *#facciamorete* (*let's act as a network*) and adopts the counter-hashtag *#portiaperti* (*open*

CSX						
2019-05	2019-06	2019-07	2019-08	2019-09	2019-10	2019-11
salvini	salvini	carolarackete	openarms	libia	lampedusa	italia
facciamorete	libia	decretosicurezzabis	salvini	marejonio	libia	libia
europa	giornatamondialedelrifugiato	salvini	decretosicurezzabis	salvini	erdogan	oceanviking
libia	lavoro	seawatch3	facciamorete	oceanviking	salvini	lavoro
milano	italia	italia	lampedusa	lavoro	italia	facciamorete
seawatch3	seawatch3	libia	gregoretti	europa	europa	sicurezza
sicurezza	facciamorete	lampedusa	libia	lampedusa	3ottobre	salvini
marejonio	innovazione	facciamorete	oceanviking	facciamorete	malta	formazione
lampedusa	roma	mediterranea	ong	italia	sostenibilità	portiaperti
lavoro	lampedusa	ong	ai	malta	cloud	europa
DX						
2019-05	2019-06	2019-07	2019-08	2019-09	2019-10	2019-11
salvini	salvini	salvini	openarms	salvini	salvini	salvini
italia	fico	lampedusa	salvini	italia	malta	italia
governo	italia	seawatch3	lampedusa	ong	italia	ong
libia	libia	italia	ong	portichiusi	lampedusa	lamorgese
portichiusi	ong	ong	italia	dalleparoleaifatti	oceanviking	libia
m5s	lampedusa	francia	bibbiano	pd	lamorgese	lega
pd	portichiusi	bibbiano	portichiusi	migrante	conte	governo
seawatch3	pd	carolarackete	pd	conte	giustizia	conte
marejonio	europa	portichiusi	richardgere	lampedusa	trieste	italiani
europa	giustizia	libia	gregoretti	macron	mattarella	oceanviking

Table 6.3: Ranking of the ten hashtags with the highest values of betweenness centrality - CSX and DX discursive communities. Some of the DX community hashtags as *#portichiusi* (*#closed ports*) denote a clear anti-migration position within this community while the CSX community is characterized by slogans such as *#portiaperti* (*#open ports*) openly promoting pro-migration positions.

ports) after the 'Partito Democratico' replaces the 'Lega' in the second government led by Giuseppe Conte.

The hashtags displayed in Table 6.4 are useful for gaining insights into the semantic positioning of the other two main discursive communities that emerged in the 2019 data set. Interestingly enough, consistent with its shifting governmental alliances, the M5S community appears to be 'semantically torn'. On the one hand, users in this community focus their attention on calls to collective action coming from the CSX community, as shown by the transversal adoption of the hashtag *#facciamorete*, and by the claims for the liberation of Captain Rackete (*#freecarola*). On the other hand, M5S members ask to *stop immigration* (*#stopimmigrazione*), supporting Matteo Salvini's positions on the topic and raising concern against the 'Partito Democratico' - particularly after its involvement in an alleged scandal about permits for placement of minors in foster care in the city of Bibbiano (*#bibbiano*). Taken altogether, these elements suggest that the M5S community does not hold a position as polarized as that of the DX and the CSX discursive communities on migration issues. However, the M5S community started distancing itself from the DX community since the government crisis in August 2019: the relevance of this political event for the M5S is shown by the presence of hashtags that refer to the crisis, as *#crisidigoverno* and *#governoconte2*.

M5S						
2019-05	2019-06	2019-07	2019-08	2019-09	2019-10	2019-11
ricerca	salvini	salvini	facciamorete	governo	quota100	pattoper-laricerca
salvini	disastrocalenda	pd	m5s	governoconte2	scuola	governo
m5s	m5s	vonderleyen	italia	conte	turchia	libia
seawatch3	piazzapulita	fico	salvini	lega	salvini	bellanova
bergamo	seawatch3	freecarola	boldrini	salvini	malta	zaiadimettiti
skytg24	governo	democrazia	lega	legammerda	disperati	leggedibilancio
governodelcam-biamento	lampedusa	facciamorete	decretosi-curezza	m5s	giule-manidaroma	lega
berlusconi	21giugno	bibbiano	crisidigoverno	lamorgese	lega	salvinivergog-nati
precari	salviniusa	quota100	salvinicazzaro	stopimmi-grazione	precari	salvini
roma	dimaio	ong	salvinitraditore	oceanviking	10ottobre	lamorgese
MINGOs						
2019-05	2019-06	2019-07	2019-08	2019-09	2019-10	2019-11
libia	giornatamondi-aledelrifugiato	seawatch3	openarms	europa	lampedusa	libia
europa	papafrancesco	libia	libia	libia	3ottobre	italia
lavoro	libia	lampedusa	lampedusa	lampedusa	migrants	31ottobre
papafrancesco	inclusione	carolarackete	europa	rohingya	manovra	europa
1maggio	sostenibilità	papafrancesco	rohingya	mediterraneo	welfare	lavoro
salvini	lampedusa	italia	gregoretti	oceanviking	europa	papafrancesco
lampedusa	italia	ioaccolgo	mediterraneo	italia	15ottobre	caporalato
opportunità	ioaccolgo	rohingya	decretosicurez-zabis	marejonio	libia	15novembre
rohingya	europa	europa	salvini	conte	papafrancesco	corridoiumani-tari
marejonio	rohingya	26giugno	tunisia	papafrancesco	gruppohera	oceanviking

Table 6.4: Ranking of the ten hashtags with the highest values of hashtag betweenness centrality - M5S and MINGOs discursive communities. The M5S community presents a peculiar mix of hashtags characterizing both the CSX (as #*facciamorete*) and the DX community (as #*bibbiano*), along with original hashtags referring to the governmental crisis in August 2019 and the formation of the new government in September 2019 (as #*governoconte2* and #*salvinitraditore*); on the other hand, the MINGOs community is characterized by an evident support for specific pro-migration topics and slogans (as proven by hashtags such as #*ioaccolgo*).

Moreover, as the 'Movimento 5 Stelle' negotiates with the 'Partito Democratico' to form a new governmental coalition, users of the M5S community place an increasing semantic distance between themselves, the 'Lega' and Matteo Salvini, calling him a 'traitor' (as shown by the hashtag #*salvinitraditore*). The MINGOs community, instead, appears to be semantically focused on the issue of migration and distant from internal political matters. Due to the high presence of NGOs specialized in human rights and migration issues, this community takes a strong pro-migration stance, through hashtags such as #*ioaccolgo* (*I host*) and #*inclusione* (*inclusiveness*). Similarly, the presence of Catholic organizations orients the terms of discussion around Pope Francis I who is semantically identified as a positive figure opposed to the negative role of Matteo Salvini.

The results of the centrality analysis add another relevant piece of information to the semantic polarization shown in Fig. 5.8. In fact, most central hashtags are shared by all discursive communities. This evidence highlights the importance of combining the results of the rank-

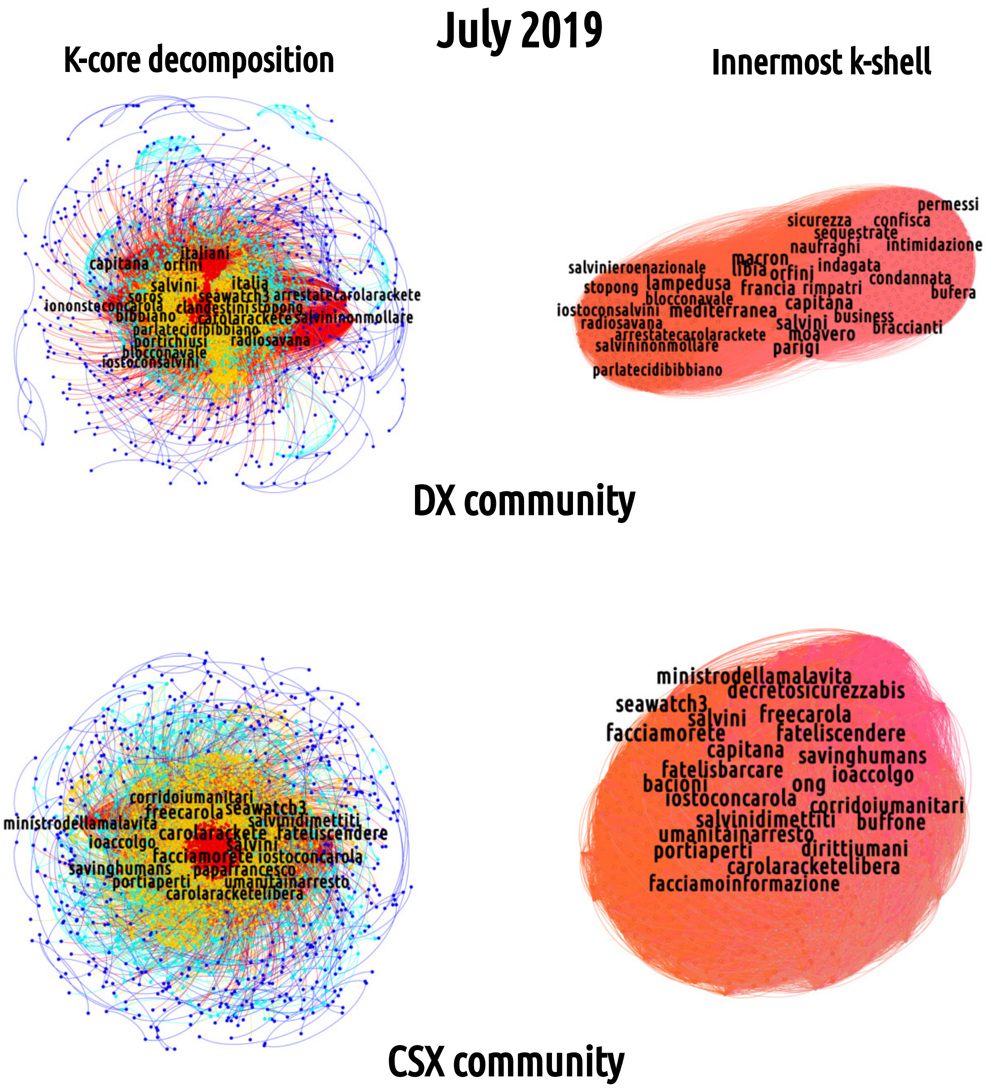


Figure 6.11: k-core decomposition of the July 2019 semantic networks for the DX (top left) and the CSX (bottom left) discursive communities. The k-core decomposition reveals the bulk of the discussion about immigration developed by these two discursive communities: while the innermost core of the DX induced semantic network (top right figure) is composed of hashtags such as *#salvininonmollare*, *#arrestatecarolarackete*, *#iostoconsalvini*, that of the CSX induced semantic network (bottom right figure) is composed of hashtags such as *#salvinidimettiti*, *#fateliscendere*, *#carolaracketelibera*.

ings of hashtags frequency with those of other network analyses, such as the study of the centrality measures. Similarly to the case of the 2018 daily semantic networks, the k-core decomposition points out the overall tendency of discussions to be hierarchically ordered. In Figs. 6.11 and 6.12 the distribution of k-core values has been divided into four quantiles that form five different regions colored from red to dark blue to indicate decreasing values of k. Although the k-core decomposition provides insights about the mesoscale organization of se-

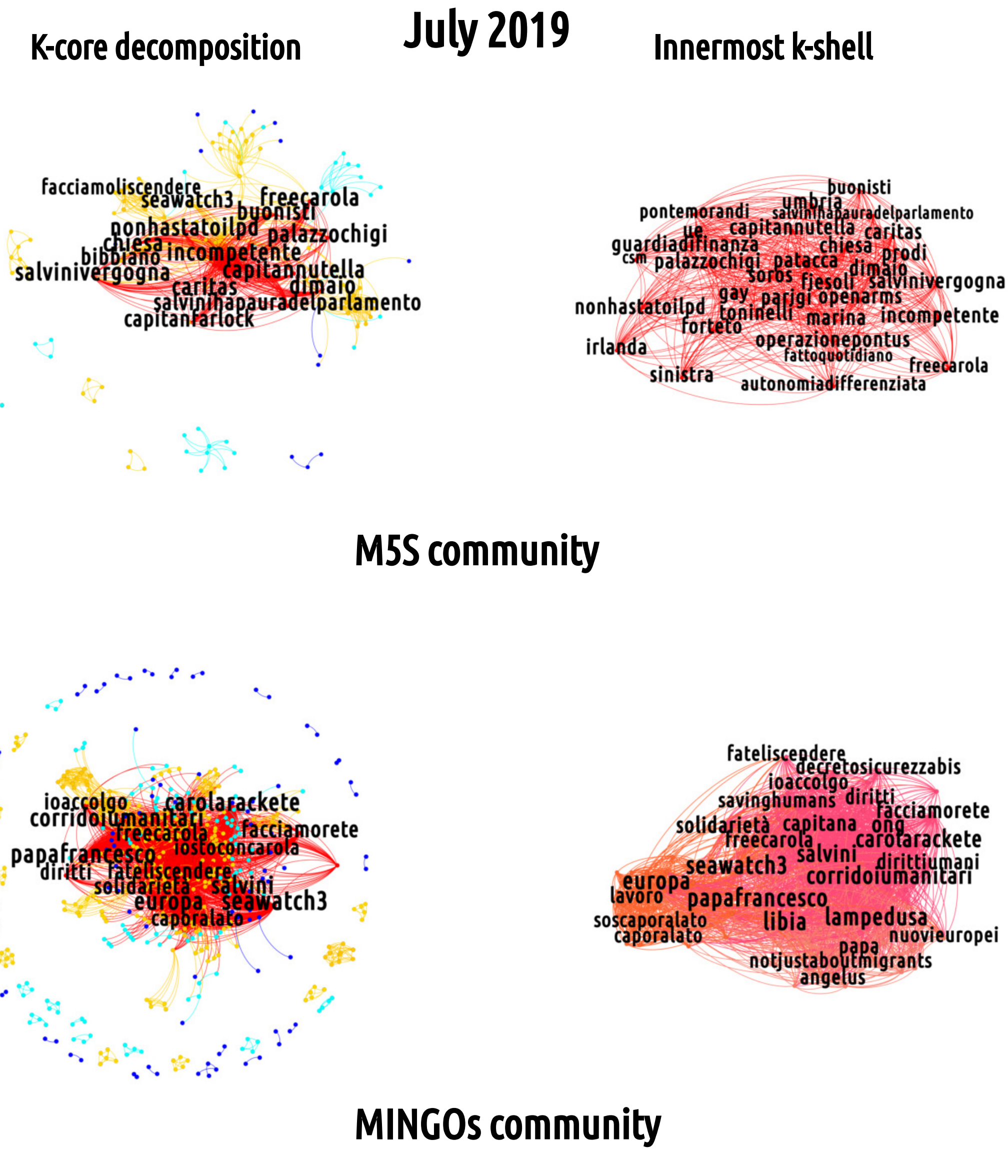


Figure 6.12: k-core decomposition of the July 2019 semantic networks for the M5S (top left) and the MINGOs (bottom left) discursive communities. The k-core decomposition reveals the bulk of the discussion about immigration developed by these two discursive communities: while the M5S community displays a mixed attitude towards immigration policies, as shown by the hashtags *#freecarola* and *#bibbiano*, the MINGOs innermost k-shell reveals strong support of pro-migration positions, as proven by the hashtags *#ioaccolgo*, *#iostoconcarola* and *#facciamorete*.

mantic networks, it cannot 'add' any significance to the partition individuated. Hence, as in Section 6.1, this analysis has been complemented by running a core-periphery decomposition via the surprise minimization. What emerges from the comparison between the k-core and the core-periphery decomposition is that the core overlaps to a large extent with the innermost k-shell, as the Jaccard correlation index is larger than 0.6 for all the semantic networks. In turn, such overlap signals that hashtags are hierarchically arranged within the Twitter discussions and, thus, that discursive communities tend to generate specific narratives that are hierarchically ordered around a finite set of thematic priorities. Analogous to what has been shown in relation to daily discursive dynamics developed during electoral campaigns, also in this case communities of hashtags revolve around a handful of political slogans and are located in the innermost k-shell of semantic networks. Conversely, secondary hashtags related to topics, actors or references are positioned in the peripheral area. A rather illustrative example is provided by the monthly induced semantic networks of July 2019 shown in Figs. 6.11 and 6.12.

As shown in Fig. 6.11, the DX and CSX induced semantic networks display a similar topological structure: both present a tightly connected bulk of hashtags (colored in red and orange) surrounded by a peripheral region (colored in light and dark blue) in which nodes are loosely interconnected. Their innermost k-shells are characterized by a set of common nodes such as *#carolarackete*, *#seawatch3* and *#salvini* suggesting that migration issues are specifically framed with respect to the episode of the Sea-Watch 3 and the controversial role of Matteo Salvini as the Minister of the Internal Affairs during this event. On the other hand, the analysis of the hashtags within the innermost core (on the top right and on the top left of Fig. 6.11) provides a hint of the polarized nature of framing practices inside these two discursive communities: within the DX community, some of the hashtags with the largest degree are *#portichiusi*, *#iostoconsalvini*, *#salvininonmollare*, *#arrestatecarolarackete* and *#nonfateliscendere*, all referring to the anti-migration slogans of the right-wing variety and supporting its leader Matteo Salvini. Conversely, the analysis of the core of the CSX induced semantic network reveals slogans like *#portiaperti*, *#carolaracketelibera* and *#fateliscendere*, *#salvinidimettiti*, *#ministrodellamalavita* which convey a radically opposing view, being openly against the closure of the Italian ports, sustaining strong pro-migration positions and calling for the Matteo Salvini's resignation from his position as Minister of Internal Affairs. As for the other discursive communities, Fig. 6.12 shows the M5S and MINGOs induced semantic networks. While these communities display a hierarchical structure that is similar to that of the other two, their peripheral region is organized in sparser structures that are also less connected with the rest of the network, suggesting the presence of multiple framing attempts, taking place at the same time, in a rather sparse way. The innermost k-shell of the M5S community reflects the same tension pointed out above with respect to the usage of hashtags that equally ask for the release of Carola Rackete, such as *#iostoconcarola* (*I stand with Carola*) and *#freecarola*, but also mark a distance from the 'Partito Democratico' along the lines of hashtags endorsed also by the DX community (e.g., *#bibbiano*). Interestingly, this monthly network shows the epitomes of the rift within the governmental coalition, which passes through framing the core issue of migration in conjunction with hashtags such as *#salvinivergogna* and *#salvinihapauradelparlamento* (respectively, *shame on Salvini* and *Salvini is afraid of the parliament*). Conversely, as noticed above, the Twitter discussion taking place in the MINGOs induced semantic networks is less centered on politics. Within this community, discussions on migration and international cooperation, as shown by the hashtags *#corridoiumanitari* and *#diritti* (respectively, *humanitarian corridors* and *rights*), are more prominent. Furthermore, the hashtags present within the community confirm the pro-migration position endorsed by its users, as proven by the presence in the innermost k-shell of the hashtags *#ioaccolgo*, *#iostoconcarola* and *#facciamorete*.

A similarly polarized configuration of framing practices that also tends to mirror fluid political alliances can be found in other monthly networks. For instance, observing the mesoscale structure of semantic networks in August 2019, it is possible to associate distinct partisanships with opposing framing practices occuring within different discursive communities. While there is a transversal tendency to build a connection between migration issues and the governmental crisis, discursive communities are populated by different slogans and keywords, supporting different ways of reading this connection. On the one side, the innermost core of the CSX and M5S induced semantic networks reveals hashtags against Matteo Salvini's decision to bring about the governmental crisis, such as *#governodelfallimento*, *#legatifrega*, *#salvinitraditore* and *#salvinidimettiti* (respectively, *government of failure*, *'Lega' fools you*, *Salvini liar* and *Salvini traitor*) thus shedding light on the progressive construction of a common ground for the future government alliance between the 'Partito Democratico' and the 'Movimento 5 Stelle'. On the opposing side, the DX community shows a strong endorsement for its leader and asks for new elections via hashtags such as *#iostoconsalvini*, *#salvininonmollare*, *#elezionisubito* and *#vogliamovotare* (respectively, *I stand with Salvini*, *Salvini don't give up*, *elections now* and *we want to vote*). Interestingly, in response to these positions, the 'former allies' from within the M5S community explicitly denounce Salvini's alleged betrayal via the hashtags *#legatifrega* and *#salvinitraditore*. Consistently, the core of the MINGOs community does not show any specific hashtags pointing to the governmental crisis. Thus, the innermost k-shell of its semantic network continues to be populated by a set of keywords rather similar to those of the July 2019 monthly semantic networks, such as *#corridoiumanitari* and *#papafrancesco*.

An interesting phenomenon that is observed when analysing monthly semantic networks is the co-existence of different types of mesoscopic structures. As pointed out in other works [132], communities are found to co-exist with k-shells in several systems: in particular, the innermost k-shell is frequently subdivided into a small number of communities further partitioning this shell. These sub-communities play an important role: within the discursive communities induced semantic networks, this role has been interpreted as a natural division of the most debated topics animating the same 'global' discussion. Running the Louvain community detection algorithm on the innermost shell (reported on the right in Figs. 6.11 and Fig. 6.12) indeed reveals the existence of these sub-communities. For instance, the core of the DX community displays two distinct sub-communities: one of the clusters concerns a general discussion on the regulation of the migration flows and the behavior of Italian institutions in handling migrants with hashtags like *#braccianti*, *#permessi* and *#confisca* (respectively, *farmworkers*, *permits* and *seizure*); the second one concerns the Sea-Watch 3 crisis and the opposition to its entrance into the Italian ports with hashtags like *#bloc-*

conavale, *#iostoconsalvini* and *#salvininonmollare*. Remarkably, nodes connecting the thematic clusters and acting as 'bridges' represent more general hashtags about politics and migration policies (e.g. *#salvini* and *#libia*). In this way, the view of the list of hashtags with largest betweenness centrality can be further refined by looking at the innermost k-shells. The DX induced semantic network for the month of July 2019 can be considered as a good example in this respect: the keywords playing the role of bridges read *#salvini*, *#capitana* (a hashtags referring to Carola Rackete as the *woman-captain* of the rescue boat Sea Watch 3) and *#libia*, a signal that these hashtags are included in several discussions within the innermost k-shell of this discursive community.

Chapter 7

Conclusions

Social media platforms have dramatically changed the patterns of news consumption and, over recent years, have become increasingly central during political events, such as electoral campaigns and societal discussions. Within a hybrid media system where networked partisanship and framing mechanisms arise spontaneously, Twitter has been shown to play a major political role exerting an agenda setting effect and contributing to polarization processes. Although the study of this platform has attracted scientists from many disciplines, researchers have so far mainly focused on users' activity, paying little attention to relational semantic mechanisms amongst users and their evolution, thus missing a particularly relevant aspect of online debates. In this context, ultimately, the behavioral rules driving online conversations remain a research area missing finer grained investigations that pay attention simultaneously to social and semantic aspects, as well as their interplay.

At the intersection of sociology, statistical physics and network science, the aim of this PhD work has been to propose a new methodological framework which addresses the semantic aspect of online political discussions in conjunction with the social dynamics. To this end, this work provides a detailed scheme of analysis that couples attention for several aspects of the networked framing occuring within discursive communities. Topic visibility, persistence and strategic semantic uses are studied in conjunction with the systematic analysis of more invisible, yet crucial, aspects concerning the production and circulation of media contents, as in the case of the identification of conversational cores of the networked framing. In doing so, the proposed approach provides a solid starting point also for understanding the symbolic aspects that nurture current dynamics of online civic participation, political partisanship and polarization which have so far catalyzed attention within the research community. All these aspects have been studied by paying particular attention to the analysis of relations between individuals, organizations and media. All these agents intervene within social dynamics gathering in online communities held together by a shared symbolic system which ultimately makes them partisan communities. In this sense, the social and semantic aspects of communication within Twitter are jointly studied in terms of the effect of a collective, recurrent and coherent behavior of online users interactions.

Compared to extant research, the current methodological approach offers two main advantages. First, the operational approach for isolating discursive communities and the corresponding community induced semantic networks does not entail any manual intervention. The semantic networks analyzed in this work follow from an identification of discursive communities that relies on an entropy-based approach, which is a methodological advancement *per se* as compared to other state-of-the-art techniques employed in socio-semantic studies. In fact, the present data-driven approach is grounded on an operational definition of online collective identities based on the Shannon entropy maximization under certain constraints which guarantees that this procedure is unbiased. This statistical framework allowed for the definition of a wide range of benchmarks through which filtering the retweeting activity of users while singling out the statistically relevant information at the desired level of detail. Above and beyond my particular case studies, the same approach can be applied to disentangle the intricacies of large-scale Twitter conversations in all domains and regardless of the language in which they take place with the only requirement of extracting data sets with a meaningful list of anchor hashtags.

Second, the semantic structures and the discursive communities are indicative of users' political affiliation yet without passing through any manual labelling of media contents. The extraction of political information from Twitter discussions occurs without any prior information on the political orientation of the users or the media contents they shared. Initial distinctions between verified and non-verified accounts are not defined bottom-up by researchers but are, in fact, assigned by the Twitter platform itself. Thus, as the classification of accounts in these two categories is regularly provided through the Streaming API by the platform itself, the solidity of the initial partition is ensured platform-wise. Despite the temporary suspension of the verification process in 2017, the robustness of discursive communities is assured not only by the ongoing grants verification on an *ad hoc* basis but also by the inclusion of non-verified users in the final communities. Similarly, Twitter discussions unfolding within each community are deduced by connecting any two hashtags if used a significantly large number of times by users, hence overcoming the limitations of studying online political conversations without considering the relational system between users. For all these reasons, the proposed semantic analysis approach allowed for several finer grained insights about the Italian Twittersphere under specific settings both at the macro level, grasping the semantic peculiarities of broader conversations taking place within discursive communities, and at the micro level, narrowing down the exploration to single and meaningful points in time during the electoral campaign period.

This PhD work offers a solid analysis of the public discourse developed online during two particularly crucial political discussions: the former occurs during run up to the 2018 Italian Elections and the two weeks after the Election day; the latter concerns more broadly the migration issue during the period May-November 2019. In both cases, the present methodological approach uses $\simeq 10^6$ tweets to define networks of statistically significant interactional links between users and hashtags at a daily and monthly time-scale. The current work results in the inference of peculiar social and semantic traits of the different political identities that guide political strategies and courses of action in the Italian political scenario. As regards the social analysis, the main finding of the present work is the operational definition of discursive communities as groups of users who share significantly similar retweeting patterns. Looking at networked partisanship from a broader view of elites on Twitter allowed for a better explanation of processes of both inter-group distancing and intra-groups consolidation. At the same time, the proposed approach of tracing communities partly detaches mechanisms of partisanship from those of formal political alliances. This is particularly evident in the separation in the 2019 migration discussion between the community of the ‘Movimento 5 Stelle’ and that of the ‘Lega’, as well as during periods in which the two parties jointly shared seats in the government: only under certain circumstances do the two communities merge in a temporary discursive coalition. However, networked partisanship cannot be thought of in isolation from political dynamics ‘on the ground’, as is well demonstrated within the left-wing community after the internal fracture of the ‘Partito Democratico’. All these features make this detection algorithm a useful tool for displaying a coherent picture of users’ behavior and political orientation: in the 2018 electoral campaign discussion, for instance, these communities are consistent with the political coalitions running for the 2018 Italian general election.

Another interesting insight that emerged from the social analysis is that social media affordances contribute differently to polarization processes. While discursive communities coalesce via retweets, they also interact, often in adversarial ways, via direct mentions. Importantly, different communities leverage on technological affordances in different ways. Collective partisan identities sustaining communities are thus formed in more institutional ways (mainly retweeting messages from parties and their leaders) or, more in line with a substantive criterion, by broadcasting messages from accounts that are more meaningfully active on electoral or migration issues. Similarly, mentions are employed to construct cross-community ties that, on the one hand, soften the isolation induced by partisan endorsement but, on the other, are often of adversarial nature. More relevantly, geometries of homophilic and cross-community ties tend to vary over time and in tight connection with relevant events on the ground - whether these are related to a specific discussion issue per se or are induced by shifting political alliances. All these communities’ characteristics are a clear indication of a coherent picture of Twitter users within them.

In regard to the semantic analysis, closer exploration of semantic networks sheds light on the more symbolic dimension of polarized partisan dynamics. One of the main findings of the analysis of the daily discursive dynamics developed during 2018 electoral campaign concerns the way the topological structure of semantic networks responds to the so-called mediated events (i.e. TV debates or the media coverage of offline events) thus revealing not only different sensibility towards media sphere but, more significantly, different identity traits that characterize each discursive political community. The topology of the CDX community is strongly dependent on these events (e.g., the mean degree of nodes increases in correspondence of specific TV shows), meaning that this group of users is more involved in the activity of retweeting during or after the immediate aftermath of the appearance of political actors particularly on television. Conversely, the activity of the M5S community appears to be much more distributed. In fact, although M5S supporters are sensitive to TV shows as well, their retweeting activity is not exclusively driven by media events but, rather, follows their preference for a generalized use of social media for organizational and political communication activities. Finally, the activity of the CSX community is characterized by a somewhat intermediate behavior: even when mediated events affect the Twitter discussion, the attention of the whole community is somehow shared among the various actors constituting the center-left alliance.

Particularly insightful is the analysis of the semantic networks at the mesoscale: what emerges is the presence of a core of topics, namely a densely connected bulk of hashtags surrounded by a periphery of loosely interconnected sub-topics. This indicates that daily semantic networks are characterized by a few relevant hashtags to which less relevant topics attach. This structure is maintained during the whole observation period and differences emerge only with respect to the number of peripheral themes entering the discussion. Moreover, the resilience of the core-periphery structure is not the same for the various discursive communities. In the context of semantic networks, the fact that the system is more or less resilient to the filtering implies that the various political groups have developed their political narrative differently, focusing their communications on a few related terms per topic or mentioning a set of omnipresent hashtags in all messages. Also in the response to the filtering procedure, M5S and CDX represent the two extremes, displaying respectively the most and the least resilient semantic network.

A core-periphery semantic structure appears also at a monthly time-scale in the discussion on the issue of migration. This analysis more markedly shows how multiple lines of polarization overlap with different discursive communities presenting distinctive approaches to the issue of migration. A few communities present a more substantive approach with a strong focus on the complex problem of migration, as in the case of the MINGOs community which genuinely endorse a pro-migrant point of view. Conversely, a more instrumental approach is evident in the communities shaped around political parties which in fact leverage on migration to contrast political adversaries. Amongst partisan communities that discuss migration in an instrumental way, right-wing and left-wing parties remain on opposite sides. This persisting arrangement is reflected in the networked framing practices in two ways. The first way is the formation of a semantic bridge between the theme of migration and the reinforcement of internal cohesion around the figure of Matteo Salvini, as in the case of the DX partisan community. Instead, at the level of cross-community ties between the DX and the CSX groups, the second way is tangible in the sustained contraposition between two opposite frames on migration: the far-rigt wing perspective oriented towards closure and the center-left wing towards openness. Conversely, the framing induced by the M5S community appears to be 'semantically torn' shifting and remaining ambivalent over time. In this tension, its positioning on migration issues remains vague and, in some sense, ancillary in comparison to a much more prominent interest for discussing political dynamics.

As already pointed out, the present approach remains limited in a twofold way. Regarding the social analysis, the discursive communities are lacking an exact indication about the actual users' political affiliation. This limitation can be relevant in the prediction of offline political outcomes such as voter turnout or election results. As for the semantic analysis, this work investigates only semantic structures formed in the digital space created by a single platform and pivoting around the hashtagging practices. Looking at Twitter discussions that form around specific anchor hashtags fails to include contributions of other social media platforms or simply of messages not including that set of keywords. Nonetheless, without any claim of exhaustivity, the present mapping of the Twitter discussions in occasion of the 2018 electoral campaign and 2019 migration issue discussions provides a useful entry point for understanding the online construction of political collective identities. Although limited to two Italian Twittersphere conversations taking place in close periods, this PhD work offers a solid methodological approach which can easily be fitted to further studies. A plethora of studies based on freely available Twitter data has shown that it is indeed possible to analyze electoral debates and societal discussions and shedding light on the political implications of non-traditional political acts such as public expressions on social media. Users employing in their tweets election and migration related hashtags did in fact contribute to framing the two political discussions along certain lines and they did so upon a platform that was not only widely used by a large segment of Italian population in that specific moment but that also plays a pivotal political communication role. Moreover, the strict filtering procedure for validating monopartite projections leads to the identification of networks with only statistically relevant users and hashtag information, as the projections guarantees that the network analysis is sound from both a methodological and interpretative point of view. Thus, albeit non representative of and non generalizable to the overall Italian population, both discursive communities and semantic networks can be thought of as a solid starting point for developing finer grained studies of voters' political opinion and behaviors.

Bibliography

[1] Statista. Global digital population as of january 2021 (in billions). https://www.statista.com/statistics/617136/digital-population-worldwide/, 2021. Accessed: 10 April 2021.

[2] Eurobarometer. Standard eurobarometer 92“Media Use in the European Union” report. *European Commission - Public Opinion*, 2019. (accessed January 7, 2021).

[3] C. Anderson. The end of theory: The data deluge makes the scientific method obsolete. https://www.uvm.edu/pdodds/files/papers/others/2008/anderson2008a.pdf, 2008. Accessed: 10 April 2021.

[4] T. Radicioni, T. Squartini, E. Pavan, and F. Saracco. Analysing twitter semantic networks: the case of 2018 italian elections. *Scientific Reports*, 11(1):1–22, 2021.

[5] C. Becatti, G. Caldarelli, R. Lambiotte, and F. Saracco. Extracting significant signal of news consumption from social networks: the case of Twitter in Italian political elections. *Palgrave Communications*, 5(91), Aug 2017.

[6] T. Radicioni, T. Squartini, E. Pavan, and F. Saracco. Networked partisanship and framing: A socio-semantic network analysis of the italian debate on migration. *PLOS ONE*, 16(8):1–24, 08 2021.

[7] G. Caldarelli, R. De Nicola, F. Del Vigna, M. Petrocchi, and F. Saracco. The role of bot squads in the political propaganda on Twitter. *Commun. Phys.*, 3(1):1–15, dec 2020.

[8] I. Paul, A. Khattar, P. Kumaraguru, M. Gupta, and S. Chopra. Elites tweet? characterizing the twitter verified user network. In *2019 IEEE 35th International Conference on Data Engineering Workshops (ICDEW)*, pages 278–285, 2019.

[9] A. Hearn. Verified: Self-presentation, identity management, and selfhood in the age of big data. *Popular Communication*, 15(2):62–77, 2017.

[10] G. Cimini, T. Squartini, F. Saracco, D. Garlaschelli, A. Gabrielli, and G. Caldarelli. The statistical physics of real-world networks. *Nature Reviews Physics*, 1(1):58–71, January 2019.

[11] H. K. Evans, S. Smith, A. Gonzales, and K. Strouse. Mudslinging on twitter during the 2014 election. *Social Media + Society*, 3(2):2056305117704408, 2017.

[12] M. D. Conover, B. Goncalves, J. Ratkiewicz, A. Flammini, and F. 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, Oct 2011.

[13] Y. Xiong, M. Cho, and B. Boatwright. Hashtag activism and message frames among social movement organizations: Semantic network analysis and thematic analysis of twitter during the #metoo movement. *Public Relations Review*, 45(1):10 – 23, 2019.

[14] P. Dahlgren. *Media and political engagement: citizens, communication, and democracy.* Cambridge University Press, United Kingdom, 2009.

[15] A. Chadwick. *The hybrid media system: Politics and power.* Oxford University Press, 2017.

[16] A. Bruns and J.-H. Schmidt. Produsage: a closer look at continuing developments. *New Review of Hypermedia and Multimedia*, 17(1):3–7, 2011.

[17] M. Castells. *Communication power.* OUP Oxford, 2013.

[18] A. Hepp. *Deep Mediatization: Key Ideas in Media & Cultural Studies.* Routledge, 2019.

[19] H. Jenkins. *Convergence Culture: Where Old and New Media Collide.* NYU Press, 2006.

[20] C. Wells, D. V. Shah, Jon C. Pevehouse, J. H. Yang, A. Pelled, F. Boehm, J. Lukito, S. Ghosh, and J. L. Schmidt. How trump drove coverage to the nomination: Hybrid media campaigning. *Political Communication*, 33(4):669–676, 2016.

[21] E. Pariser. *The Filter Bubble: What the Internet Is Hiding from You.* Penguin Group , The, 2011.

[22] M. Cinelli, E. Brugnoli, A. L. Schmidt, F. Zollo, W. Quattrociocchi, and A. Scala. Selective exposure shapes the facebook news diet. *PLOS ONE*, 15(3):1–17, 03 2020.

[23] M. Cinelli, G. De Francisci Morales, A. Galeazzi, W. Quattrociocchi, and M. Starnini. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9), 2021.

[24] A. L. Schmidt, F. Zollo, M. Del Vicario, A. Bessi, A. Scala, G. Caldarelli, H. E. Stanley, and W. Quattrociocchi. Anatomy of news consumption on facebook. *Proceedings of the National Academy of Sciences*, 114(12):3035–3039, 2017.

[25] E. Dubois and G. Blank. The echo chamber is overstated: the moderating effect of political interest and diverse media. *Information, Communication & Society*, 21(5):729–745, 2018.

[26] J. Bright. Explaining the emergence of echo chambers on social media: The role of ideology and extremism. *SSRN Electronic Journal*, 09 2016.

[27] C. R. Sunstein. The law of group polarization. *Journal of Political Philosophy*, 10(2):175–195, 2002.

[28] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi. Fame for sale: Efficient detection of fake Twitter followers. *Decis. Support Syst.*, 2015.

[29] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini. The rise of social bots. *Commun. ACM*, 59(7):96–104, June 2016.

[30] T. R. Keller and U. Klinger. Social bots in election campaigns: Theoretical, empirical, and methodological implications. *Political Communication*, 36(1):171–189, 2019.

[31] B. Latour. *Reassembling the Social: An Introduction to the Actor-Network Theory*. Oxford University Press, 2005.

[32] T. Gillespie, P. J. Boczkowski, and K. A. Foot. *Materiality and Media in Communication and Technology Studies: An Unfinished Project*, pages 21–51. The MIT Press, 2013.

[33] C. Shao, G. Ciampaglia, O. Varol, A. Flammini, F. Menczer, and K.-C. Yang. The spread of low-credibility content by social bots. *Nature Communications*, 9, 11 2018.

[34] T. Bucher. Want to be on the top? algorithmic power and the threat of invisibility on facebook. *New Media & Society*, 14(7):1164–1180, 2012.

[35] L. A. Granka. The politics of search: A decade retrospective. *The Information Society*, 26(5):364–374, 2010.

[36] D. Nikolov, M. Lalmas, A. Flammini, and F. Menczer. Quantifying biases in online information exposure. *CoRR*, abs/1807.06958, 2018.

[37] H. Gil de Zúñiga, N. Jung, and S. Valenzuela. Social Media Use for News and Individuals' Social Capital, Civic Engagement and Political Participation. *Journal of Computer-Mediated Communication*, 17(3):319–336, 04 2012.

[38] C.R. Sunstein. *Echo Chambers: Bush V. Gore, Impeachment, and Beyond*. Princeton University Press, 2001.

[39] M. Del Vicario, F. Zollo, G. Caldarelli, A. Scala, and W. Quattrociocchi. Mapping social dynamics on facebook: The brexit debate. *Social Networks*, 50:6 – 16, 2017.

[40] J. A. Tucker, A. Guess, P. Barbera, C. Vaccari, A. Siegel, S. Sanovich, D. Stukal, and B. Nyhan. Social media, political polarization, and political disinformation: A review of the scientific literature, Mar 2018.

[41] F. Zollo, A. Bessi, M. Del Vicario, A. Scala, G. Caldarelli, L. Shekhtman, S. Havlin, and W. Quattrociocchi. Debunking in a world of tribes. *PLOS ONE*, 12(7):1–27, 07 2017.

[42] M. Yarchi, C. Baden, and N. Kligler-Vilenchik. Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. *Political Communication*, 0(0):1–42, 2020.

[43] Magdalena Wojcieszak. 'don't talk to me': effects of ideologically homogeneous online groups and politically dissimilar offline ties on extremism. *New Media & Society*, 12(4):637–655, 2010.

[44] S. Iyengar, G. Sood, and Y. Lelkes. Affect, Not Ideology: A Social Identity Perspective on Polarization. *Public Opinion Quarterly*, 76(3):405–431, 09 2012.

[45] J. N. Druckman, E. Peterson, and R. Slothuus. How elite partisan polarization affects public opinion formation. *American Political Science Review*, 107(1):57–79, 2013.

[46] M. Wojcieszak. *Polarization, Political*, pages 1–7. American Cancer Society, 2016.

[47] Matthew S. Levendusky. Why do partisan media polarize viewers? *American Journal of Political Science*, 57(3):611–623, 2013.

[48] K. Arceneaux and M. Johnson. *More a Symptom Than a Cause: Polarization and Partisan News Media in America*, chapter Part IV - Polarization in the media, page 309–336. Cambridge University Press, 2015.

[49] P. Di Maggio, M. Nag, and D. Blei. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of u.s. government arts funding. *Poetics*, 41(6):570–606, 2013. Topic Models and the Cultural Sciences.

[50] P. Mancini. The italian public sphere: a case of dramatized polarization. *Journal of Modern Italian Studies*, 18(3):335–347, 2013.

[51] W. Lance Bennett and A. Segerberg. The logic of connective action. *Information, Communication & Society*, 15(5):739–768, 2012.

[52] E. Pavan and A. Mainardi. Striking, marching, tweeting. studying how online networks change together with movements. *Partecipazione e Conflitto*, 11(2), 2018.

[53] P. Barberá, J. Jost, J. Nagler, J. Tucker, and R. Bonneau. Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological science*, 26, 08 2015.

[54] H. T.P. Williams, J. R. McMurray, T. Kurz, and F. Hugo Lambert. Network analysis reveals open forums and echo chambers in social media discussions of climate change. *Global Environmental Change*, 32:126–138, 2015.

[55] S. J. Jackson, M. Bailey, and B. F. Welles. *#HashtagActivism: Networks of race and gender justice*. MIT Press, 2020.

[56] Z. Papacharissi. Affective publics and structures of storytelling: sentiment, events and mediality. *Information, Communication & Society*, 19(3):307–324, 2016.

[57] R. Pöyhtäri, M. Nelimarkka, K. Nikunen, M. Ojala, M. Pantti, and J. Pääkkönen. Refugee debate and networked framing in the hybrid media environment. *International Communication Gazette*, 0(0):1748048519883520, 0.

[58] S. Meraz and Z. Papacharissi. Networked gatekeeping and networked framing on #egypt. *The International Journal of Press/Politics*, 18(2):138–166, 2013.

[59] E. Siapera, M. Boudourides, S. Lenis, and J. Suiter. Refugees and network publics on twitter: Networked framing, affect, and capture. *Social Media + Society*, 4(1):2056305118764437, 2018.

[60] E. Pavan. The integrative power of online collective action networks beyond protest. exploring social media use in the process of institutionalization. *Social Movement Studies*, 16(4):433–446, 2017.

[61] M. Conover, J. Ratkiewicz, M. Francisco, B. Gonçalves, F. Menczer, and A. Flammini. Political polarization on twitter. In *Fifth International AAAI Conference on Weblogs and Social Media.*, 01 2011.

[62] A. J. Morales, J. Borondo, J. C. Losada, and R. M. Benito. Measuring political polarization: Twitter shows the two sides of venezuela. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(3):033114, 2015.

[63] D. A. Siegel. Social networks and collective action. *American Journal of Political Science*, 53(1):122–138, 2009.

[64] L. Fleming, S. Mingo, and D. Chen. Collaborative brokerage, generative creativity, and creative success. *Administrative Science Quarterly*, 52(3):443–475, 2007.

[65] S. D. Mcclurg. Social networks and political participation: The role of social interaction in explaining political participation. *Political Research Quarterly*, 56(4):449–464, 2003.

[66] A. Melucci. The process of collective identity. *Social movements and culture*, 4:41–63, 1995.

[67] X. E. Barandiaran, A. Calleja-López, and E. Cozzo. Defining collective identities in technopolitical interaction networks. *Frontiers in Psychology*, 11:1549, 2020.

[68] S. Pei, L. Muchnik, J. S. Andrade, Jr., Z. Zheng, and H. A. Makse. Searching for superspreaders of information in real-world social media. *Scientific Reports*, 4:5547, July 2014.

[69] S. González-Bailón, J. Borge-Holthoefer, and Y. Moreno. Broadcasters and hidden influentials in online protest diffusion. *American Behavioral Scientist*, 57(7):943–965, 2013.

[70] M. T. Bastos, D. Mercea, and A. Charpentier. Tents, tweets, and events: The interplay between ongoing protests and social media. *Journal of Communication*, 65(2):320–350, 2015.

[71] C. Padovani and E. Pavan. Global governance and icts: exploring online governance networks around gender and media. *Global Networks*, 16(3):350–371, 2016.

[72] N. Gaumont, M. Panahi, and D. Chavalarias. Reconstruction of the socio-semantic dynamics of political activist twitter networks—method and application to the 2017 french presidential election. *PLOS ONE*, 13(9):1–38, 09 2018.

[73] F. Giglietto, L. Iannelli, L. Rossi, A. Valeriani, N. Righetti, F. Carabini, G. Marino, S. Usai, and E. Zurovac. Mapping italian news media political coverage in the lead-up of 2018 general election. *SSRN Electronic Journal*, 2018.

[74] F. Giglietto and Y. Lee. To be or not to be charlie: Twitter hashtags as a discourse and counter-discourse in the aftermath of the 2015 charlie hebdo shooting in france. In *Proceedings of the 5th Workshop on Making Sense of Microposts co-located with the 24th International World Wide Web Conference*, pages 33–37, 2015.

[75] R. Marchetti and D. Ceccobelli. Twitter and television in a hybrid media system. *Journalism Practice*, 10(5):626–644, 2016.

[76] R. Recuero, G. Zago, M. T. Bastos, and R. Araújo. Hashtags functions in the protests across brazil. *SAGE Open*, 5(2):2158244015586000, 2015.

[77] A. Chiaramonte, V. Emanuele, N. Maggini, and A. Paparo. Populist success in a hung parliament: The 2018 general election in italy. *South European Society and Politics*, 23(4):479–501, 2018.

[78] R. Bracciale and C. Cepernich. Hybrid 2018 campaigning: the social media habits of italian political leaders and parties. *Italian Political Science*, 13(1):36–50, 2018.

[79] A. Jungherr. Twitter use in election campaigns: A systematic literature review. *Journal of Information Technology & Politics*, 13:72–91, 03 2016.

[80] C. Vaccari and A. Valeriani. Follow the leader! direct and indirect flows of political communication during the 2013 italian general election campaign. *New Media & Society*, 17(7):1025–1042, 2015.

[81] M. Berry, I. Garcia-Blanco, and K. Moore. Press coverage of the refugee and migrant crisis in the eu: A content analysis of five european countries. Technical report, United Nations High Commissioner for Refugees, 2016.

[82] G. Mazzoleni and R. Bracciale. Socially mediated populism: the communicative strategies of political leaders on Facebook. *Palgrave Communications*, 4(1):1–10, December 2018.

[83] G. Iocco, M. Lo Cascio, and D. C. Perrotta. 'close the ports to african migrants and asian rice!': The politics of agriculture and migration and the rise of a 'new' right-wing populism in italy. *Sociologia Ruralis*, 60(4):732–753, 2020.

[84] M. Huberty. Can we vote with our tweet? on the perennial difficulty of election forecasting with social media. *International Journal of Forecasting*, 31(3):992 – 1007, 2015.

[85] J. Pfeffer, K. Mayer, and F. Morstatter. Tampering with twitter's sample api. *EPJ Data Science*, 7(1):50, 2018.

[86] V. Cosenza. Social media in italia: utenti e tempo di utilizzo nel 2018. `https://vincos.it/2019/04/10/social-media-in-italia-utenti-e-tempo-di-utilizzo-nel-2018/`, 2019. Accessed: 20 December 2020.

[87] M. Andretta and E. Pavan. *Mapping Protest on the Refugee Crisis: Insights from Online Protest Event Analysis*, pages 299–324. 02 2018.

[88] Toffee - datasets. `https://toffee.imtlucca.it/datasets`, 2021. Accessed: 13 September 2021.

[89] M. Newman. *Networks: An Introduction*. Oxford University Press, Inc., USA, 2010.

[90] G. Caldarelli and M. Catanzaro. *Networks: A Very Short Introduction*. Very Short Introductions. OUP Oxford, 2012.

[91] P. Landi, H. O. Minoarivelo, Å. Brännström, C. Hui, and U. Dieckmann. Complexity and stability of ecological networks: a review of the theory. *Population Ecology*, 60(4):319–345, 2018.

[92] M. Bardoscia, Paolo Barucca, S. Battiston, F. Caccioli, G. Cimini, D. Garlaschelli, F. Saracco, Tiziano Squartini, and G. Caldarelli. The physics of financial networks. *ArXiv*, abs/2103.05623, 2021.

[93] M. O. Jackson. *Social and Economic Networks*. Princeton University Press, 2008.

[94] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Rev. Mod. Phys.*, 87:925–979, Aug 2015.

[95] A. Bessi, F. Zollo, M. Del Vicario, M. Puliga, A. Scala, G. Caldarelli, B. Uzzi, and W. Quattrociocchi. Users polarization on facebook and youtube. *PLOS ONE*, 11(8):1–24, 08 2016.

[96] F. Saracco, R. di Clemente, A. Gabrielli, and T. Squartini. Randomizing bipartite networks: the case of the World Trade Web. *Scientific Reports*, 5:10595, June 2015.

[97] M. J. Straka, G. Caldarelli, and F. Saracco. Grand canonical validation of the bipartite international trade network. *Phys. Rev. E*, 96:022306, Aug 2017.

[98] F. Saracco, M. J. Straka, R. Di Clemente, A. Gabrielli, G. Caldarelli, and T. Squartini. Inferring monopartite projections of bipartite networks: an entropy-based approach. *New Journal of Physics*, 19(5):053022, may 2017.

[99] M. J. Straka, G. Caldarelli, T. Squartini, and F. Saracco. From ecology to finance (and back?): A review on entropy-based null models for the analysis of bipartite networks. *Journal of Statistical Physics*, 173:1252–1285, 2018.

[100] Twitter Help Center. About verified accounts. `https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts`, Accessed February 20, 2021.

[101] A. Hutchinson. Twitter hits another hiccup with public requests for verification, pauses roll-out of new process. `https://www.socialmediatoday.com/news/twitter-hits-another-hiccup-with-public-requests-for-verification-pauses-r/605021/`, 2021. Accessed: 13 September 2021.

[102] C. Welch. Twitter says it doesn't 'have the bandwidth' to fix verification right now. `https://www.theverge.com/2018/7/17/17583624/twitter-cant-fix-verification-because-elections`, 2018. Accessed: 13 September 2021.

[103] M. R. Morris, S. Counts, A. Roseway, A. Hoff, and J. Schwarz. Tweeting is believing? understanding microblog credibility perceptions. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, CSCW '12, page 441–450, New York, NY, USA, 2012. Association for Computing Machinery.

[104] T. A. Small. What the hashtag? *Information, Communication & Society*, 14(6):872–895, 2011.

[105] W. H. Gomaa and A. A. Fahmy. A survey of text similarity approaches. *International Journal of Computer Applications*, 68(13):13–18, April 2013. Full text available.

[106] Z. Neal. The backbone of bipartite projections: Inferring relationships from co-authorship, co-sponsorship, co-attendance and other co-behaviors. *Social Networks*, 39:84 – 97, 2014.

[107] M. Ángeles Serrano, M. Boguñá, and A. Vespignani. Extracting the multiscale backbone of complex weighted networks. *Proceedings of the National Academy of Sciences*, 106(16):6483–6488, 2009.

[108] S. Gualdi, G. Cimini, K. Primicerio, R. Di Clemente, and D. Challet. Statistically validated network of portfolio overlaps and systemic risk. *Scientific reports*, 6:39467, 2016.

[109] M. Tumminello, S. Miccichè, F. Lillo, J. Piilo, and R. N. Mantegna. Statistically validated networks in bipartite complex systems. *PLOS ONE*, 6(3):1–11, 03 2011.

[110] Y. Hong. On computing the distribution function for the poisson binomial distribution. *Computational Statistics & Data Analysis*, 59(C):41–51, 2013.

[111] S. Fortunato and D. Hric. Community detection in networks: A user guide. *Physics Reports*, 659:1 – 44, 2016. Community detection in networks: A user guide.

[112] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, oct 2008.

[113] A. Lancichinetti and S. Fortunato. Community detection algorithms: A comparative analysis. *Phys. Rev. E*, 80:056117, Nov 2009.

[114] U. N. Raghavan, R. Albert, and S. Kumara. Near linear time algorithm to detect community structures in large-scale networks. *Physical review E*, 76(3):036106, 2007.

[115] E. Pavan A. Rapini. Antifascism retweeted. semantic networks around #antifa* in the italian twittersphere. In *Annual Conference of the Association of Italian Political Communication (AssoComPol)*, 2019.

[116] T. Highfield and A. Bruns. Compulsory voting, encouraged tweeting? australian elections and social media. In G Enli, A Bruns, A O Larsson, E Skogerbo, and C Christensen, editors, *The Routledge companion to social media and politics*, pages 338–350. Routledge, United States of America, 2016.

[117] A. Hanna. Computer-aided content analysis of digitally enabled movements. *Mobilization: An International Quarterly*, 18(4):367–388, 2013.

[118] N. Aruguete. The agenda setting hypothesis in the new media environment. *Comunicacion Y Sociedad*, pages 35–58, 2017.

[119] G. Caldarelli, R. de Nicola, M. Petrocchi, M. Pratelli, and F. Saracco. Analysis of online misinformation during the peak of the COVID-19 pandemics in Italy. *arXiv e-prints*, oct 2020.

[120] A. Feller, M. Kuhnert, T. Sprenger, and I. Welpe. Divided they tweet: The network structure of political microbloggers and discussion topics. *Proceedings of the International AAAI Conference on Web and Social Media*, 5(1):474–477, Aug. 2011.

[121] C. Mudde. *Populist Radical Right Parties in Europe*. Cambridge University Press, 2007.

[122] E. Brugnoli, M. Cinelli, F. Zollo, W. Quattrociocchi, and A. Scala. Lexical convergence inside and across echo chambers. *CoRR*, abs/1903.11452, 2019.

[123] R. Page. *Collective Identities and Co-tellership in Facebook Comments*, page 101–119. Cambridge University Press, 2018.

[124] G. Caldarelli. *Scale-free networks: complex webs in nature and technology*. Oxford University Press, 2007.

[125] Massimo Stella. Cognitive network science for understanding online social cognitions: A brief review. *arXiv preprint arXiv:2102.12799*, 2021.

[126] Y. Lin, B. Keegan, D. Margolin, and D. Lazer. Rising tides or rising stars?: Dynamics of shared attention on twitter during media events. *PLOS ONE*, 9(5):1–12, 05 2014.

[127] G. Bianconi, R. K. Darst, J. Iacovacci, and S. Fortunato. Triadic closure as a basic generating mechanism of communities in complex networks. *Physical Review E*, 90(4):042806, October 2014.

[128] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, New York, NY, USA, 2010.

[129] Y. Kong, G. Shi, R. Wu, and Y. Zhang. k-core: Theories and applications. *Physics Reports*, 832:1 – 32, 2019. k-core: Theories and Applications.

[130] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse. Identification of influential spreaders in complex networks. *Nature Physics*, 6(11):888–893, Aug 2010.

[131] J. v. L. de Jeude, G. Caldarelli, and T. Squartini. Detecting core-periphery structures by surprise. *Europhysics Letters*, 125(6):68001, 2019.

[132] I. Malvestio, A. Cardillo, and N. Masuda. Interplay between k-core and community structure in complex networks. *Scientific reports*, 10(1):1–12, 2020.

[133] E. Marchese, G. Caldarelli, and T. Squartini. Detecting mesoscale structures by surprise. *arXiv preprint arXiv:2106.05055*, 2021.

[134] K. M. Carley and D. S. Kaufer. Semantic connectivity: An approach for analyzing symbols in semantic networks. *Communication Theory*, 3(3):183–213, 1993.

[135] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69:026113, Feb 2004.

[136] S. P. Borgatti and M. G. Everett. Models of core/periphery structures. *Social Networks*, 21(4):375–395, 2000.

Appendix A

Lists of verified users within discursive communities

In the following Subsections, the complete list of verified users within each identified discursive community is reported for both data sets. In particular, Table A.1 and Table A.2 refer respectively to the 2018 Italian general elections data set and to 2019 Italian migration discussion data set.

A.1 2018 Italian general elections

User screen name	User ID	Discursive community
1giornodapecora	67006079	CDX
Adnkronos	25676606	CDX
AffInt	106699882	CDX
Affaritaliani	35525950	CDX
Alice_Weidel	833398209053605888	CDX
Antonio_Tajani	529247064	CDX
Bertolaso2016	4905672815	CDX
CalabriaTw	412371227	CDX
Capezzone	144210805	CDX
ChiaraMaff	368947547	CDX
DeborahBergamin	92556620	CDX
EGardini	404836757	CDX
Europarl_IT	37621971	CDX
FioriMarcello	2350203541	CDX
FratellidItaIia	1024976264	CDX
GabriGiammanco	714417878	CDX
GiorgiaMeloni	130537001	CDX
Giorgiolaporta	16145985	CDX
GiovanniToti	2331718804	CDX
GruppoFICamera	1344131071	CDX
IAIonline	237334959	CDX
IsolaDeiFamosi	228841112	CDX
LegaSalvini	13514762	CDX
LiciaRonzulli	318316637	CDX
LombardiaOnLine	56674519	CDX
ManfredWeber	20399735	CDX
Marcozanni86	112873259	CDX
MariaLatella	74693664	CDX
MatrixCanale5	1969190317	CDX

MaxSalini	601773356	CDX
MediasetTgcom24	331617619	CDX
NicolaPorro	399004979	CDX
Noiconsalvini	2900789860	CDX
PE_Italia	897627558	CDX
PatricielloAldo	454821447	CDX
QuiMediaset_it	384760605	CDX
QuintaColonnaTv	3044664286	CDX
Quirinale	732819391	CDX
Radio1Rai	210501383	CDX
RadioRadicale	74707629	CDX
RaiNews	14060262	CDX
RaiPortaaPorta	406880559	CDX
RaiTre	1433286055	CDX
RaiUno	1208394690	CDX
Raiofficialnews	538147915	CDX
RudyBandiera	14684382	CDX
SkyTG24	5893702	CDX
Tg3web	69959408	CDX
TgrMarche	1647500934	CDX
ansaeuropa	296371084	CDX
antoniopalmieri	6827962	CDX
berlusconi	920277002858500096	CDX
borghi_claudio	337767301	CDX
classcnbc	372208149	CDX
coldiretti	626360641	CDX
comilara	156729839	CDX
discoradioIT	487403439	CDX
elio_vito	1557514573	CDX
europainitalia	90139268	CDX
forza_italia	147543162	CDX
fulviomartuscie	813738259	CDX
ispionline	91138156	CDX
limesonline	9095292	CDX
m_hunziker	378611487	CDX
matteosalvinimi	270839361	CDX
pedroelrey	9734552	CDX
pomeriggio5	3030837185	CDX
renatobrunetta	20430345	CDX
rtl1025	56341776	CDX
s_parisi	351675859	CDX
simonefurlan1	507652780	CDX
social_mediaset	849195252	CDX
sole24ore	420351046	CDX
sputnik_italia	2817304093	CDX
zaiapresidente	1324893229	CDX
AAtheMerciless	416922961	CSX
ALDEParty	42378378	CSX
Agenzia_Dire	469398862	CSX
Agenzia_Italia	72248630	CSX
AlessiaMorani	429312499	CSX
AndreaRomano9	284706444	CSX
Andst7	541882699	CSX
AnnaAscani	492325083	CSX
AnnalisaChirico	117407713	CSX
BeaLorenzin	515229378	CSX
BeppeSala	4524412653	CSX
CarloCalenda	2416067982	CSX
CarloStagnaro	332894847	CSX
ChangeItalia	636955791	CSX
ComuneMI	728604452	CSX

Corriere	395218906	CSX
Daniele_Manca	29236274	CSX
DantiNicola	829032654	CSX
DarioNardella	434505068	CSX
DavidSassoli	54988878	CSX
EnricoLetta	419622371	CSX
Ettore_Rosato	258861907	CSX
FGoria	6361512	CSX
FerdiGiugliano	503250806	CSX
HuffPostItalia	543774554	CSX
LaStampa	29416653	CSX
Linkiesta	215728045	CSX
LottiLuca	1949844973	CSX
LucaTentoni1	2282325169	CSX
Maumol	23364501	CSX
PaoloGentiloni	406869976	CSX
Pierferdinando	36374014	CSX
PieroPelu	1281472218	CSX
PietroDeMatteis	109288334	CSX
Radicali	150196030	CSX
RiccardoLuna	17214580	CSX
RobTallei	253636558	CSX
RossPellecchia	1095820980	CSX
Serv_Pubblico	370009808	CSX
Tommasocerno	66966529	CSX
VincenzoDeLuca	110960339	CSX
alemannoEU	18720978	CSX
alessiobalbi	23077590	CSX
alessioviola	107582221	CSX
alexbarbera	409361259	CSX
annamasera	28098604	CSX
bobogiac	523058143	CSX
boccadutri	490373717	CSX
bordomichele	62038434	CSX
carlogabardini	324594735	CSX
chedisagio	44487069	CSX
comunefi	420186908	CSX
dariofrance	61154684	CSX
davidallegranti	210573372	CSX
davidealgebris	484158845	CSX
disinformatico	7111232	CSX
emenietti	7697992	CSX
eurodeputatipd	79385667	CSX
eziomauro	365462238	CSX
ferrazza	8226422	CSX
francescocosta	6566922	CSX
gabrieligm	14210532	CSX
giorgio_gori	368950045	CSX
giornalettismo	53154047	CSX
giusmo1	388284615	CSX
graziano_delrio	484982241	CSX
gualtierieurope	103152299	CSX
ilfoglio_it	49954018	CSX
ilpost	133790890	CSX
insopportabile	39941540	CSX
ivanscalfarotto	30248306	CSX
lorepregliasco	454423746	CSX
lucasofri	20910977	CSX
lucfontana	358754219	CSX
mante	394613	CSX
marcocappato	128870883	CSX

marcodimaio	48484178	CSX
mariannamadia	588975097	CSX
masechi	27456576	CSX
matteorenzi	18762875	CSX
mattiafeltri	482897816	CSX
maumartina	415571726	CSX
meb	588200416	CSX
nadiaferrigo	226134529	CSX
nicola_pinna	364785591	CSX
nzingaretti	403544693	CSX
pdnetwork	13294452	CSX
pfmajorino	271596933	CSX
pierofassino	100218289	CSX
pietroraffa	61877999	CSX
repubblica	18935802	CSX
riccardomagi	283026832	CSX
ricpuglisi	225168723	CSX
riotta	73971750	CSX
riparteilfuturo	988659876	CSX
robertapinotti	102672182	CSX
robertosaviano	14516324	CSX
sebmessina1	383116512	CSX
serracchiani	35298549	CSX
simonabonafe	391945446	CSX
smenichini	40006630	CSX
straneuropa	334777310	CSX
valeriafedeli	480645077	CSX
vittoriozucconi	419918470	CSX
AsiaArgento	241027939	LEU
Assolombarda	323149893	LEU
Confindustria	732136568421556224	LEU
MassimGiannini	3241806322	LEU
PietroGrasso	1071332641	LEU
Tommasolabate	122014351	LEU
_arianna	16562213	LEU
civati	14108472	LEU
fabiochiusi	58768691	LEU
lauraboldrini	221902171	LEU
lucatelese	100610543	LEU
robertobernabo	132303430	LEU
rossipresidente	97734603	LEU
valigiablu	146448681	LEU
Agenzia_Ansa	150725695	M5S
AndCol81	1407978600	M5S
AndreaScanzi	376218450	M5S
AndreaVallascas	1386423176	M5S
AngeloTofalo	337171830	M5S
Azzurra_C	87934587	M5S
BarillariM5S	107517248	M5S
Bernini_P	19863725	M5S
DadoneFabiana	708276706	M5S
DalilaNesci	473107684	M5S
DanielaAiuto	2466169957	M5S
DanielaDonno	136840065	M5S
DaniloToninelli	960924277	M5S
DeBortoliF	335953039	M5S
DgPilot81	202726940	M5S
EleonoraEvi	1135141640	M5S
FMCastaldo	1368850908	M5S
F_DUva	238909679	M5S
FedericaDaga	361373768	M5S

Ferraresi_V	2256679742	M5S
FicarraePicone	398904362	M5S
FrancBusinarolo	1072331106	M5S
Gallinella_F	1637833237	M5S
GianlucaVasto	318499634	M5S
GiuliaDiVita	379256309	M5S
GiuliaSarti86	1053093684	M5S
Isa_Adinolfi	2195622679	M5S
La7tv	121424128	M5S
LauraBottici	1943781798	M5S
LuigiGallo15	920435114	M5S
M5S_Europa	2530314205	M5S
MFantinati	49275896	M5S
ManlioDS	208642171	M5S
MatteoPedrosi	212923650	M5S
Maxdero	195689744	M5S
MirkoBusto	974102796	M5S
Mlucialorefice	1623798294	M5S
Mov5Stelle	289400495	M5S
NicolaMorra63	1314728936	M5S
OttoemezzoTW	2359413432	M5S
PaoloParentela	381611850	M5S
PediciniM5S	631767111	M5S
PiazzapulitaLA7	58453980	M5S
Roma	459558301	M5S
SVignaroli	1107008174	M5S
SorialGiorgio	1915200546	M5S
StefaniaPetyx	425380217	M5S
StefanoFeltri	391485689	M5S
TgLa7	828717014	M5S
DAGOSPIA	70971449	M5S
a_padellaro	274087932	M5S
berenice0104	93917881	M5S
c_appendino	963073442	M5S
carlaruocco1	978488840	M5S
carlosibilia	127025568	M5S
ciropellegrino	5847962	M5S
dellorco85	1638278257	M5S
diMartedi	2695436575	M5S
e_cappelletti	2210614638	M5S
enpaonlus	461067562	M5S
espressonline	85852409	M5S
fanpage	213254229	M5S
fattoquotidiano	52424550	M5S
g_brescia	1326743389	M5S
gparagone	384927198	M5S
ignaziocorrao	143393223	M5S
loscibo	165443253	M5S
lucafrusone	262823808	M5S
luigidimaio	48062712	M5S
marcotravaglio	320603385	M5S
matteodallosso	13065292	M5S
micillom5s	1715119512	M5S
mirellaliuzzi	11945512	M5S
nicola_bianchi	133366702	M5S
petergomezblog	274093178	M5S
redazioneiene	239784688	M5S
riccardo_fra	1052596340	M5S
robertalombardi	61727028	M5S
silviachimienti	1888044751	M5S
stanzaselvaggia	416476798	M5S

tancredipalmeri	167704836	M5S
tatianabasilio1	969081366	M5S
valedenina	38182836	M5S
vilmamoronese	126970118	M5S
vitocrimi	62003557	M5S
vitopetrocelli	318400179	M5S
you_trend	404064077	M5S
Avvenire_NEI	132880191	MEDIA
BentivogliMarco	491327107	MEDIA
ReginaCatrambon	27448416	MEDIA
Rvaticanaitalia	121480284	MEDIA
TIM_music	342100068	MEDIA
TgrRaiPuglia	4850084878	MEDIA
antoniospadaro	6062112	MEDIA
chetempochefa	58748101	MEDIA
civcatt	246220040	MEDIA
francescoseghez	65444000	MEDIA

Table A.1: Verified users list within discursive communities in the 2018 data set. Each user screen name is reported together with its user ID which uniquely identifies a Twitter user.

A.2 2019 Italian migration discussion

User screen name	User ID	Discursive community
AffInt	106699882	CSX
AlessiaMorani	429312499	CSX
Am_Parente	701094295	CSX
AndreaMarcucci	57282569	CSX
AndreaOrlandosp	496437886	CSX
AndreaRomano9	284706444	CSX
AnnaAscani	492325083	CSX
BeaLorenzin	515229378	CSX
BentivogliMarco	491327107	CSX
BeppeSala	4524412653	CSX
BrunoAstorre	2900717633	CSX
CarloCalenda	2416067982	CSX
CarloStagnaro	332894847	CSX
Comincini	28398786	CSX
ComuneNapoli	83800849	CSX
CottarelliCPI	940499085886480384	CSX
DamianoZoffoli	141151403	CSX
DantiNicola	829032654	CSX
DarioStefano	295754318	CSX
DavidSassoli	54988878	CSX
Davide	625733	CSX
DeBortoliF	335953039	CSX
EP_President	818837224003289088	CSX
Emagorno	1114681735	CSX
EnricoLetta	419622371	CSX
Erriders	262421906	CSX
Ettore_Rosato	258861907	CSX
F_Boccia	461229005	CSX
FerdiGiugliano	503250806	CSX
FraMirabelli	545136488	CSX
FrancescoBonif1	418814614	CSX
GBenamati	168586183	CSX
GassmanGassmann	1380510872	CSX
GermaniaItalia	2711832192	CSX

GiovanniLegnini	334611957	CSX
IlContiAndrea	417469423	CSX
IsabellaDeMonte	1510780328	CSX
ItaliaViva	1059361525109010433	CSX
LaStampa	29416653	CSX
LauraGaravini	321939235	CSX
LiaQuartapelle	216361540	CSX
Linkiesta	215728045	CSX
LinoGuanciale	1546467320	CSX
LucaBizzarri	28822067	CSX
MarleneSchiappa	14237828	CSX
MassimGiannini	3241806322	CSX
MatteoRichetti	395183088	CSX
MonicaCirinna	334337968	CSX
NicolaCaputo	83041759	CSX
PaoloGentiloni	406869976	CSX
Radicali	150196030	CSX
RobertoBurioni	732817452569141248	CSX
RossomandoPd	2927013813	CSX
Serv_Pubblico	370009808	CSX
ShooterHatesYou	94324642	CSX
SilviaCostaEU	1336254342	CSX
SimonaMalpezzi	428521248	CSX
SirDistruggere	413426849	CSX
TNannicini	443198044	CSX
TeresaBellanova	606259626	CSX
TizianaFerrario	50689313	CSX
Twiperbole	43086900	CSX
UNHCRLibya	902045651726368768	CSX
USCGFlorence	895925394301493248	CSX
ValeriaValente_	1288696170	CSX
VincenzoDeLuca	110960339	CSX
alex_orlowski	55862897	CSX
alexbarbera	409361259	CSX
annamasera	28098604	CSX
antoniomisiani	336975968	CSX
beppesevergnini	33196091	CSX
bordomichele	62038434	CSX
brandobenifei	87306292	CSX
caritas_milano	130837695	CSX
carlogubi	36353564	CSX
caterinabiti	38183286	CSX
cescverducci	827908404	CSX
chedisagio	44487069	CSX
chiaragribaudo	1030715161	CSX
ckyenge	140133595	CSX
cozzolino62	467604110	CSX
danieleviotti	22651555	CSX
dariofrance	61154684	CSX
davidallegranti	210573372	CSX
davidealgebris	484158845	CSX
davidefaraone	41391760	CSX
demagistris	29416670	CSX
edopatriarca	1171546866	CSX
elenabonetti	1134357498	CSX
emenietti	7697992	CSX
emmedibi	58020070	CSX
eurodeputatipd	79385667	CSX
evagiovannini	215087363	CSX
eziomauro	365462238	CSX
fabfazio	1005827778	CSX

fabiochiusi	58768691	CSX
fabriggio	4156055777	CSX
ferrazza	8226422	CSX
francescocosta	6566922	CSX
francescoseghez	65444000	CSX
gabrieligm	14210532	CSX
gennaromigliore	154082631	CSX
giannigipi	253012857	CSX
giorgio_gori	368950045	CSX
giornalettismo	53154047	CSX
giosiferrandino	402097529	CSX
graziano_delrio	484982241	CSX
gualtierieurope	103152299	CSX
guerini_lorenzo	1870754606	CSX
heather_parisi	973582794	CSX
il_pucciarelli	370278925	CSX
ilfoglio_it	49954018	CSX
ivanscalfarotto	30248306	CSX
lauraboldrini	221902171	CSX
lorepregliasco	454423746	CSX
lucasofri	20910977	CSX
lucfontana	358754219	CSX
lucianodalfonso	238681240	CSX
lucianonobili	130851801	CSX
m_giuffrida	143536359	CSX
makkox	16273541	CSX
mancaimola	897264492	CSX
mante	394613	CSX
marattin	616886078	CSX
marcocappato	128870883	CSX
marcodimaio	48484178	CSX
marcofurfaro	46928097	CSX
mariannamadia	588975097	CSX
matteograndi	44574685	CSX
matteorenzi	18762875	CSX
mattiafeltri	482897816	CSX
maumartina	415571726	CSX
meb	588200416	CSX
mercedesbresso	16905927	CSX
nadiaferrigo	226134529	CSX
nzingaretti	403544693	CSX
paolodecastro	9234892	CSX
paoloroversi	24856912	CSX
patriziaprestip	399010858	CSX
pdnetwork	13294452	CSX
pfmajorino	271596933	CSX
pierofassino	100218289	CSX
pietroraffa	61877999	CSX
pinapic	339672122	CSX
raffaellapaita	455562995	CSX
raistolo	526837239	CSX
rampi	14110067	CSX
riccardomagi	283026832	CSX
ricpuglisi	225168723	CSX
riotta	73971750	CSX
robertapinotti	102672182	CSX
robertosaviano	14516324	CSX
rossipresidente	97734603	CSX
rubio_chef	1269037446	CSX
s_parisi	351675859	CSX
sandrogozi	76769340	CSX

serracchiani	35298549	CSX
simonabonafe	391945446	CSX
sonolucadini	2325244646	CSX
stanzaselvaggia	416476798	CSX
straneuropa	334777310	CSX
toiapatrizia	86719842	CSX
valeriafedeli	480645077	CSX
vannaio	1160083688	CSX
virginiomerola	230817595	CSX
vittoriozucconi	419918470	CSX
vladiluxuria	608534375	CSX
Adnkronos	25676606	DX
Agenzia_Ansa	150725695	DX
Agenzia_Italia	72248630	DX
AnnalisaChirico	117407713	DX
Ariachetira	465390684	DX
Capezzone	144210805	DX
Conad	709759006630350850	DX
DPCgov	982240377636700161	DX
Daniele_Manca	29236274	DX
EASO	325486395	DX
EGardini	404836757	DX
ESA_Italia	117017147	DX
Esercito	437337751	DX
Fontana3Lorenzo	455378815	DX
FratellidItaIia	1024976264	DX
GDF	542914304	DX
GiorgiaMeloni	130537001	DX
Giorgiolaporta	16145985	DX
GiovanniToti	2331718804	DX
ItalianAirForce	1280597143	DX
La7tv	121424128	DX
LegaSalvini	13514762	DX
Leonardo_IT	615029235	DX
MEPvistisen	952935540738928641	DX
Marcozanni86	112873259	DX
MatrixCanale5	1969190317	DX
MediasetTgcom24	331617619	DX
NicolaPorro	399004979	DX
Noiconsalvini	2900789860	DX
OmnibusLa7	546791475	DX
Paolo_Bargiggia	2865436081	DX
QuiMediaset_it	384760605	DX
RaffaeleFitto	702661167	DX
RedCrossEU	484724670	DX
RobertoMaroni_	495277374	DX
SM_Difesa	1919237072	DX
SernagiottoRemo	552711664	DX
SkyTG24	5893702	DX
StefanoFassina	301472434	DX
TgLa7	828717014	DX
UniCredit_IT	4872114203	DX
UniCredit_PR	270982064	DX
UnioneSarda	97643497	DX
Viminale	2585782921	DX
VittorioSgarbi	201658727	DX
Carabinieri	2965811733	DX
DAGOSPIA	70971449	DX
_MiBACT	33502178	DX
agorarai	317769975	DX
ansaeuropa	296371084	DX

borghi_claudio	337767301	DX
coldiretti	626360641	DX
crocerossa	123487422	DX
diMartedi	2695436575	DX
emergenzavvf	702594325625085952	DX
euronewsit	266674649	DX
fpugliese_conad	1513014650	DX
guardiacostiera	142968828	DX
lucatremolada	28315879	DX
lumorisi	433409100	DX
masechi	27456576	DX
matteosalvinimi	270839361	DX
msn_italia	16328363	DX
poliziadistato	20423071	DX
pomeriggio5	3030837185	DX
prattandwhitney	20814874	DX
sputnik_italia	2817304093	DX
tg2rai	526530582	DX
tvsvizzera	2306281964	DX
you_trend	404064077	DX
zaiapresidente	1324893229	DX
Antonio_Tajani	529247064	FI
CalabriaTw	412371227	FI
DeborahBergamin	92556620	FI
GabriGiammanco	714417878	FI
GiusyVersace	447056150	FI
GruppoFICamera	1344131071	FI
IrenePivetti	1127586013	FI
JerryCala	569946597	FI
LiciaRonzulli	318316637	FI
MaxSalini	601773356	FI
PatricielloAldo	454821447	FI
StefanoFeltri	391485689	FI
berlusconi	920277002858500096	FI
corradoformigli	246910613	FI
forza_italia	147543162	FI
msgelmini	420332560	FI
renatobrunetta	20430345	FI
roncellamare	253655087	FI
simonefurlan1	507652780	FI
Affaritaliani	35525950	M5S
AgeaLaura	1977797946	M5S
AlfonsoBonafede	1413168613	M5S
BI_Italia	702901779512610816	M5S
BarillariM5S	107517248	M5S
CatalfoNunzia	572128310	M5S
DalilaNesci	473107684	M5S
DaniloToninelli	960924277	M5S
EleonoraEvi	1135141640	M5S
FMCastaldo	1368850908	M5S
FedericaDaga	361373768	M5S
FedericoDinca	122445010	M5S
FondFeltrinelli	2190353826	M5S
GianlucaVacca	124236178	M5S
GianlucaVasto	318499634	M5S
GiuliaGrilloM5S	236565724	M5S
GiuliaLupo2	1019242497103548416	M5S
GiuseppeConteIT	999578121123848192	M5S
INPS_it	1579745006	M5S
Isa_Adinolfi	2195622679	M5S
ItalyMFA	599114492	M5S

ItalyinMKD	768046646584348673	M5S
ItalyinSerbia	713362911662587904	M5S
LauraBottici	1943781798	M5S
LuigiGallo15	920435114	M5S
M5S_Baroni	2459996144	M5S
M5S_Camera	1354935949	M5S
M5S_Europa	2530314205	M5S
M5S_Senato	1343975432	M5S
MFantinati	49275896	M5S
ManlioDS	208642171	M5S
MartaGrande87	1269713336	M5S
MinLavoro	2291193420	M5S
MinisteroDifesa	384192964	M5S
MinisteroSalute	1904010924	M5S
Mlucialorefice	1623798294	M5S
Montecitorio	2278995820	M5S
Mov5Stelle	289400495	M5S
NicolaMorra63	1314728936	M5S
Palazzo_Chigi	963938472	M5S
PaolaTavernaM5S	497797578	M5S
Patty_LAbbate	986629276458995712	M5S
PediciniM5S	631767111	M5S
Roberto_Fico	22834067	M5S
Roma	459558301	M5S
SergioCosta_min	1003507623214370816	M5S
Sergio_Vaccaro1	228860861	M5S
WeWorldOnlus	1853887003	M5S
ZolezziAlberto	1712358318	M5S
andrea_cioffi	281066500	M5S
ariccardi9	986968199949377536	M5S
baffone5stelle	1509825860	M5S
beppe_grillo	19067940	M5S
c_appendino	963073442	M5S
carlaruocco1	978488840	M5S
carlosibilia	127025568	M5S
crippa5stelle	1283276454	M5S
dellorco85	1638278257	M5S
fanpage	213254229	M5S
fattoquotidiano	52424550	M5S
g_brescia	1326743389	M5S
gianluc_ferrara	2828160313	M5S
gparagone	384927198	M5S
ignaziocorrao	143393223	M5S
lofioramonti	839373926	M5S
lucatelese	100610543	M5S
luigidimaio	48062712	M5S
marcoaffronte	46416760	M5S
marcotravaglio	320603385	M5S
micillom5s	1715119512	M5S
minGiustizia	2867200594	M5S
mirellaliuzzi	11945512	M5S
mitgov	1171761254	M5S
petergomezblog	274093178	M5S
piersileri	506454157	M5S
riccardo_fra	1052596340	M5S
robertalombardi	61727028	M5S
romanopaolo	52071233	M5S
sfnlcd	2330261468	M5S
spatua	319766967	M5S
vilmamoronese	126970118	M5S
virginiaraggi	1530798872	M5S

1giornodapecora	67006079	MINGOs
AGisotti	593857148	MINGOs
ActionAidItalia	34571662	MINGOs
Agenzia_Dire	469398862	MINGOs
Avvenire_Nei	132880191	MINGOs
BeatriceCovassi	2565885655	MINGOs
CAI150	895692594	MINGOs
CaFoscari	108979965	MINGOs
CaroRackete	1149753717564088322	MINGOs
Cild2014	2829501504	MINGOs
Corriere	395218906	MINGOs
EY_Italy	286487408	MINGOs
Einaudieditore	72587784	MINGOs
Europarl_IT	37621971	MINGOs
FiorellaMannoia	135859697	MINGOs
Focolare_org	238081172	MINGOs
FranceauVatican	824615814460559360	MINGOs
Francescorocca	160321413	MINGOs
GDS_it	324818479	MINGOs
GinoStrada	4009984403	MINGOs
GioMelandri	95630884	MINGOs
HolySeePress	2242580173	MINGOs
HuffPostItalia	543774554	MINGOs
IAIonline	237334959	MINGOs
InOndaLa7	499843550	MINGOs
Internazionale	15254807	MINGOs
ItalianNavy	252039915	MINGOs
ItalyinTunisia	595396304	MINGOs
LaityFamilyLife	429721587	MINGOs
MSF_ITALIA	24896038	MINGOs
MarroneEmma	438212411	MINGOs
MattiaBriga	398742952	MINGOs
MauroCasciari	124020256	MINGOs
MiurSocial	882235051	MINGOs
NichiVendola	37631819	MINGOs
OIMItalia	1303472983	MINGOs
OttoemezzoTW	2359413432	MINGOs
PE_Italia	897627558	MINGOs
Panna975	308073259	MINGOs
PiazzapulitaLA7	58453980	MINGOs
PietroGrasso	1071332641	MINGOs
PontAcadLife	808317193918287872	MINGOs
Pontifex	500704345	MINGOs
Pontifex_de	523150102	MINGOs
Pontifex_fr	500713106	MINGOs
Pontifex_it	500711588	MINGOs
Pontifex_pt	965741665	MINGOs
Quirinale	732819391	MINGOs
Radio1Rai	210501383	MINGOs
Radio24_news	126677638	MINGOs
Radio3tweet	100185915	MINGOs
RadioRadicale	74707629	MINGOs
RadiocorriereTv	4706221216	MINGOs
RaiDue	1204076264	MINGOs
RaiNews	14060262	MINGOs
RaiRadio2	94335911	MINGOs
RaiTre	1433286055	MINGOs
Raicomspa	3303600730	MINGOs
Raiofficialnews	538147915	MINGOs
RefugeesChief	4353234441	MINGOs
ReginaCatrambon	27448416	MINGOs

RespSocialeRai	1969148101	MINGOs
RocioMMorales	281214024	MINGOs
RollingStoneita	76913673	MINGOs
Rvaticanaitalia	121480284	MINGOs
SIAE_Official	479602890	MINGOs
SViadiDamasco	634799120	MINGOs
SalernoSal	369400663	MINGOs
SaveChildrenIT	29178622	MINGOs
SimoneCosimi	1408543448	MINGOs
Striscia	2320863902	MINGOs
Tg3web	69959408	MINGOs
TgrAltoAdige	3492840315	MINGOs
TgrCalabria	3401651157	MINGOs
TgrLiguria	706498833430269952	MINGOs
TgrRai	804354408	MINGOs
TgrRaiFVG	4755350878	MINGOs
TgrRaiPuglia	4850084878	MINGOs
TgrRaiToscana	4343971877	MINGOs
TgrRaiTrentino	2319707989	MINGOs
TgrSardegna	3292845743	MINGOs
TgrSicilia	4861946812	MINGOs
TgrVeneto	3292873475	MINGOs
Tommasolabate	122014351	MINGOs
Transport_EU	296253874	MINGOs
TvTalk_Rai	372237180	MINGOs
UNHCRItalia	993225440	MINGOs
UNICEF_Italia	38632492	MINGOs
UNOPS	204698478	MINGOs
Unicatt	284056610	MINGOs
Unomattina	390118819	MINGOs
VauroSenesi	329492939	MINGOs
_Nico_Piro_	429255979	MINGOs
_PaoloRuffini	94146300	MINGOs
_arianna	16562213	MINGOs
ambasciatasvizz	627163262	MINGOs
amnestyitalia	41370439	MINGOs
andreabettini	27260996	MINGOs
antoniospadaro	6062112	MINGOs
bancaetica	247194049	MINGOs
bknsty	2098681	MINGOs
chetempochefa	58748101	MINGOs
civati	14108472	MINGOs
civcatt	246220040	MINGOs
cooperazione_it	1136408917	MINGOs
cpcatrambone	2230241172	MINGOs
davidebanzato	401715194	MINGOs
direzioneprc	24184892	MINGOs
eleonoraforenza	491722490	MINGOs
elvira_serra	421147323	MINGOs
emergency_ong	23615081	MINGOs
espressonline	85852409	MINGOs
europainitalia	90139268	MINGOs
europeaid	19722343	MINGOs
fannicanelles	224940210	MINGOs
fladig	19594408	MINGOs
francescoassisi	115690700	MINGOs
giusmo1	388284615	MINGOs
greenMe_it	36043253	MINGOs
ignaziomarino	58430129	MINGOs
il_piccolo	96987099	MINGOs
ilpost	133790890	MINGOs

ispionline	91138156	MINGOs
lanuovasardegna	121470821	MINGOs
limesonline	9095292	MINGOs
mannocchia	494364900	MINGOs
martaottaviani	61567461	MINGOs
martaserafini	29226134	MINGOs
mattino5	1734567211	MINGOs
micheledisalvo	264467370	MINGOs
moas_eu	2226850327	MINGOs
nicola_pinna	364785591	MINGOs
openpolis	83212310	MINGOs
oss_romano	444921429	MINGOs
paola_saluzzi	476661302	MINGOs
philipdisalvo	203007837	MINGOs
pif_iltestimone	445008112	MINGOs
redazioneiene	239784688	MINGOs
reportrai3	271476597	MINGOs
repubblica	18935802	MINGOs
rtl1025	56341776	MINGOs
silvioderossi	1299371	MINGOs
sole24ore	420351046	MINGOs
sunderland_jude	316466692	MINGOs
tagadala7	3610762348	MINGOs
tigella	7912442	MINGOs
unisiena	435081321	MINGOs
unitorvergata	526559780	MINGOs
valigiablu	146448681	MINGOs
vaticannews_it	291294443	MINGOs
welikeduel	908637973436084224	MINGOs
yaxle	53645670	MINGOs

Table A.2: Verified users list within discursive communities in the 2019 data set. Each user screen name is reported together with its user ID which uniquely identifies a Twitter user.

Appendix B

The Exponential Random Graph formalism

When studying real-world networks, individuating properties that deviate from a properly defined null model is of utmost importance [98]. By analyzing these deviations, non-trivial properties of real-world networks are likely to be deduced. Maximum entropy models are the best candidate for this kind of task since no information about the system is required except for the set of properties $\mathbf{c}^*$ employed as constraints [10]. All the other statistical property of the real-world network $\mathbf{G}^*$ can be validated against a null hypotesis. In the context of pattern validation, the process of building a statistically validated projection of a bipartite network has been widely studied [96].

In the following, the notation corresponding to the previously defined users bipartite networks notation can be easily replaced with that of the user-hashtag bipartite networks.

B.1 Bipartite Configuration Model

The Bipartite Configuration Model (BiCM) extends the Binary Configuration Model (BCM) to the class of bipartite networks. In the same way, BiCM preserves the degrees of nodes in both layers $\top$ and $\perp$. Thus, the corresponding hamiltonian can be rewritten as the sum of two degree sequences, that is:

$$H(\mathbf{G}, \boldsymbol{\theta}) = \sum_{i=1}^{N_\perp} \theta_i k_i + \sum_{\alpha=i}^{N_\top} \eta_\alpha k_\alpha \tag{B.1}$$

where k_i and k_α are the degrees of nodes i and α on the layers $\perp$ and $\top$ while $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ are the Lagrangian multipliers associated with the constraints. It is possible to show that the Shannon entropy maximization leads to writing the probability of the generic matrix $\mathbf{G}$ in the following way:

$$P(\mathbf{G}|\boldsymbol{\theta},\boldsymbol{\eta}) = \frac{e^{-H(\mathbf{G})}}{Z(\boldsymbol{\theta})} \tag{B.2}$$

where $Z(\boldsymbol{\theta})$ is the partition function. Due to the linear constraints reported in Eq. B.1, $P(\mathbf{G}|\boldsymbol{\theta},\boldsymbol{\eta})$ can be rewritten in a factorized form as the product of probability coefficients $p_{i\alpha}$:

$$P(\mathbf{G}|\boldsymbol{\theta},\boldsymbol{\eta}) = \prod_{i=1}^{N_\perp}\prod_{\alpha=1}^{N_\top}\left(\frac{x_i y_\alpha}{1+x_i y_\alpha}\right)^{m_{i\alpha}}\left(\frac{1}{1+x_i y_\alpha}\right)^{1-m_{i\alpha}} = \prod_{i=1}^{N_\perp}\prod_{\alpha=1}^{N_\top} p_{i\alpha}^{m_{i\alpha}}(1-p_{i\alpha})^{1-m_{i\alpha}} \tag{B.3}$$

where $x_i \equiv e^{-\theta_i}$ and $y_\alpha \equiv e^{-\eta_\alpha}$. Thus, the quantity $p_{i\alpha} = \frac{x_i y_\alpha}{1+x_i y_\alpha}$ represents the probability that a link connects nodes i and α on the layers $\top$ and $\perp$. The values of x_i and y_α are estimated by maximizing the probability of observing the given $\mathbf{G}^*$, i.e. through the maximization of the likelihood functional reported in Eq. 4.4. Thus, the Lagrangian multipliers are derived by solving a system of equations:

$$\begin{aligned} k_i^* &= \langle k_i \rangle = \sum_{\alpha=1}^{N_\top} \frac{x_i y_\alpha}{1+x_i y_\alpha} \\ k_\alpha^* &= \langle k_\alpha \rangle = \sum_{i=1}^{N_\perp} \frac{x_i y_\alpha}{1+x_i y_\alpha} \end{aligned} \tag{B.4}$$

where $\{k_i^*\}_{i=1}^{N_\perp}$ and $\{k_\alpha^*\}_{\alpha=1}^{N_\top}$ are the observed degree sequences.

B.2 Bipartite Partial Configuration Model

In bipartite networks, a 'partial' version of the BiCM can be defined by constraining the degree sequence $\mathbf{k}^*$ of a single layer. Starting with this definition, the Bipartite Partial Configuration Model (BiPCM) preserves the degree of the layer $\top$ resulting in the hamiltonian:

$$H(\mathbf{G},\boldsymbol{\theta}) = \sum_{\alpha=1}^{N_\top} \eta_\alpha k_\alpha \tag{B.5}$$

where k_α is the degree of node α on the layer $\top$. The probability of the generic matrix $\mathbf{G}$ factorizes in the form:

$$P(\mathbf{G}|\boldsymbol{\eta}) = \prod_{i=1}^{N_\perp}\prod_{\alpha=1}^{N_\top} p_{i\alpha}^{m_{i\alpha}}(1-p_{i\alpha})^{1-m_{i\alpha}} \tag{B.6}$$

Since the hamiltonian in Eq. B.5 is now simplified, the corresponding coefficient $p_{i\alpha}$ assumes the form:

$$p_{i\alpha} = p_{i\alpha}(\eta_\alpha) = \frac{e^{-\eta_\alpha}}{1+e^{-\eta_\alpha}} = \frac{y_\alpha}{1+y_\alpha} \tag{B.7}$$

The likelihood function $\mathcal{L} = \ln P(\mathbf{G}^*)$ has to be maximized in order to evaluate y_α whose values are now easily obtained:

$$k_\alpha^* = \langle k_\alpha \rangle = \sum_{i=1}^{N_\perp} \frac{y_\alpha}{1+y_\alpha} \Rightarrow p_{i\alpha} = \frac{k_\alpha^*}{N_\perp} \tag{B.8}$$

Accordingly, the probability of observing each motif defined by the nodes α and β has the same probability $p(V^i_{\alpha\beta}) = \frac{k_\alpha^* k_\beta^*}{N_\top^2}$ regardless of the node α. This implies that the probability distribution of the number of V-motifs shared by the nodes α and β is now a Binomial distribution defined as:

$$f_{\text{Bin}}(V_{\alpha\beta} = n) = \binom{N_\top}{n} \left(\frac{k_\alpha^* k_\beta^*}{N_\top^2}\right)^n \left(1 - \frac{k_\alpha^* k_\beta^*}{N_\top^2}\right)^{N_\top - n} \tag{B.9}$$

B.3 Bipartite Random Graph Model

The Bipartite Random Graph Model (BiRGM) extends the Random Graph Model (RGM) to the bipartite networks. The present model is easily obtained by constraining the expected total number of edges E leading to an ensemble in which all pairs of nodes are connected with a fixed probability coefficient $p_{\text{BiRG}} = \frac{L}{N_\top \cdot N_\perp}$ where $N_\top$ and $N_\perp$ are the total number of nodes, respectively, on the layers $\top$ and $\perp$. It is worth noting that fixing the probability coefficient p_{BiRG} to all nodes connections is a global constraint which implies that the degree distribution follows a binomial law.

Since the probability coefficients of each pair of nodes is equal, it is straightforward to conclude that the probability of a single V-motif $V^i_{\alpha\beta}$ assumes the form:

$$p(V^i_{\alpha\beta}) = p^2_{\text{BiRG}} = \left(\frac{E}{N_\top \cdot N_\perp}\right)^2 \tag{B.10}$$

Thus, the probability distribution of the number of V-motifs shared by nodes α and β is simply a Binomial distribution:

$$f_{\text{Bin}}(V_{\alpha\beta} = n) = \binom{N_\top}{n} \left(p^2_{\text{BiRG}}\right)^n \left(1 - p^2_{\text{BiRG}}\right)^{N_\top - n} \tag{B.11}$$

Appendix C

Analysis of network mesoscale structures

Mesoscale structures play a fundamental role for understanding the resilience of the system or when investigating the emergence of collective behaviors. An example of mesoscale structure is the *community structure* characterizing a huge amount of real-world networks. Although there is not a univocal definition of community [111], in general this structure refers to a network organization where groups of nodes are more thightly connected to each other than to members of other groups.

Other examples of mesoscale structures have been studied to reveal a finer grained organization of nodes. For instance, the *core-periphery structure* is a well-known configuration consisting of a tightly connected group of nodes, called *core*, surrounded by a region of low-degree nodes preferentially connected to the core, called *periphery*. The next Sections are devoted to the description of the community and the core-periphery detection algorithms which are employed in the analysis of the semantic and users networks.

C.1 Community detection: the Louvain algorithm

Amongst the community detection techniques, the most popular class of algorithms employ a function indicating the quality of a partition over the space of all clusterings. The quality function most commonly-used is the *modularity*, originally defined by Newman and Girvan [135]. The modularity measures the optimal value of a given partition by comparing the empirical number of edges with the number predicted by a properly defined benchmark model.

One of the modularity community detection algorithms is the Louvain algorithm [112] which has been run in this work to detect the presence of communities within the monopartite networks. This algorithm works by searching for the partition that attains the maximum value of the following formula:

$$Q = \frac{1}{2E} \sum_{\alpha,\beta} \left[m_{\alpha\beta} - \frac{k_\alpha k_\beta}{2E} \right] \delta_{c_\alpha, c_\beta} \tag{C.1}$$

where E is the total number of edges. In the expression above, $m_{\alpha\beta}$ is the generic entry of the network adjacency matrix $\mathbf{M}$ while the factor $\frac{k_\alpha k_\beta}{2L}$ is the probability of a connection between nodes α and β according to the Chung-Lu model. This model corresponds to a random network structure where the edges of the network are rewired to preserve the degree of all nodes, on average. The Louvain algorithm performs a greedy optimisation of Q by assigning each node to the community of their neighbors with the largest Q and creating a smaller weighted network whose nodes are the clusters found in the previous step. The procedure is repeated until the largest modularity is reached. Due to the dependence of the order of the nodes taken as an input, after having applied a nodes reshuffling, the Louvain algorithm has been run 100 times to prevent the modularity function from reaching a local maximum.

C.2 Core-periphery detection procedure

The first formalization of the core-periphery structure was realized by Borgatti and Everett [136] who define a score function indicating the deviation of a given network partition from an ideal core-periphery configuration defined as a fully connected core region and a periphery region which is only linked to the core. Differently from this approach, core-periphery detection can be carried out upon adopting another class of algorithms which employ benchmark models to compare with the original network structure.

An algorithm to detect a statistically significant core-periphery structure is proposed in [131]. The procedure described in this work prescribes the search for the network partition minimizing a quantity called *bimodular surprise*, i.e.

$$\mathscr{S}_{\|} = \sum_{i \geq e^*_\bullet} \sum_{j \geq e^*_\circ} \frac{\binom{V_\bullet}{i}\binom{V_\circ}{j}\binom{V-(V_\bullet+V_\circ)}{E-(i+j)}}{\binom{V}{E}}; \tag{C.2}$$

The quantity above is the multinomial version of the *surprise*, originally proposed to carry out a *community detection* exercise. In the present case, E is the total number of edges observed in monopartite projections, while V is the total number of possible edges, i.e. $V = \frac{N(N-1)}{2}$. The quantities marked with $\bullet$ ($\circ$) refer to the corresponding core (periphery) quantities: for example, $V_\bullet$ is the total number of possible core edges, $V_\circ$ is the total number of possible periphery edges, $e^*_\bullet$ is the number of observed edges within the core and $e^*_\circ$ is the number of observed edges within the periphery. The presence of three different binomial coefficients allows three different 'species' of edges to be accounted for: the binomial coefficient $\binom{V_\bullet}{i}$ enumerates the number of ways i edges can be redistributed *within* the core, the binomial coefficient $\binom{V_\circ}{j}$ enumerates the number of ways j edges can be redistributed *within* the periphery and the binomial coefficient $\binom{V-(V_\bullet+V_\circ)}{E-(i+j)}$ enumerates the number of ways the remaining $E-(i+j)$ edges can be redistributed *between* the two, i.e. over the remaining $V-(V_\bullet+V_\circ)$ node pairs.

From a technical point of view, $\mathscr{S}_{\|}$ is the p-value of a multivariate hypergeometric distribution, describing the probability of $i+j$ successes in E draws (without replacement), from a finite population of size V that contains exactly $V_\bullet$ objects with a first specific feature and $V_\circ$ objects with a second specific feature, wherein each draw is either a 'success' or a 'failure': analogously to the univariate case, $i+j \in [e^*_\bullet + e^*_\circ, \min\{E, V_\bullet + V_\circ\}]$. The method results in the most statistically significant core-periphery structure compatible with the network under analysis.