Computational Framework for the Assessment of New Forms of Political Organization in Social Media

Pablo Aragón

Escuela Técnica Superior de Ingeniería Informática

Universidad Nacional de Educación a Distancia

This dissertation is submitted for the degree of

Máster en Inteligencia Artificial Avanzada

Supervised by:

Andreas Kaltenbrunner

Luis Manuel Sarro Baro

June 2016

To my beloved Marion

Acknowledgements

Este Trabajo Fin de Máster (TFM) es el resultado de una etapa que se inicia en mi llegada a Barcelona en julio de 2011. Desde entonces son muchas las personas que generosamente han compartido ideas y experiencias que, directa o indirectamente, explican el trabajo aquí realizado. La razón me aconsejó que, para maximizar la visibilidad de este artículo en el ámbito académico, convenía redactarlo en inglés. Sin embargo, esta sección de agradecimientos a todas estas personas está escrita con el corazón. Permítame entonces el lector su redacción en castellano.

En primer lugar, este estudio ha contado con el privilegio de ser dirigido por *Andreas Kaltenbrunner*. He de agradecer a Andreas todos sus comentarios y sugerencias, no sólo en este estudio ni exclusivamente en el ámbito académico. También quisiera agradecer a *Luis Sarro*, profesor de la UNED, su co-dirección del TFM, especialmente por sus propuestas para adecuar el formato paper original al formato tesis. En el formato paper tuve la suerte de contar con mis colegas *Yana Volkovich* y *David Laniado* como colaboradores. He de agradecer a Yana su empuje para creer en esta idea y su meticulosa revisión del texto. A David tengo que agradecerle sus excelentes sugerencias y su generosa amistad. Además, muchas ideas que motivan este estudio surgen del la línea de trabajo del grupo de Digital Humanities en Eurecat (anteriormente grupo Social Media en Barcelona Media). Por tanto, quisiera reconocer la contribución de mis colegas *Carlos Ullod*, *David Comas*, *Ruth García-Gavilanes*, *Jessica G. Neff*, *Karolin Kappler*, *Matteo Manca*, *Ariadna M. Fernández* y *Ricardo Ruiz de Querol*. Y

por supuesto, muchas gracias al resto de personas de Eurecat - Barcelona Media por todos estos años.

En segundo lugar, la inspiración para este estudio también parte del trabajo colectivo realizado en *DatAnalysis15M*, una de las experiencias más enriquecedoras en las que he tenido la suerte de participar. Muchas gracias a los nodos de esta red de investigación: *Javier Toret*, *Arnau Monterde*, *Antonio Calleja*, *Miguel Aguilera*, *Xabier Barandiaran*, *Alberto Lumbreras*, *Óscar Marín*, *Eunate Serrano*, *Emanuele Cozzo*, *Ignacio Morer*, *Luce Prignano*, *Juan Linares*, *Ariadna Fernández* y *Pablo Rey*. Quisiera destacar la contribución de Toret, una fuente constante de estímulos y compromiso. También he de agradecer de manera especial las enseñanzas de Arnau sobre la reapropiación de la academia para promover cambios. Ese espíritu de cambio se ha visto reforzado por el proyecto europeo *D-CENT*, que precisamente ha financiado este estudio. Quisiera agradecer la dedicación y entrega de su coordinadora *Francesca Bria*, de investigadores e ingenieros tan excepcionales como *Robert Bjarnason*, *Gunnar Grimsson*, *Jaromil*, *Marco Sachy*, *Stefano Lucarelli*, *Klara Jaya Brekke*, y del resto de personas que han alimentado el proyecto. A su vez, este proyecto ha dado pie al lanzamiento de otra red de investigación: *BetaDemIC*. En ella he conocido a personas tan sabias y generosas como *Alberto Abellán*, *Pedro Álvarez Roxu*, *Miguel Arana*, *José Luis Aznarte*, *Alejandra Baciero*, *Yago Bermejo*, *María G. Navarro*, *Tomás Gómez*, *Gonzalo Polavieja*, *Nuria del Río*, *Eduardo Robles*, *Eduardo Romanos*, *David Ruescas*, *Saya Sauliere*, *Pablo Soto* e *Iván Villarubia*. También he de agradecer a mis colaboradores *Mariluz Congosto* e *Ismael Peña-López* por sus ideas y recursos aportados. Muchas gracias a todas por vuestras contribuciones, en lo académico pero sobre todo en lo humano.

En tercer lugar, como comenté en el inicio, este máster va inexorablemente ligado a mi vida en Barcelona y a las personas que me acompañan en ella. He de agradecer a mis Hermanos Dalton *Gabriel Cisneros*, *Mario Benítez* y *Roberto Velasco* por evitar varias

sobredosis de estudio en la sangre (“es parte de mi alma”). Este agradecimiento también se amplía consecuentemente a la extensión Dalton en *Familia Maña*: *Aída Sanz*, *Cristy Escuer* y *Anna Giordanni*. Mi primeros años en Barcelona se asemejaron a los de un cangrejo ermitaño, cambiando periódicamente de piso. En estos cambios he tenido la ocasión de forjar amistades con personas tan maravillosas como *Nuria Alarcón*, *Nuria Coma*, *Úrsula Marcilla*, *Alba Cambronero* y *Alba Martí*. En paralelo, los cursos de catalán se convirtieron espontáneamente en una oportunidad de compartir grandes momentos con *Gonzalo Correa*, *Andrés Cruz*, *Cris Cruz*, *Alexis Ferrer*, *Javi Izquierdo*, *Ramona Galatola*, *Elisa Raposo*, *Marta Ribas* y *Natalia Rodríguez*. Muchas gracias a todas vosotras por vuestra amistad.

En cuarto lugar, en este máster, como en muchos procesos de mi vida, mi familia ha sido un puntal destacado. Tengo que agradecer a mi padre *José Antonio*, a mi madre *María del Carmen*, a mi hermana *Lucía*, a mis abuelos *Tasio* y *Carmen*, y al resto de mi familia por apoyarme.

Por último, y por ello de manera más trascendental, si hay una persona fundamental en esta etapa y en las que quedan por llegar es *Marion*. Agradecer su amor, paciencia y generosidad incondicional exige una extensión superior al texto de este TFM. Muchas gracias por cambiar mi vida y por dejar que cambiase la tuya, no hay día en el que no me sienta la más afortunada de las personas. *Merci beaucoup mon amour*.

Abstract

Social media has become a key mechanism for the organization of grassroots movements. In the 2015 Barcelona City Council election, Barcelona en Comú, an emerging grassroots party, was the most voted one. This candidacy was devised by activists involved in the Spanish 15M movement in order to turn citizen outrage into political change. On the one hand, the 15M movement is based on a decentralized structure. On the other hand, political science literature postulates that parties historically develop oligarchical leadership structures. This tension motivates to examine whether Barcelona en Comú preserved a decentralized structure or adopted a conventional centralized organization.

This thesis proposes a computational framework to (1) better identify the Twitter networks of political parties by setting a confidence parameter in a popular community detection algorithm of the state of the art, and (2) analyze each party network by measuring the hierarchical structure, small-world phenomenon and coreness. The evaluation of the framework on the Twitter networks of the parties that ran for this election shows that the extended community detection algorithm effectively set boundaries between party networks. Although most parties consist of one single cluster, in Barcelona en Comú two well-defined groups co-exist: a cluster dominated by the party leader and the collective accounts, and another cluster formed by the movement activists. While the former group is highly centralized like traditional parties, the latter one stands out for its decentralized, cohesive and resilient structure.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

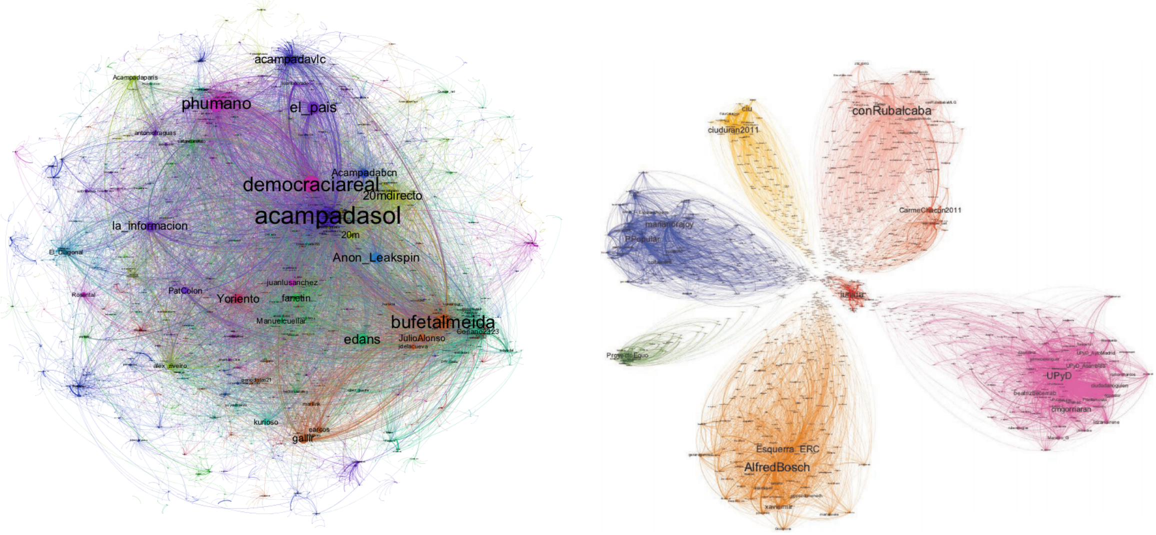
1.1 Motivation and related work

In the last years a new global wave of citizen protests has emerged: the Arab Spring, the 15M movement in Spain, Occupy Wall Street, #YoSoy132 in Mexico, Occupy Gezi in Turkey, the Brazilian movement #VemPraRua, Occupy Central in Hong Kong, etc. All these movements share common characteristics such as the claim for new models of democracy, the strategic usage of social media (e.g. Twitter), and the occupation of physical spaces. Also, all of them have encountered difficulties in modifying the institutional agenda and, hence, the public policies. The 2015 Barcelona City Council election is one of the first cases in which one of these movements has got to "occupy" the public institutions by building Barcelona en Comú (BeC), a political party that won the elections. BeC was conceived as the confluence of (1) minor and/or emerging parties and, to a large extent, (2) collectives and activists, with no political party affiliation, who played a prominent role in the 15M movement.

The 15M movement, also referred to as #SpanishRevolution or the "Indignados" movement, emerged in May 2011 and has been defined as a "networked social movement of the digital age" [9]. Networked social movements, like the Arab Spring, the 15M and Occupy

Wall Street, are claimed to be “a network of networks, they can afford not to have an identifiable center, and yet ensure coordination functions, as well as deliberation, by interaction between multiple nodes” [9]. Other authors have formulated similar hypotheses defining this new model of social movement as a “change from logic of collective action, associated with high levels of organizational resources and the formation of collective identities, to a logic of connective action, based on personalized content sharing across media networks” [4]. Some voices have refused these theoretical assumptions and argued that “a handful of people control most of the communication flow” and, consequently, the existence of leaders in such movements could not be denied [19]. Empirical studies revealed that the 15M network on Twitter is characterized by its “decentralized structure, based on coalitions of smaller organizations” in spite of “a small core of central users is still critical to trigger chains of messages of high orders of magnitude” [22]. Decentralization has been also observed in [49] in which the 15M network is defined as polycentric (see Figure 1.1a).

The 15M network properties (i.e. decentralization, polycentrism) could be perceived as a striking contrast to conventional political organizations, in particular, political parties. The Iron Law of Oligarchy [33] postulates that political parties, like any complex organization, self-generate an elite (i.e. “Who says organization, says oligarchy”). Although some scholars have criticized the idea that organizations will intrinsically build oligarchical leadership structures [31, 44, 12], many political and social theorists have supported that, historically, small minorities hold the most power in political processes [42, 36, 34]. Regarding Spanish politics, a study of the 2011 national election campaign on Twitter revealed that “minor and new parties tend to be more clustered and better connected, which implies a more cohesive community” [2] (see Figure 1.1b). Nevertheless, all the diffusion networks of parties in that study were strongly centralized around their candidate and/or party profiles. Later studies analyzed the interactions on Twitter between the 15M nodes and political parties, and conclude that networked social movements are *para-institutions*: perceived as institutions but preserv-



(a) 15M movement in May 2011. Source [49].

(b) Members of Spanish political parties in the 2011 national election. Source [2].

Fig. 1.1 Political retweet networks.

ing an internal networked organization [43]. However, these conclusions were formulated by analyzing the networks when no elections were held, before institutionalization began. Campaigns are competitive processes that might favor the centralization of an organization around candidates. Indeed, it has been proven that the network properties of political parties change when elections arrive [16]. Given that Barcelona en Comú emerged from the 15M and this networked movement is characterized by a decentralized structure, the research question that motivates this study is the following: ***Has Barcelona en Comú preserved a decentralized structure or has it adopted a conventional centralized organization ruled by an elite?***

Previous hypotheses [48] about Podemos, a member party of the Barcelona en Comú candidacy and as well inspired by the 15M movement, postulate an organization formed by a *front-end* ("spokesmen/spokeswomen who are visible from the media perspective") and a *back-end* ("muscle of the organization, barely visible from the media perspective"). However, there are no empirical validations of this hypothesis.

1.2 Objectives

Motivated by the research question, this thesis proposes a computational framework for community detection and cluster characterization on Twitter political networks.

1.2.1 Community detection

The identification of the sub-network corresponding to each party is based on the partisan structure of Twitter networks. Data-driven political science has revealed the recurrent existence of boundaries between ideological online communities, in particular, political parties. A study of the 2004 U.S. Presidential election depicted a divided blogosphere in which liberals and conservatives barely generated links between the two communities [1]. Similarly, the network of retweets for the 2010 U.S. congressional midterm elections exhibited a highly segregated partisan structure where connections between left- and right-leaning users were extremely limited [11]. Both studies have been taken as relevant empirical validations of the so-called *cyber-balkanization*, a social phenomenon that occurs when Internet users form isolated groups around specific topics (e.g. political interests). This concept is closely related to the idea of *echo chambers*, in which people are "mainly listening to louder echoes of their own voices" [47] and, therefore, reinforce division in social media. Indeed, online polarization is not only a particular feature of U.S. politics but also a social behaviour observed in a diverse range of countries, e.g. Canada [24] and Germany [13]. In Spain, previous studies of Twitter networks related to recent elections also showed evidence of online polarization, e.g. in the 2010 Catalan election [10] and in the 2011 Spanish elections [6]. In contrast, it is important to note that some relevant nodes do not follow the cyber-balkanization trend. Some Twitter accounts, e.g. media and social organizations, are not aligned with political parties but users from political clusters might diffuse their content. Therefore, community detection must must exclude such nodes to truly identify the network of each political party.

Although users from political parties might diffuse their content, the identification of the party structure must exclude such nodes for not being taken into account in the cluster characterization process.

Contribution to the state of the art Given the network topology described above, community detection is proposed in this thesis by the refinement of a state of the art technique: the Louvain method [5]. This method, that attempts to optimize the modularity of a partition of the network, is one of the most popular ones because of its efficiency in large networks. Nevertheless, the greedy design of the method, based on a random ordering of nodes and edges, makes that clusters often include nodes which might overlap several communities (e.g. a media account retweeted by members from different political clusters). For this reason, this thesis defines a new version of the Louvain method by setting a confidence parameter that effectively avoids this drawback. The results are also compared to the ones obtained through the Clique Percolation Method (CPM) [41], a state of the art method for community detection in overlapping communities. The proposed method outperforms CPM in terms of computational complexity and clustering recall.

1.2.2 Cluster characterization

Clusters are subgraphs and, therefore, clusters can be characterized by their network structure. Network science is an academic field, emerged from the Euler's work on graph theory in the 18th century, that is attracting increasing interest in the last decades. Among the wide array of network metrics in the state of the art, some studies have featured some of them as the most suitable ones to characterize the organization of political networks. In particular, a recent study on a online platform for politicians suggested the analysis of the internal social network of political parties through metrics related to hierarchical structure, information efficiency, and social resilience [16].

Contribution to the state of the art

This thesis proposes cluster characterization by refining the analysis suggested in [16]. First, hierarchical structure was originally measured by the in-degree centralization [15]. This thesis mathematically proves that such metric is not suitable for social graphs because of the heavy tailed distribution of the in-degree. The experiments conducted in this work show that inequality of the in-degree distribution, measured by the Gini coefficient [20], better captures the hierarchical structure of a political network. Second, information efficiency in social graphs is explained by their small-world phenomenon [50]. Therefore, this thesis proposes the inclusion of the clustering coefficient as a reliable metric of how efficiently nodes are able to diffuse information in a political network. Third, social resilience was originally computed through the k-core decomposition. The coreness is certainly an suitable metric of social resilience but this thesis proposes not only the maximum value, as done in [16], but also the distribution and average k-index of every political network.

1.3 Publications

This work was accepted as a full paper at the International Association for the Advancement of Artificial Intelligence Conference on Web and Social Media (ICWSM-2016)[1]. The article was subject to strict double-blind peer review process and this conference edition had a full paper acceptance rate around 17%.

Fig. 1.2 Banner of the ICWSM-16 conference.

[1] Web of the conference at: http://www.icwsm.org/2016/

In addition, this research is funded by the EU project D-CENT[2] (FP7 CAPS 610349) and the results are contained in the project deliverable *"D2.3 - When a movement becomes a party"*.

Fig. 1.3 Banner of the D-CENT project.

1.4 Structure of the thesis

The thesis is structured in four chapters. This first chapter has introduced the motivation, related work, objectives, contributions to the state of the art, and academic publications. In Chapter 2 the computation framework is described by defining the metrics that are used for community detection and cluster characterization. Chapter 3 presents the evaluation of the computational framework on a dataset of tweets related to the 2015 Barcelona City Council election. Finally, Chapter 4 summarizes the conclusions and future work derived from this thesis.

[2]Web of the project at: http://dcentproject.eu/

Chapter 2

Computational Framework

This chapter describes the computational framework that, given a network of retweets, detects the major clusters (i.e. political parties) and characterizes their social structures along three dimension: hierarchical structure, small-world phenomenon and coreness.

2.1 Community detection

Social networks are characterized by the ability of nodes to cluster into groups with a high density of intra-group edges and a much lower density of inter-group edges. Community detection methods are designed to give a partition of the graph in clusters such that (1) each node belongs to a cluster and (2) the majority of edges occurs between nodes from the same cluster. The state of the art includes a diverse range of methods for this goal, e.g:

- Hierarchical clustering
- Modularity-based methods
- Methods based on statistical inference
- Methods to find overlapping communities

According to [14], one of the most popular algorithms is the one proposed by Girvan and Newman [40] which follow these four steps:

1. Computation of the centrality for all edges;
2. Removal of edge with largest centrality
3. Recalculation of centralities on the running graph;
4. If the the graph still contains edges, jump to step 2.

In this method centrality is measured by the edge betweenness: the number of shortest paths between all node pairs that run along the edge. That is to say that the method finds the community structure by removing the edges according their importance in linking communities (see Figure 2.1).

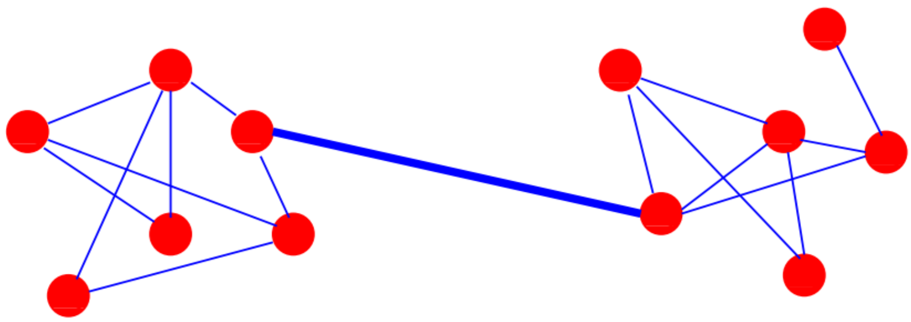


Fig. 2.1 Edge betweenness is higher for edges connecting communities. In the figure, the edge in the middle has a much higher betweenness than all other edges, because all shortest paths connecting nodes of the two communities run through it. Source [14].

Although this technique effectively finds the community structure of a given graph, the $O(mn)$ complexity of the calculation of the betweenness of all edges makes this method impracticable for large networks. This situation motivates the definition of a method based on a much faster algorithm: the Louvain method [5]. Below this section introduces the Louvain method and the refinement proposed in this thesis by setting a confidence interval. The Clique Percolation Method [41], a popular method to find overlapping communities, is also described as it will be used to compare the obtained results.

2.1.1 Louvain method

Many previous studies have relied on the Louvain method [5] because of its high performance in terms of accuracy, and its efficiency ($O(m)$ computational complexity). This method is based on a greedy algorithm that attempts to optimize the modularity of a partition of a given network. Modularity function measures the density of edges inside communities in comparison to edges between communities [38]. Given a network, the modularity value, between -1 and 1, is defined as:

$$Q = \frac{1}{2m}\Sigma_{ij}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta(c_i, c_j)$$

where A_{ij} is the edge weight between nodes i and j; k_i and k_j are the degrees of the nodes i and j, respectively; m represents the total number of edges in the network; c_i and c_j are the indexes of communities of those nodes; and δ is the Kronecker delta.

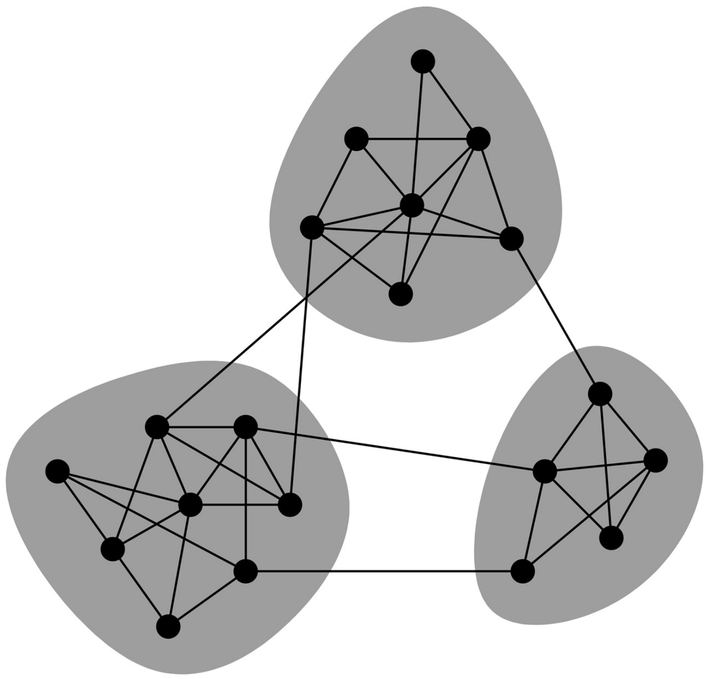


Fig. 2.2 Modularity is based on the idea that the nodes in many networks fall naturally into groups or communities, sets of nodes (shaded) within which there are many edges, with only a smaller number of edges between nodes of different groups. Source [39].

The Louvain method follows a two-step approach. First, each node is assigned to its own community. Then, for each node i, the change in modularity is measured for moving i from its own community into the community of each neighbor j:

$$\Delta Q = \left[\frac{S_{in} + k_{i,in}}{2m} - \left(\frac{S_{tot} + k_i}{2m}\right)^2\right] - \left[\frac{S_{in}}{2m} - \left(\frac{S_{tot}}{2m}\right)^2 - \left(\frac{k_i}{2m}\right)^2\right],$$

where S_{in} is the sum of all the weights of the intra-edges of the community where i being moved into, S_{tot} is the sum of all the weights of the edges to nodes of the community, k_i is the degree of i, $k_{i,in}$ is the sum of the weights of the edges between i and other nodes in the community, and m is the sum of the weights of all edges in the network. Once this value is measured for all communities that i is linked to, the algorithm sets i into the community that produces the largest increase in modularity. If no increase is possible, i remains in its original community. This process is applied until modularity cannot be increased and a local maximum of modularity is achieved. Then, the method groups the nodes from the same community and builds a new network where nodes are the communities from the previous step. Both steps are repeated until modularity cannot be increased.

2.1.2 Louvain method with a confidence interval

The Louvain method is a greedy algorithm and has a random component, so each execution produces a different result. To obtain robust results, avoiding dependency on a particular execution of the algorithm, this thesis introduces the following modification to identify the main clusters of the network in a robust way.

First, it runs N executions of the Louvain algorithm, which produce N different partitions of the network into clusters. Then it selects the bigger clusters for each partition, and identifies each cluster through the most representative node. In particular, since it is expected that

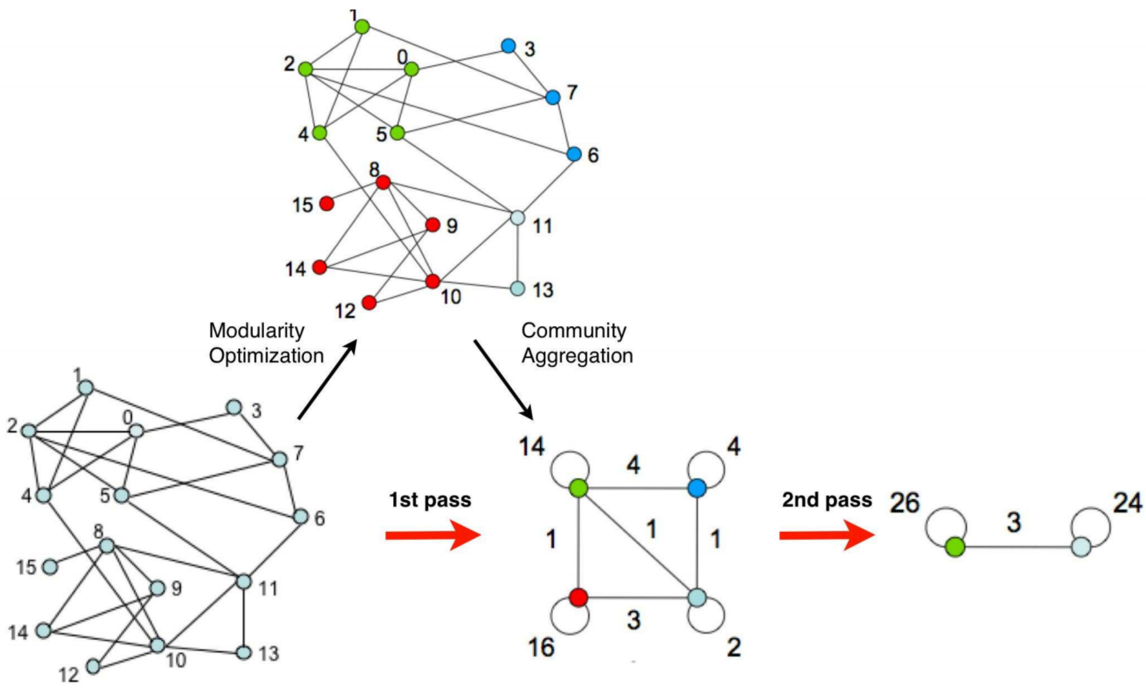


Fig. 2.3 Visualization of the steps of the Louvain method. Source [5].

the main clusters will represent the political parties, it identifies each cluster with the most relevant node according to its centrality measure PageRank [8] (expected to be the account of a political party or a political party leader). Finally, the method assigns to each cluster all the nodes that appear in that cluster in at least a fraction $(1-\varepsilon)$ of the partitions created, that is to say that ε represents the error. This procedure allows to validate the results of the community detection algorithm, and to guarantee that all the nodes that are assigned to a cluster do actually belong to it with high confidence. The remaining nodes, that cannot be assigned in a stable way to any of the main clusters, are left out from all the clusters (in this study, $\varepsilon = 0.95$).

The computational complexity of this algorithm is given by the complexity of the original Louvain method ($O(m)$) and the number of executions (N), that is to say a computational complexity of $O(Nm)$.

PageRank

As commented above, the identification of the most relevant node of each major cluster is done by PageRank. This metric is a global characteristic of a node participation in some network and could be seen as a characteristic of node's success and popularity [8]. It is defined as a stationary distribution of a random walk on the directed graph. At each step, with probability c, the random walk follows a randomly chosen outgoing edge from a node, and with probability $(1-c)$ the walk starts afresh from a node chosen uniformly among all nodes. The constant c is called damping factor, and takes values between 0 and 1 (traditionally c=0.85). PageRank can be summarized in the following formula:

$$PR(i) = c \sum_{j \to i} \frac{1}{d_j} PR(j) + \frac{1-c}{n},$$

where $PR(i)$ is the PageRank of node i, d_j is out-degree of node j, the sum is taken over all nodes j that link to node i, and n is the number of nodes in the network. Unlike in- and out-degree which are local characteristics, the PageRank is a global characteristic of a node. In other words, adding/removing an edge between two nodes could affect PageRank values of many nodes.

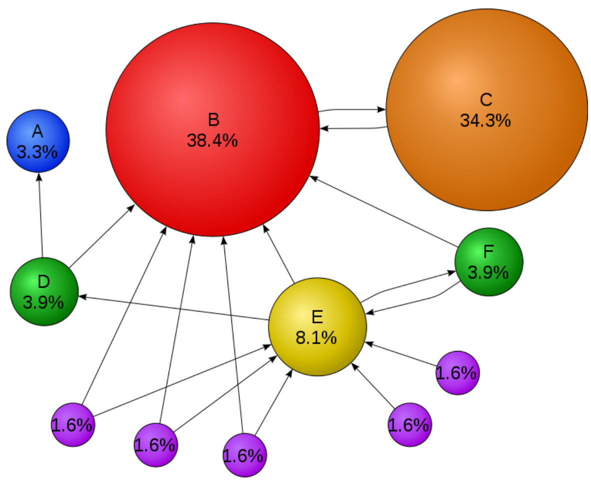


Fig. 2.4 PageRank values (as percentages) for a simple network. Source: Wikipedia.

2.1.3 Clique Percolation Method

The Clique Percolation Method (CPM) [41] is, according to [14], the most popular technique to find overlapping communities. The method is based on the idea that intra-group edges, due to their high density, are likely to form cliques (subsets of nodes in which every pair of nodes are connected). In contrast, inter-group edges unlikely form cliques. Given a parameter k, the Clique Percolation Method follows these three steps:

1. Find all the cliques of size k.
2. Construct a clique graph: two cliques are adjacent if they have $k-1$ nodes in common.
3. Each community corresponds to each connected component of the clique graph (see Figure (see Figure 2.5).

The interest of this method lies in the approach of community detection within overlapping communities, as the method proposed in this thesis. Nevertheless, it is important to note the $O(exp(n))$ computational complexity of CPM. This is explained by the complexity of finding large cliques in graphs. Indeed, the maximum clique problem is one the Karp's 21 NP-complete problems [26].

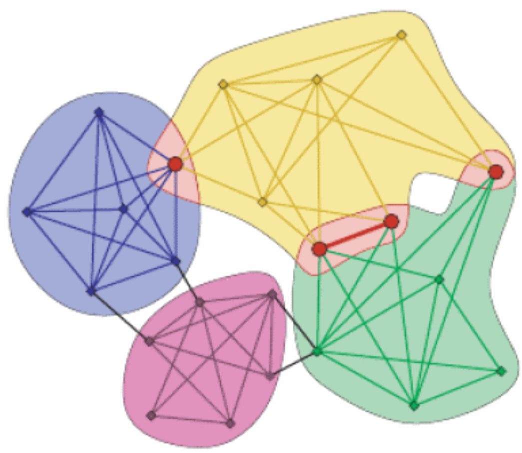


Fig. 2.5 The example shows communities spanned by adjacent 4-cliques. Overlapping nodes are shown by the bigger dots. Source: [41].

2.2 Cluster characterization

Inspired by the social dimensions and metrics suggested in [16], this thesis proposes some new metrics to better characterize the topology of the intra-network of each cluster.

2.2.1 Hierarchical structure

The hierarchical structure is quantified on the in-degree distribution of each cluster. The in-degree of node i is the total number of edges onto node i. By counting how many nodes have each in-degree value, the in-degree distribution $P(k_{in})$ is equal to the fraction of nodes in the graph with such in-degree k_{in}. The cumulative in-degree distribution $P(K \geq k_{in})$ represents the fraction of nodes in the graph whose in-degree is greater than or equal to k_{in}.

The original framework [16] used a existing method to measure degree centralization introduced by [15]. Degree centralization is based on two concepts:

1. How the centrality of the most central node exceeds the centrality of all other nodes
2. Setting the value as a ratio by comparing to a star network:

$$C_{in} = \frac{\sum_{i=1}^{n} [k_{\max}^{in} - k_i^{in}]}{\max \sum_{i=1}^{n} [k_{\max}^{in} - k_i^{in}]},$$

   where k_i^{in} is the in-degree of node i, k_*^{in} is the maximum in-degree of the network and $\max \sum_{i=1}^{n} [k_*^{in} - k_i^{in}]$ is the maximum possible sum of differences for a graph with the same number of nodes (a star network).

In contrast, as demonstrated later, this thesis proposes to evaluate the hierarchical structure by applying the Gini coefficient, a statistical metric to quantify the level of inequality given a distribution [20]. It was initially formulated in Economics to measure the income distribution by using the Lorenz curve. The Gini coefficient is equal to

$$G_{in} = A/(A+B),$$

where A is the area between the line corresponding perfect equality and B is the area under the Lorenz curve. If the Lorenz curve is expressed by the function $y = L(x)$, B is calculated as $B = 1 - 2\int_0^1 L(x)dx$ and $A = 1/2 - B$. In the context of network topology, the Gini coefficient is applied to characterize the hierarchical structure of a network based on the inequality of its in-degree distribution.

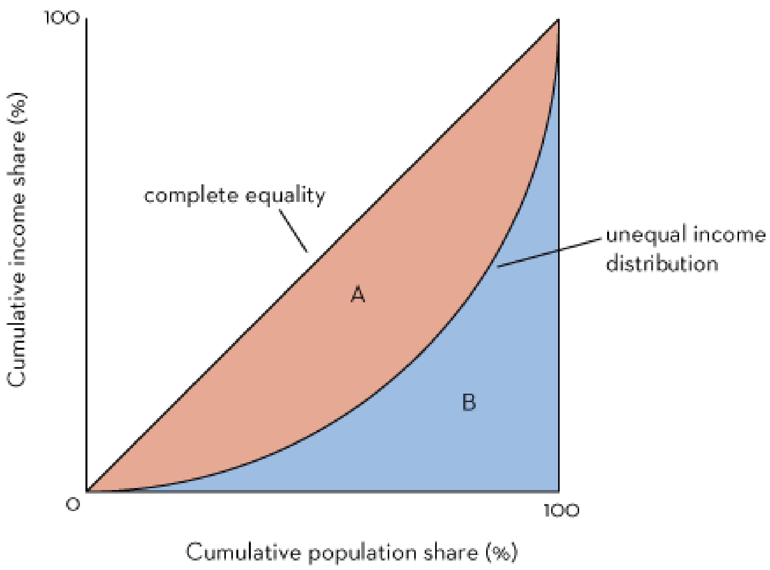


Fig. 2.6 Graphical representation of the Gini coefficient. Source: http://persquaremile.com/.

2.2.2 Small-world phenomenon

The small-world phenomenon states that most nodes of a network are reachable from any other node in a small number of steps and explains information efficiency in social networks. To assess the small-world phenomenon in each cluster, the clustering coefficient and the average path length are computed. Small-world networks tend to have a small average path length and a clustering coefficient significantly higher than expected by random chance [50]. The clustering coefficient measures the extent of nodes to cluster together by calculating the number of triangles in the network. For every node i it set N_i to be the neighborhood, i.e. $N_i = \{j \in V : (i, j) \in E\}$, and defines the local clustering coefficient as

$$Cl_i = \frac{2|(j,k) \in E : j,\, k \in N_i|}{k_i(k_i - 1)}.$$

Then, following [50], the clustering coefficient is just the average of the local clustering coefficients: $Cl = \sum_i Cl_i/n$, where n is the number of nodes in the network. To calculate the average path length, for every pair of nodes i and j, it set d_{ij} to be the smallest number of steps among all paths between i and j. This metric is applied to the clusters identified by the new version algorithm for community detection and, by definition, there is always a path between any pair of nodes in every cluster. The average path length is defined as follows:

$$l = \sum_{i \neq j} d_{ij}/n(n-1)$$

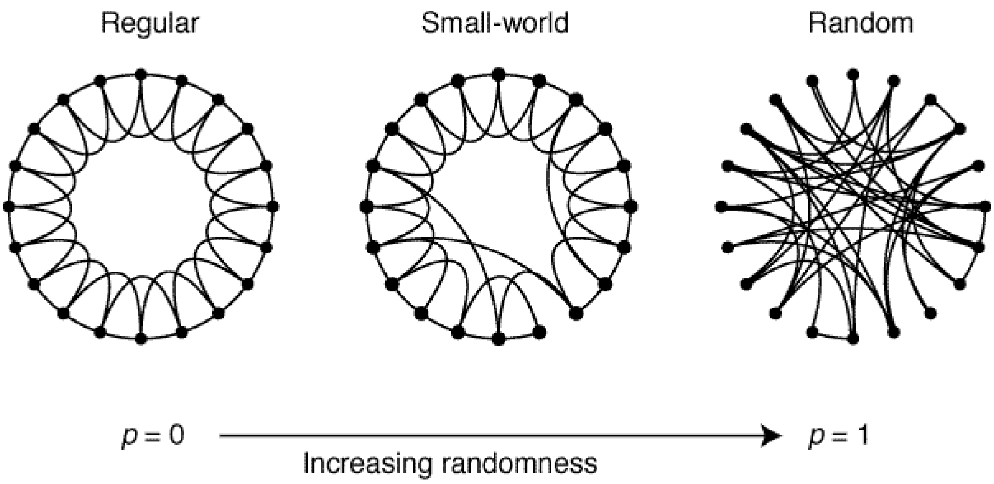


Fig. 2.7 Random rewiring procedure for interpolating between a regular ring lattice and a random network, without altering the number of nodes or edges in the graph. Source: [50].

2.2.3 Coreness

Coreness has been employed in previous literature as a metric of the resilience of a network [17]. The resilience of a social network is the ability of a social group to withstand external stresses. To measure coreness of the intra-network of each cluster the k-core decomposition is applied in order to evaluate the distributions of the nodes within each k-core.

Given a network, a sub-network H induced by the subset of nodes C is defined. H is a k-core of the network if and only if for every node in C: $deg_H(i) \geq k$, and H is the maximum sub-graph which fulfils this condition. The degree of the node i in the sub-graph H is denoted as $deg_H(i)$. A node has k-index equal to k if it belongs to the k-core but not to the $(k+1)$-core. In simple words, k-core decomposition starts with $k = 1$ and removes all nodes with degree equal to 1. The procedure is repeated iteratively until no nodes with degree 1 remain. Next, all removed nodes are assigned k-index to be 1. It continues with the same procedure for $k = 2$ and obtains nodes with indexes equal 2, and so on. The process stops when the last node from the network is removed at the ${k_{max}}^{th}$ step. The variable k_{max} is then the maximum shell index of the graph.

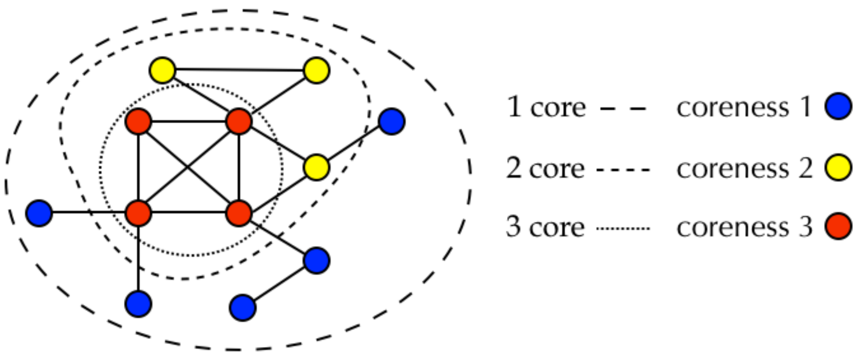


Fig. 2.8 K-core decomposition of a given graph.
Source: https://chaoslikehome.wordpress.com/.

Chapter 3

Evaluation

This chapter includes the description of the Twitter dataset, the evaluation of the computational framework, and the discussion of the results.

3.1 Data preparation

Data were collected from Twitter in relation to the campaign for the 2015 Barcelona City Council election (May 1-26, 2015) by the definition of a list of Twitter accounts of the seven main political parties:

- Barcelona en Comú (BeC)[1],
- Convergència i Unió (CiU)[2],
- Ciudadanos (Cs)[3],
- Capgirem Barcelona (CUP)[4],
- Esquerra Republicana de Catalunya (ERC) [5],

[1] https://en.wikipedia.org/wiki/Barcelona_en_Com%C3%BA
[2] https://en.wikipedia.org/wiki/Convergence_and_Union
[3] https://en.wikipedia.org/wiki/Citizens_(Spanish_political_party)
[4] https://en.wikipedia.org/wiki/Popular_Unity_Candidates
[5] https://en.wikipedia.org/wiki/Republican_Left_of_Catalonia

- Partit Popular de Catalunya (PP)[6],
- Partit dels Socialistes de Catalunya (PSC)[7].

The lists also includes the Twitter accounts for corresponding candidates for Mayor and each member party for the coalitions CiU, BeC and CUP. The users of the list can be found in Table 3.1. The Twitter Streaming API provided 373,818 retweets of tweets that (1) were created by, (2) were retweeted by, or (3) mentioned a user from the list. Figure 3.1 shows the distribution of the retweets over time, and reveals that the most active dates were the election day (March 24) and the one of the televised debate between candidates (March 21). In contrast, the day preceding the election, known as the reflection day, shows an notable decrease in Twitter activity. This distribution is similar to the one observed in previous studies about Spanish politics on this social network [2].

Table 3.1 Twitter accounts of the selected political parties and candidates.

Political Party	Party account(s)	Candidate account
BeC	@bcnencomu @icveuiabcn @podem_bcn @equobcn @pconstituentbcn	@adacolau
CiU	@cdcbarcelona @uniobcn	@xaviertrias
Cs	@cs_bcna	@carinamejias
CUP	@capgirembcn @cupbarcelona	@mjlecha
ERC	@ercbcn	@alfredbosch
PP	@ppbarcelona_	@albertofdezxbcn
PSC	@pscbarcelona	@jaumecollboni

[6] https://en.wikipedia.org/wiki/People%27s_Party_of_Catalonia
[7] https://en.wikipedia.org/wiki/Socialists%27_Party_of_Catalonia

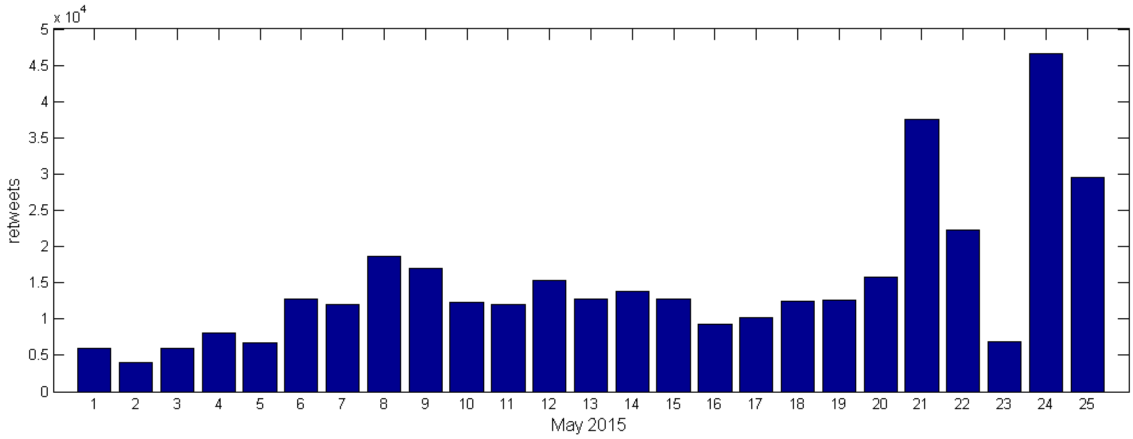


Fig. 3.1 Distribution of retweets over time.

It is important note that the sampling criteria are based on specific accounts instead of hashtags. Some studies have detected differences in the tagging practice of politicians [30]. Previous work has observed that some parties adopt a small set of hashtags during campaigns and some other parties generate new hashtags every day in order to locate them in the list of trending topics. Therefore, sampling messages from a list of campaign hashtags would likely lead to an unbalanced dataset. For this reason, these sampling criteria represent a better approach to capture the diffusion practices of the communities around parties. From the collection of retweets, a directed weighted graph is built which comprises a set of nodes (users) and a set of edges (retweets between any pair of users). Each edge in the graph represents that the source user retweeted a message posted by the the target user. To exclude anecdotal interactions between users which might not be enough of a signal to infer endorsement [18] and to highlight the structure of the expected clusters, the network only contains the interactions between any pair of nodes that occurred at least 3 times: an edge from user A to user B implies that user A has retweeted at least 3 times user B in the dataset. Nodes without edges after this process are removed. The resulting network comprises 6,492 nodes and 16,775 edges.

3.2 Results

This section describes the results of the computational framework for community detection and the characterization of the major clusters.

3.2.1 Community detection

A first execution of the standard Louvain method is done finding 151 clusters and achieving a remarkable value of modularity ($Q = 0.727$). Figure 3.2 shows a clear difference between the eight largest clusters (size $\in [232, 1981]$) and the remaining 143 clusters (size $\in [2, 62]$). In order to label these eight clusters, a manual inspection of the most relevant users from each cluster is done. Relevant users are defined by their PageRank value within the full network (the top five users for each cluster are listed in Table 3.2).

The results indicate that the standard Louvain method identifies a single cluster for almost each party: BeC = c_1, c_4; ERC = c_2; CUP = c_3; Cs = c_5; CiU = c_6; PP = c_7 and PSC = c_8. The only exception for such rule is that BeC is composed of two clusters. The manual inspection of the users from these two clusters reveals that cluster c_1 is formed by the official accounts of the party (e.g. @bcnencomu, @ahorapodemos), allied parties (e.g. @ahoramadrid), the candidate (@adacolau) and a large community of peripheral users. Cluster c_4 is composed of activists engaged in the digital communication for the campaign (e.g. @toret, @santidemajo, @galapita). That is to say that the most visible accounts from the media perspective belong to c_1 while c_4 is formed by party activists, many of whom are related to the 15M movement. For this reason, from now on, the analysis distinguishes these clusters as *BeC-p* and *BeC-m*: *party* and *movement*, respectively.

In this single execution of the standard Louvain method, accounts related to media appear in almost every political party cluster. As defined above, a new version of the Louvain method

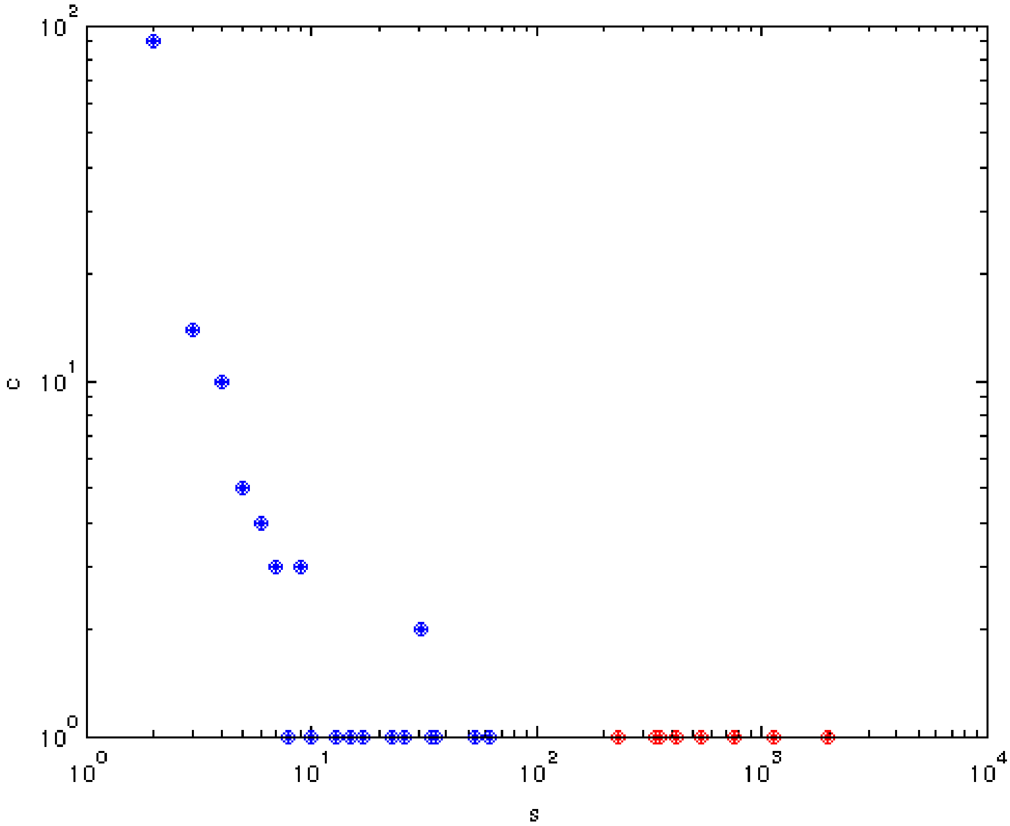


Fig. 3.2 Distribution of the number of clusters (c) by size (s). Red markers are used to indicate the eight largest clusters.

Table 3.2 Top 5 users for the eight largest clusters in the first execution of the standard Louvain method according to their PageRank (PR) value (clusters are ordered by size).

Id	Label	User	PageRank	Role
c_1	BeC-p	@bcnencomu	0.092	BeC party
c_1	BeC-p	@adacolau	0.029	BeC candidate
c_1	BeC-p	@ahoramadrid	0.009	BeC allied party
c_1	BeC-p	@ahorapodemos	0.009	BeC member party
c_1	BeC-p	@elperiodico	0.005	media
c_2	ERC	@ercbcn	0.016	ERC party
c_2	ERC	@alfredbosch	0.011	ERC candidate
c_2	ERC	@naciodigital	0.009	media
c_2	ERC	@arapolitica	0.007	media
c_2	ERC	@esquerra_erc	0.004	ERC party
c_3	CUP	@cupbarcelona	0.016	CUP party
c_3	CUP	@capgirembcn	0.008	CUP party
c_3	CUP	@albertmartnez	0.005	media
c_3	CUP	@encampanya	0.003	media
c_3	CUP	@mjlecha	0.002	CUP candidate
c_4	BeC-m	@toret	0.014	BeC member
c_4	BeC-m	@santidemajo	0.005	BeC member
c_4	BeC-m	@sentitcritic	0.005	media
c_4	BeC-m	@galapita	0.005	BeC member
c_4	BeC-m	@eloibadia	0.005	BeC member
c_5	Cs	@carinamejias	0.007	Cs candidate
c_5	Cs	@cs_bcna	0.006	Cs party
c_5	Cs	@ciudadanoscs	0.004	Cs party
c_5	Cs	@soniasi02	0.003	Cs member
c_5	Cs	@prensacs	0.002	media
c_6	CiU	@xaviertrias	0.012	CiU candidate
c_6	CiU	@ciu	0.004	CiU party
c_6	CiU	@bcn_ajuntament	0.003	institutional
c_6	CiU	@ramontremosa	0.002	CiU member
c_6	CiU	@cdcbarcelona	0.002	CiU member party
c_7	PP	@btvnoticies	0.011	media
c_7	PP	@cati_bcn	0.003	media
c_7	PP	@albertofdezxbcn	0.003	PP candidate
c_7	PP	@maticatradio	0.002	media
c_7	PP	@ppbarcelona_	0.002	PP party
c_8	PSC	@elsmatins	0.006	media
c_8	PSC	@pscbarcelona	0.003	PSC party
c_8	PSC	@sergifor	0.003	media
c_8	PSC	@jaumecollboni	0.002	PSC candidate
c_8	PSC	@elpaiscat	0.002	media

is designed to study the ecosystem of each political party, i.e including only nodes that are reliably assigned to them. The algorithm is applied by running the Louvain method 100 times and assigning to each cluster only the nodes that fall into that cluster more than 95 times ($N = 100, 1-\varepsilon = 0.95$). By inspecting the results of the 100 executions, a constant presence of eight major clusters, much bigger than the other clusters, is observed. The composition of these clusters is also quite stable: 4,973 nodes (82.25%) are assigned to the same cluster in over 95 executions.

The boundaries between ideological online communities are visible in Figure 3.3. For a better readability of the network, the network visualization only considers the giant component of the graph and applies the Force Atlas 2 layout [25] to enforce cluster graph drawing. As one could expect in any polarized scenario, the largest number of interaction links occur within the same cluster. There exists, however, a notably large number of links between the two clusters of BeC (BeC-p and BeC-m). Figure 3.4 presents the sub-network formed by the nodes and links of both clusters. The network visualization also includes the node labels of the most relevant nodes of the BeC party cluster from Table 3.2: @bcnencomu, @adacolau, @ahoramadrid, @ahorapodemos (the media account @elperiodico was excluded of the cluster by the new version of the Louvain method because it did not fall to the cluster more than 95 times). To further prove the low levels of interactions between major parties an interaction matrix A is defined, where $A_{i,j}$ counts all retweets that accounts from cluster i made for the tweets from users of cluster j. Since the clusters are of the different size, $A_{i,j}$ is normalized by the sum of the all retweets made by the users assigned to cluster i. Figure 3.5 shows matrix A for all the clusters and confirms that a vast majority of retweets were made between users from the same cluster (main diagonal). This is also true in the case of the two clusters of Barcelona en Comú although there is a presence of communication between movement and party clusters, with a prevalence from the movement to the party (BeC-m $\rightarrow$ BeC-p $= 0.18$, the largest value out of the main diagonal).

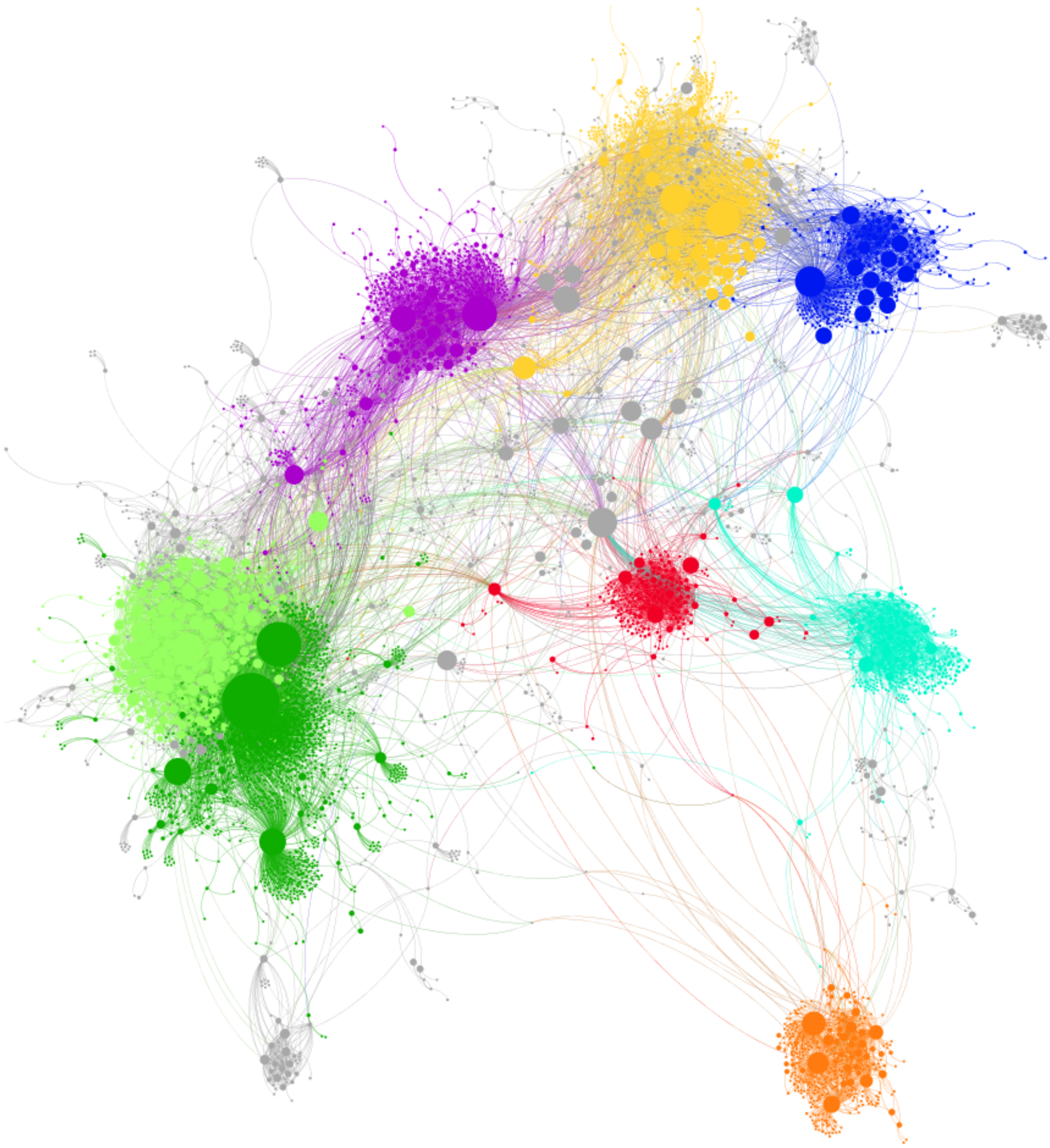


Fig. 3.3 Network of retweets (giant component). Clusters are represented by color: BeC-p (dark green); BeC-m (light green); ERC (yellow); PSC (red); CUP (violet); Cs (orange); CiU (dark blue); PP (cyan). The nodes out of these clusters are gray-colored.

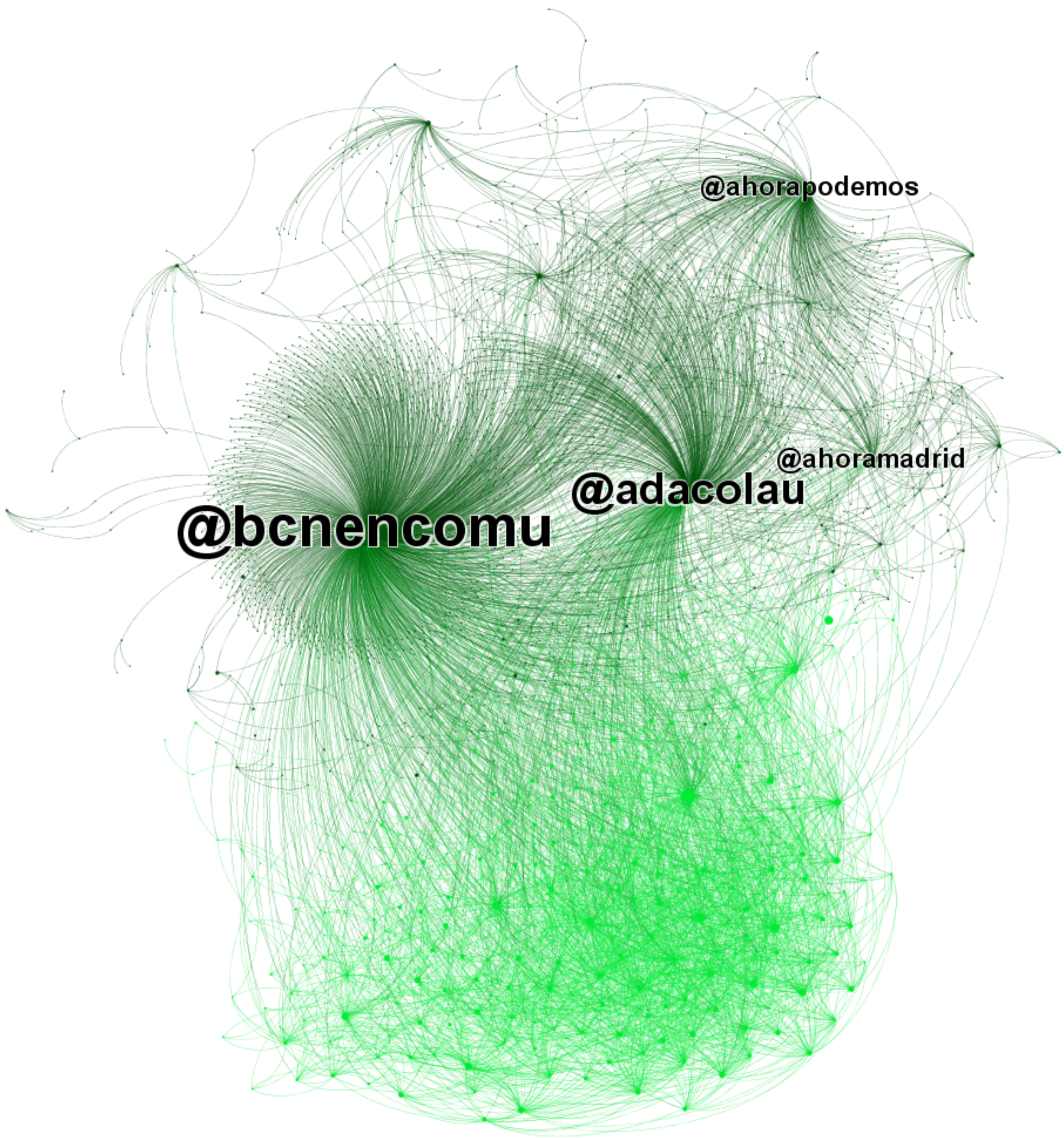


Fig. 3.4 Sub-network of retweets of BeC-p (dark green) and BeC-m (light green). The labels of the most relevant nodes of Bec-p from Table 3.2 are showed (the media account @elperiodico was excluded of the cluster by the new version of the Louvain method because it did not fall to the cluster more than 95 times).

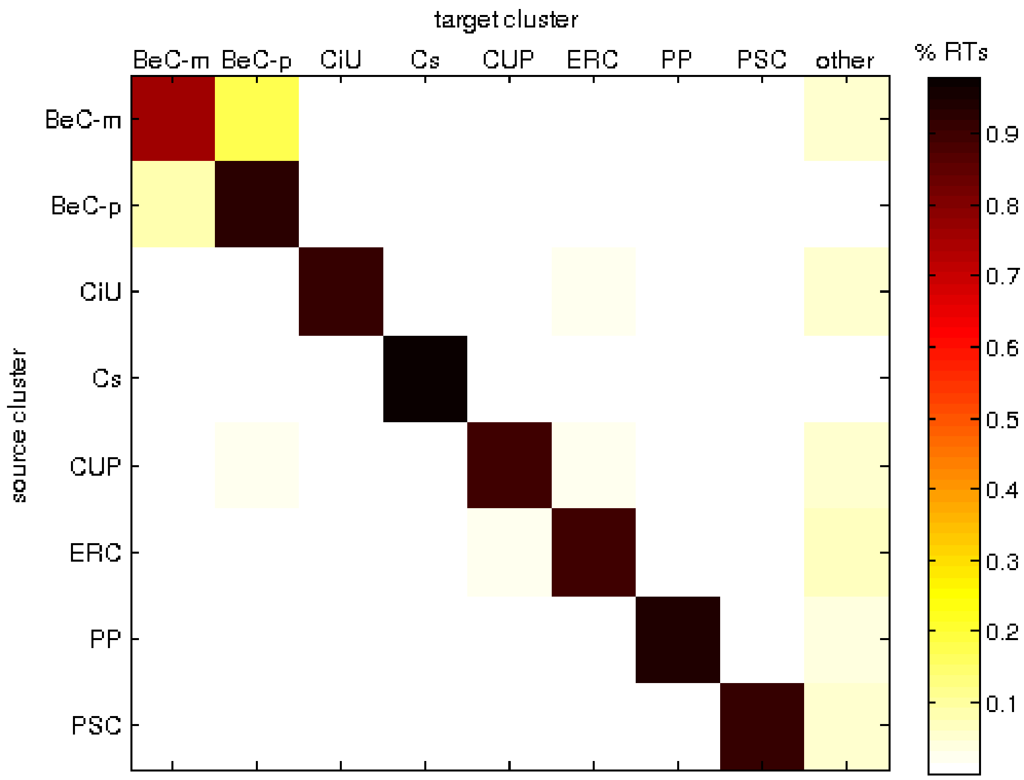


Fig. 3.5 Normalized weighted adjacency matrix of the network of clusters.

As mentioned above, the new version of the Louvain method proposed in this thesis only assigned a node to one of the eight largest clusters if it belongs to a particular cluster more than 95 of 100 times. The final inclusion/exclusion of the most relevant nodes to a cluster was manually inspected in order to assess the performance of this new version. For preserving the political affinity of users, Table 3.3 only presents the 20 most relevant nodes which were not assigned to any cluster, their role, and how many times they belong to each cluster over the 100 executions. The results prove that the suggested confidence interval of the new version for community detection effectively prevented the inclusion of media accounts in the intra-network of political parties. The list also includes two citizens, a politician from CIU with some connection to ERC, the political grassroots party Terrassa en Comú (allied with BeC), and two citizen organizations: Plataforma de Afectados por la Hipoteca (co-founded by the candidate of BeC) and Vaga de Totes.

Table 3.3 Most relevant nodes which could not be reliably assigned to any of the major clusters indicating the number of occurrences in each cluster.

User	Role	bec-m	bec-p	ciu	cs	cup	erc	pp	psc	other
@btvnoticies	media	0	0	0	0	1	0	86	13	0
@elperiodico	media	0	90	0	3	0	1	0	1	5
@elsmatins	media	0	0	0	0	0	93	0	7	0
@naciodigital	media	0	0	1	0	38	61	0	0	0
@tv3cat	media	0	0	0	0	3	54	0	19	24
@encampanya	media	1	0	0	0	36	0	0	0	63
@rocsalafaixa	citizen	0	0	7	0	1	92	0	0	0
@bernatff	media	0	0	1	0	38	61	0	0	0
@jordi_palmer	media	0	0	1	0	38	61	0	0	0
@mariamariekke	citizen	50	50	0	0	0	0	0	0	0
@puntcattv3	media	0	0	0	0	0	44	0	56	0
@ramontremosa	politician	0	0	90	0	0	10	0	0	0
@santimdx5	media	91	0	0	0	0	0	0	0	9
@mtudela	media	0	0	7	0	1	92	0	0	0
@pah_bcn	citizen organization	89	0	0	0	0	0	0	0	11
@324cat	media	0	0	0	0	3	52	0	13	32
@terrassaencomu	political party	2	92	0	0	0	0	0	0	6
@sicomtelevision	media	1	8	0	0	90	0	0	0	1
@xriusenoticies	media	0	0	35	0	0	0	0	65	0
@vagadetotes	citizen organization	78	0	0	0	22	0	0	0	0

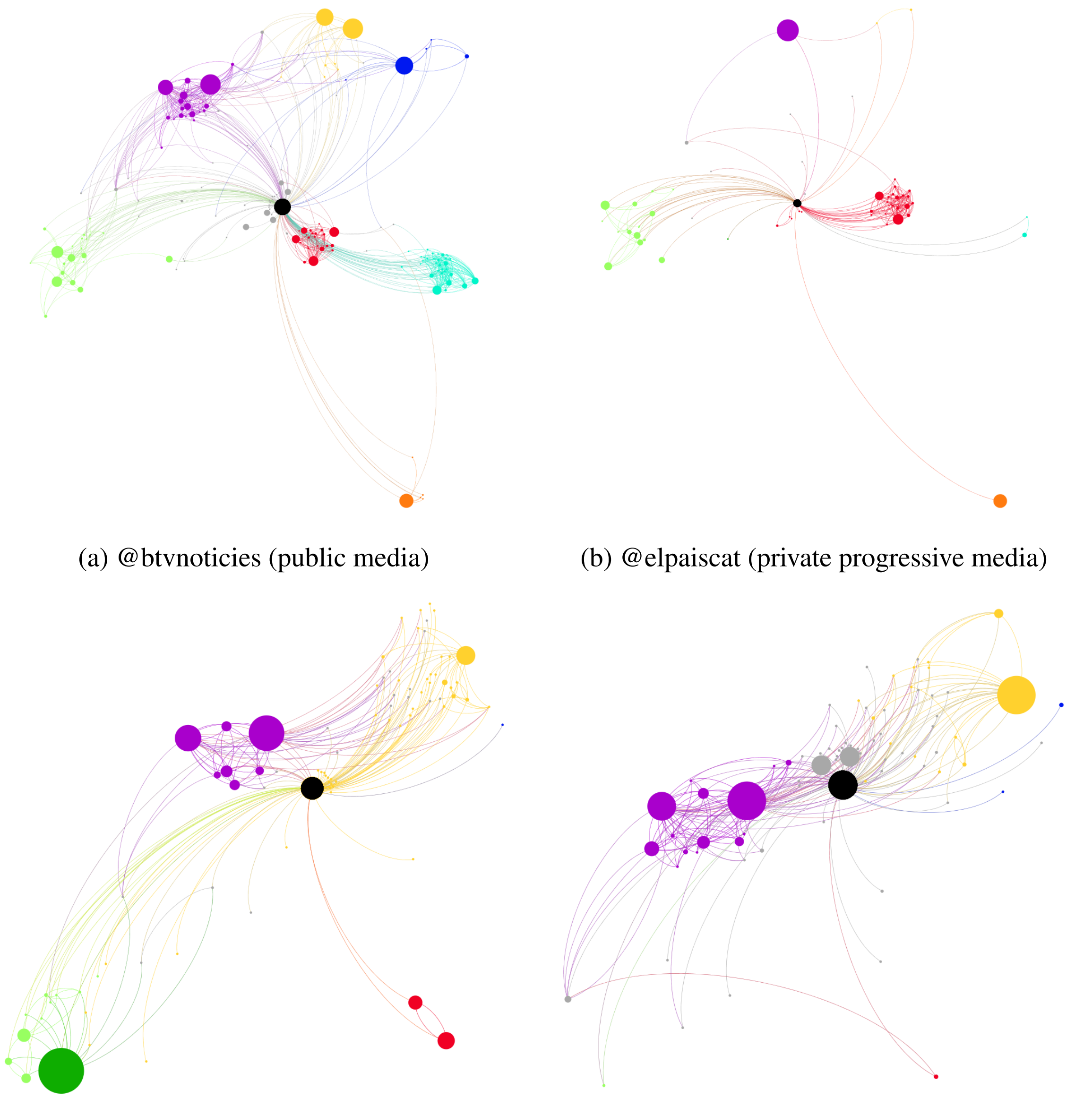
Network of weak ties

Among the nodes which could not be reliably assigned to any of the major clusters, many accounts correspond to traditional mass media outlets. To analyze this finding in more detail, an execution of the standard Louvain method is taken to identify the most relevant users, according to PageRank, in the sub-network formed only by edges between nodes from different clusters, i.e. "weak ties" [23]. Table 3.4 presents the 25 most relevant users in this sub-network and confirms that media played a key role in connecting different clusters. Since media accounts rarely retweet content from other accounts, a great amount of weak ties consists of users from political party clusters retweeting content published by media accounts.

To deepen how media built bridges among clusters, the ego-networks of four of the most relevant media accounts within the retweet networks are analyzed. Figure 3.6a corresponds to the ego-network of @btvnoticies, the local and publicly owned television channel, that is retweeted by users from every cluster. This behaviour might be explained by the fact that this TV channel organized the debate among most of the candidates. In contrast, the other three accounts are private media: @elpaiscat, @arapolitica, and @naciodigital. Figure 3.6b clearly shows that tweets from @elpaiscat, progressive media, are mostly diffused by users from BeC and PSC, progressive parties. On the other hand, Figures 3.6c and 3.6d reveal that @arapolitica and @naciodigital, Catalan nationalist media, are mainly retweeted by users from the pro-independence Catalan parties CUP and ERC.

Table 3.4 Most relevant nodes by PageRank in the sub-network formed by edges between nodes from different clusters.

User	Page Rank	Role
@btvnoticies	0.014	media
@bcnencomu	0.012	party
@sicomtelevision	0.010	media
@cupbarcelona	0.007	party
@elsmatins	0.007	media
@capgirembcn	0.006	party
@tv3cat	0.006	media
@324cat	0.006	media
@xaviertrias	0.005	candidate
@puntcattv3	0.005	media
@revolucio1984	0.004	citizen
@sergifor	0.004	media
@nuriapujadas	0.004	media
@annatorrasfont	0.004	media
@arapolitica	0.004	media
@maticatradio	0.003	media
@cati_bcn	0.003	media
@elpaiscat	0.003	media
@encampanya	0.003	media
@albertmartnez	0.002	media
@naciodigital	0.002	media
@adacolau	0.002	candidate
@ramontremosa	0.002	party member
@alfredbosch	0.002	candidate
@directe	0.001	media



(a) @btvnoticies (public media)

(b) @elpaiscat (private progressive media)

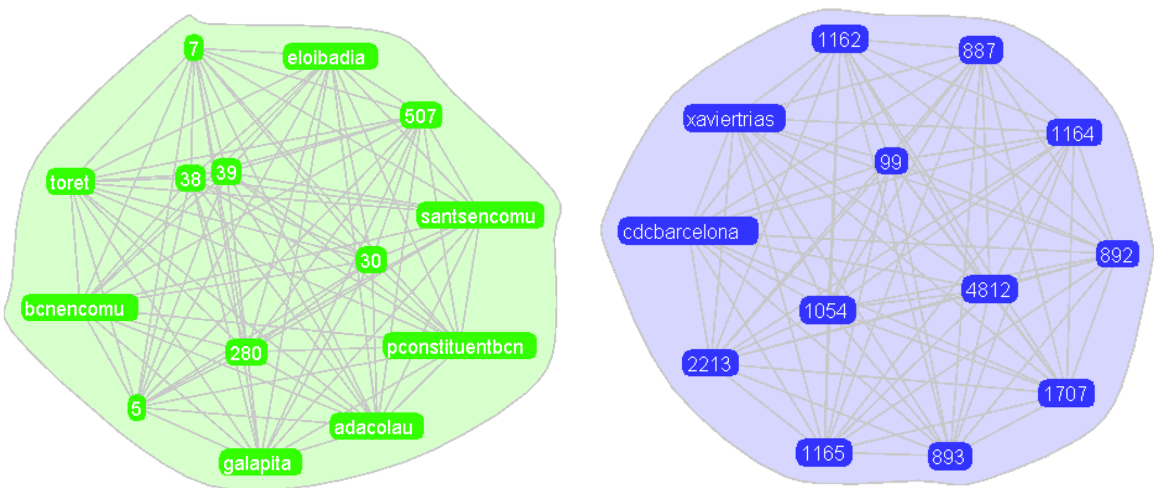
(c) @arapolitica (private Catalan nationalist media)

(d) @naciodigital (private Catalan nationalist media)

Fig. 3.6 Ego-networks of 4 media accounts within the network of retweets. Central nodes (i.e. corresponding media accounts) are black-colored. Clusters are represented by color: BeC-p (dark green); BeC-m (light green); ERC (yellow); PSC (red); CUP (violet); Cs (orange); CiU (dark blue); PP (cyan). The nodes outside of these clusters are gray-colored.

Comparison to the Clique Percolation Method

The new version of the Louvain is motivated by the fuzzy community structure of political networks, as the one of the campaign for the 2015 Barcelona City Council election. These networks are usually formed by overlapping communities and the proposed algorithm improves the standard Louvain method by identifying clusters in a more stable way. The definition of the computational framework in the previous chapter noted that there are some community detection methods in the state of the art for overlapping communities. In particular, the Clique Percolation Method (CPM) is the most popular one according to [14]. This method is applied on the network by using the CFinder software package[8]. Figure 3.8 presents the number of clique graphs obtained through the CPM at every value of k. As expected, the number of clique graphs tends to decrease as k increases. At its maximum value ($k = 13$), the method only detects two clique graphs: one formed by users from BeC and another formed by users from CiU (see Figure 3.7).



(a) BeC. (b) CiU.

Fig. 3.7 Clique graphs obtained through the CPM at its maximum possible value ($k = 13$). Accounts of non-public citizens are anonymized by showing a numerical ID.

[8] Available at: http://www.cfinder.org/.

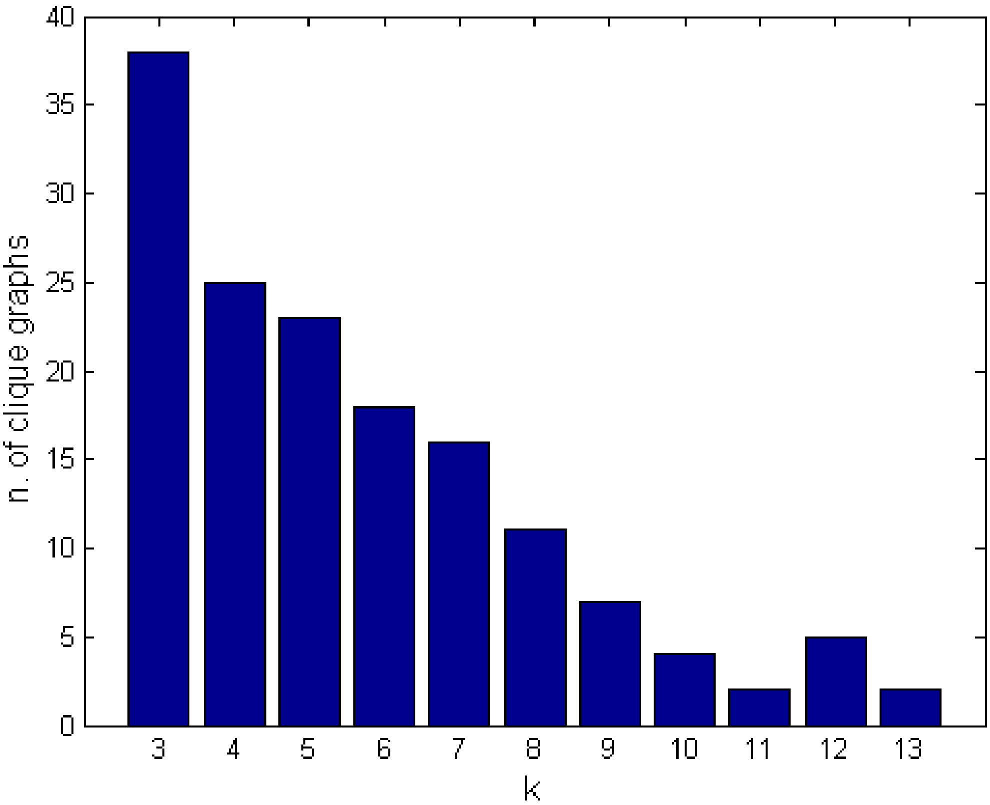
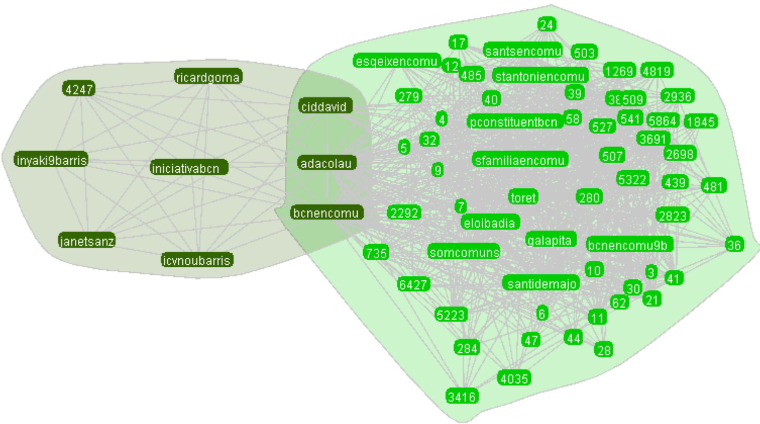


Fig. 3.8 Number of clique graphs obtained through the Clique Percolation Method at every value of k.

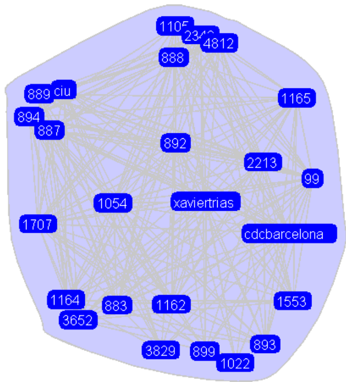
While the Louvain method was able to identify every party cluster, CPM at its maximum value only detects two party clusters. This is explained by the different size and structure of the party networks. For this reason, the communities at different values of k have been examined. When $k = 9$, CPM identifies seven clique graphs. The inspection of the nodes of each of them reveals that two of them are related to BeC, one is related to a municipal police trade union and the rest are related to each of the political parties CiU, CUP, Cs and PP. For PSC and ERC, CPM identifies clique graphs when $k = 8$ and $k = 7$ respectively. To compare these results with the clusters from the Louvain method with confidence interval, Table 3.5 indicates how many nodes of the each clique graph occurred in each cluster, and reveals that:

- All the nodes of the clique graphs related to CiU, Cs, CUP, ERC and PSC are part of the corresponding clusters from the Louvain method with a confidence interval.
- Only one node from PP clique graph was not in PP political cluster.
- The nodes from the clique graph related to a trade union of municipal police (GU) were not in a political cluster.
- The largest BeC clique graph (BeC_1) is mainly formed by nodes from the BeC movement cluster. The smallest clique graph (BeC_2) is composed by two nodes from the BeC party cluster and seven nodes from the BeC movement cluster.

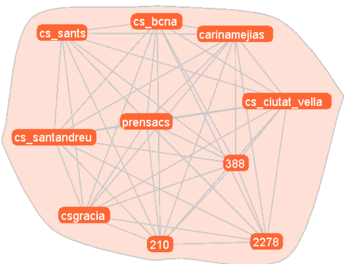
Figure 3.9 presents all these clique graphs to better understand their composition. The figure shows an overlap between the two BeC clique graphs which is composed of three nodes: @bcnencomu (party account), @adacolau (candidate) and @ciddavid (party member). It is interesting to observe that, although the rest of nodes of the smallest clique graph belongs to the movement cluster, all of them are related to Iniciativa per Catalunya Verds, the main pre-existing party that converged in Barcelona en Comú. In other works, CPM also identifies a clique graph related to the institutional elite of BeC and a much larger clique graph related to the grassroots elements of BeC.



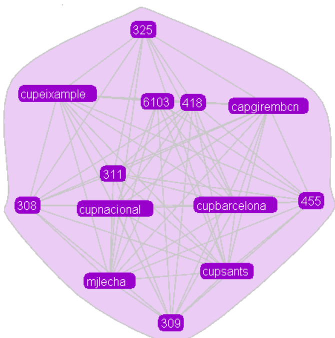
(a) BeC-p (left) & BeC-p (right) (k=9).



(b) CiU ($k = 9$).

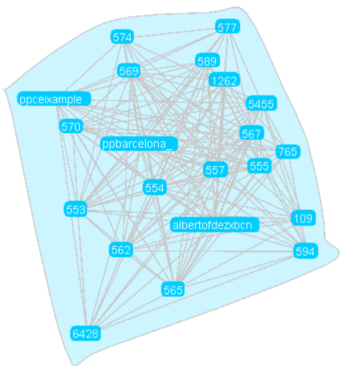


(c) Cs ($k = 9$).

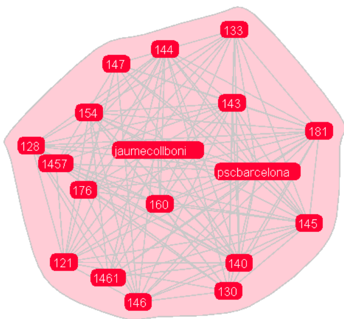


(d) CUP ($k = 9$).

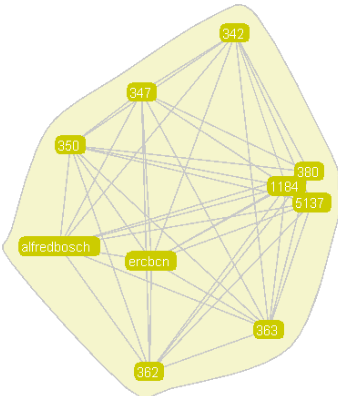
(e) Guardia Urbana ($k = 9$).



(f) PP (k=9).



(g) PSC ($k = 8$).



(h) ERC ($k = 7$).

Fig. 3.9 Clique graphs obtained through the Clique Percolation Method. The seven first graphs are the ones when k equals to 9. The two last graphs are the largest clique graphs for PSC ($k = 8$), and ERC ($k = 7$). Accounts of non-public citizens are anonymized by showing a numerical ID.

Table 3.5 Clusters obtained through Clique Percolation Method, k value of k-clique graph, and number of nodes which occur in the clusters obtained through the Louvain method with confidence interval ($LMCI_x$). The largest number of each row is bold.

CPM	k	$LMCI_{bec-m}$	$LMCI_{bec-p}$	$LMCI_{ciu}$	$LMCI_{cs}$	$LMCI_{cup}$	$LMCI_{erc}$	$LMCI_{pp}$	$LMCI_{psc}$	$LMCI_{other}$
bec_1	9	**60**	3	0	0	0	0	0	0	2
bec_2	9	**7**	2	0	0	0	0	0	0	0
ciu	9	0	0	**25**	0	0	0	0	0	0
cs	9	0	0	0	**10**	0	0	0	0	0
cup	9	0	0	0	0	**13**	0	0	0	0
erc	7	0	0	0	0	0	**7**	0	0	0
pp	9	0	0	0	0	0	0	**20**	0	1
psc	8	0	0	0	0	0	0	0	**18**	0
gu	9	0	0	0	0	0	0	0	0	**11**

In conclusion, the results from applying CPM are consistent with the ones obtained through the community detection algorithm proposed in this thesis. However the suggested method have substantial advantages over CPM:

- Given a graph composed of n nodes and m edges, the computational complexity of CPM is $O(exp(n))$, since finding a maximum clique in a graph is a NP-complete problem. In contrast, the computational complexity of the suggested method by running N executions of the standard Louvain method is $O(Nm)$.
- The different size and structure of the political networks make that CPM at the maximum value of k only detects two major cluster. On the other hand, the proposed method is able to identify every party cluster.
- The clusters obtained through CPM are k-cliques and, therefore, such clusters are dense graphs formed by the core of the party network structure. Social networks are characterized by their heavy-tailed degree distribution so the clique graphs exclude the large amount of less active users. Recent studies have proved that these are the nodes which compose the critical periphery in the growth of protest movements [3]. For this reason, the inclusion of these nodes, as done in the proposed method, becomes essential for the following characterization of clusters.

3.2.2 Cluster characterization

The eight clusters detected by the community detection algorithm are then characterized in terms of hierarchical structure, small-world phenomenon and coreness.

Hierarchical structure

To evaluate the hierarchical structure, the in-degree inequality of each cluster is measured by calculating Gini coefficient. In-degree centralization, originally suggested in [16], is also computed.

From results in Table 3.6 a notable divergence between both metrics is seen: the inequality values of CiU and PP are similar ($G_{in} = 0.893$ and $G_{in} = 0.876$, respectively), but PP centralization ($C_{in} = 0.378$) is far from the maximum centralization value exhibited by CiU ($C_{in} = 0.770$). For Barcelona en Comú, BeC-m emerges as the least inequal and the least centralized structure, while BeC-p forms the most inequal cluster ($G_{in} = 0.995$). The Lorenz curve of the in-degree distribution of the clusters is presented in Figure 3.10 to visually validate the different levels of inequality among clusters that are presented in Table 3.6.

Table 3.6 Inequality based on the Gini Coefficient (G_{in}) and centralization (C_{in}) of the in-degree distribution of each cluster, and ratio between the maximum in-degree and the number of nodes (r).

Cluster	G_{in}	C_{in}	r
BeC-p	0.995	0.639	0.639
Cs	0.964	0.476	0.480
ERC	0.954	0.452	0.454
CUP	0.953	0.635	0.636
CiU	0.893	0.770	0.774
PP	0.876	0.378	0.389
PSC	0.818	0.565	0.578
BeC-m	0.811	0.290	0.302

As showed in Table 3.6, it is easy to demonstrate that for networks with a heavy tailed in-degree distribution the in-degree centralization formulated in [15] is approximately equal to the ratio between the maximum in-degree and the number of nodes:

$$C_{in} = \frac{\sum_{i=1}^{n}[k_{\max}^{in} - k_i^{in}]}{\max \sum_{i=1}^{n}[k_{\max}^{in} - k_i^{in}]} \approx \frac{(n-1)\cdot k_{\max}^{in}}{(n-1)\cdot(n-1)} \approx \frac{k_{\max}^{in}}{n}$$

This is caused by the differences of several orders of magnitude between the maximum and average in-degree, common situation for social graphs. Therefore, this metric is not a good one to capture hierarchical structure for social diffusion graphs, and Gini coefficient for in-degree inequality represents a more reliable measure.

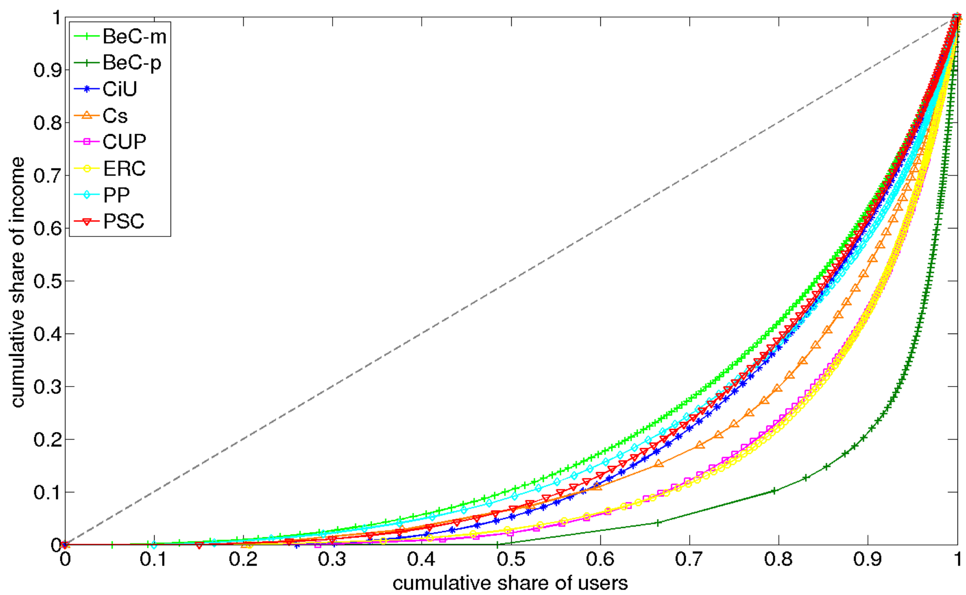


Fig. 3.10 Lorenz curve of the in-degree distribution of each cluster.

Small-world phenomenon

Broadly speaking, the efficiency of a social network is explained by its small-world phenomenon, i.e. phenomenon of users being linked by a mutual acquaintance. To assess the small-world phenomenon in each party, the average path length and the clustering coefficient are computed.

Table 3.7 reveals that BeC-m has the highest clustering coefficient ($Cl = 0.208$) closely followed by PP and PSC, the two smallest clusters by size. On the contrary the clustering coefficient of BeC-p is almost 0. This finding is explained by the topology of BeC-p, roughly formed by stars whose center nodes are the most visible Twitter accounts of Barcelona en Comú: the party accounts and the candidate.

No remarkable patterns regarding the average path length are observed. It is lower than 3 for the majority of the party clusters with the PSC cluster having the lowest value ($l = 2.29$). In the same time ERC, CiU and BeC-m expose the longest average path length (5.43, 4.66, 3.35 respectively), that might signal the lower information especially in the case of ERC.

Table 3.7 Number of nodes (N) and edges (E), clustering coefficient (Cl) and average path length (l) of the intra-network of each cluster.

Cluster	N	E	Cl	l
BeC-m	427	2 431	0.208	3.35
PP	301	1 163	0.188	2.73
PSC	211	810	0.182	2.29
CiU	337	1 003	0.114	4.66
Cs	352	832	0.073	2.57
CUP	635	1 422	0.037	2.57
ERC	866	1 899	0.027	5.43
BeC-p	1 844	2 427	0.002	2.48

Coreness

The coreness of a network is closely related to its social resilience, i.e. the ability of a social group to withstand external stresses. To measure social resilience for a social network, the k-core decomposition of each cluster is performed in order to evaluate the distributions of the nodes within each k-core. The more nodes are in the most inner cores, i.e. the ones with the larger k-indexes, and the larger is the maximal k-index, then the more resilient the cluster is.

Table 3.8 presents the maximal and average k-indexes for each cluster and Figure 3.11 visually shows the corresponding distributions. As in the case of hierarchical structure and small-world phenomenon, BeC-m ($k_{max} = 17$, $k_{avg} = 5.90$) and BeC-p ($k_{max} = 5$, $k_{avg} = 1.33$) are the highest and lowest values respectively. In comparison to the other parties there are clear differences between node distributions for both, BeC-m and BeC-p, and the rest (the largest concentration of the nodes is in the first k-cores and considerable part is in the inner most cores). Therefore, the movement group of Barcelona en Comú is an online social community with an extreme ability to withstand or recover. In the same time the party group of Barcelona en Comú seems to only focus on the core users.

Table 3.8 Maximal and average k-index (standard deviation in parentheses) for the intra-network of each cluster.

cluster	k_{max}	k_{avg}
BeC-m	17	5.90 (5.46)
PP	12	4.02 (3.99)
PSC	11	3.85 (3.55)
CiU	13	3.10 (3.44)
ERC	8	2.25 (1.85)
Cs	10	2.42 (2.42)
CUP	10	2.19 (2.22)
BeC-p	5	1.33 (0.71)

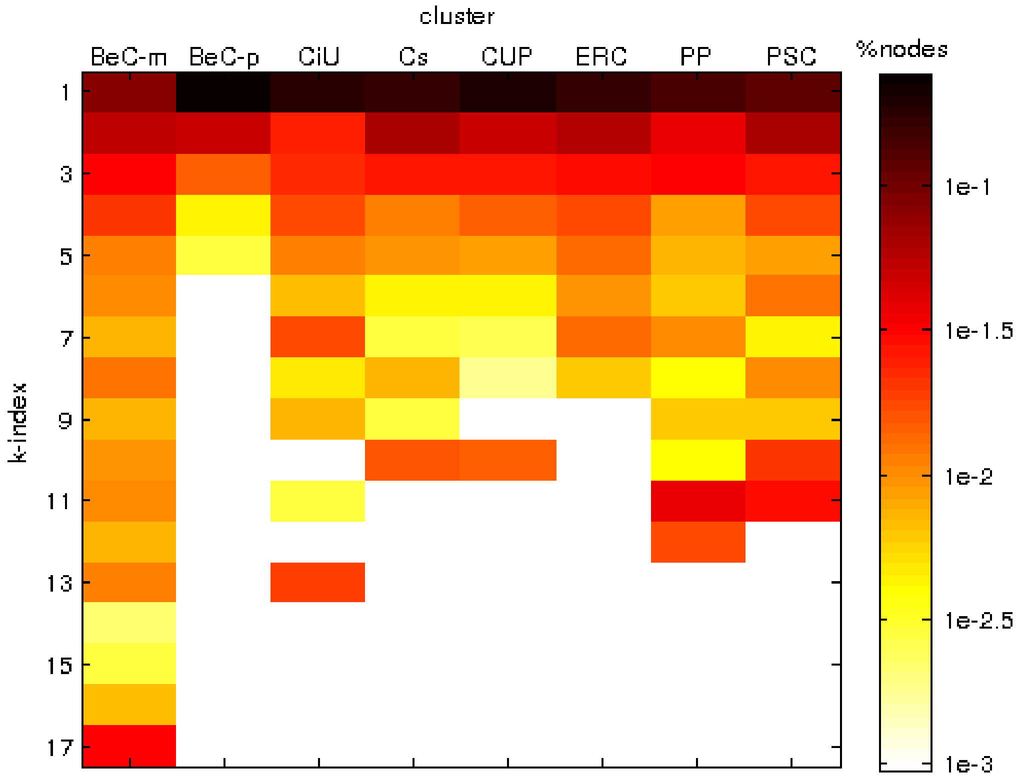


Fig. 3.11 Distribution of the nodes per cluster (column) and k-index (row). Cells are colored to form a heat map indicating the density (log scale).

3.3 Discussion

This section presents the discussion of the results of the computational framework on the Twitter party networks for the 2015 Barcelona City Council election.

3.3.1 Institutionalization of movement

The computational framework has been designed to provide an answer to the research question that deals with the kind of organizational structure that Barcelona en Comú developed for the election campaign. On the one hand, the cited literature [22, 49] provided evidence of the decentralization of the 15M movement, which inspired the Barcelona en Comú candidacy. On the other hand, many political scientists [33, 42, 36, 34] argued that parties are historically ruled by elites and, therefore, result in centralized organizations. Furthermore, the historical models of political parties reviewed in [27] (i.e. *Caucus parties*, *Mass parties*, *Catch-all parties*, and *Cartel parties*) always assumed organization around elites. All of these observations motivated to study whether Barcelona en Comú preserved a decentralized structure or adopted a conventional centralized organization.

The results depict a movement-party structure in which the two components form well-defined clusters. In comparison to the clusters of the rest of political parties, BeC movement community emerges as the least hierarchical, best clustered and most resilient one. In contrast, the BeC party community is the most hierarchical, least clustered and least resilient one. The centralization of the party cluster points to the candidate and official accounts, the subjects that are commonly associated with the elite. However, unlike the rest of political parties, there is a co-existence of both party and movement clusters. This co-existence is consistent with the hypothesis expressed in [48] when defining Podemos, member party of Barcelona en Comú, as the conjugation of a front-end and a back-end.

This thesis has provided hints about the characterization of the organization of political parties according to their online diffusion networks. Some authors have reported that the Internet played a key role in the organization of the 15M movement for building “a hybrid space between the Internet social networks and the occupied urban space” [9]. According to [49], this hybrid space is the result of *techno-political* practices: “the tactical and strategic use of technological devices (including social networks) for organization, communication and collective action”. Are techno-politics the origin of this particular movement-party partition of Barcelona en Comú? Recently, political scientists have postulated the emergence of *cyber parties* “with its origins in developments in media and information and communication technologies” [32]. Although the results of this stuy cannot ensure that the Internet and social media are the only reason behind this new form of political organization, in this particular context some party activists reported that ICT becomes essential for campaigning [45]. Therefore, a close link between techno-politics and the structure of Barcelona en Comú might exist.

3.3.2 Polarization in social media

The identification of the different clusters was made possible by the high level of polarization that the network exhibited, as initially expected. The observed bridges between clusters (i.e. “weak ties" [23]) were mostly built by accounts related to media. As noted above, media do not retweet messages from other accounts, therefore, most weak ties correspond to tweets from media accounts that were retweeted by users from political party clusters. This means that media play a key role in generating messages that build a public sphere. Some theorists suggest that the best response to group polarization is the usage of “mechanisms providing a public sphere” [46]. The most relevant account in the sub-network of weak ties was @btvnoticies, the local and publicly owned television. Indeed, according to its ego-network, @btvnoticies was retweeted from every party while the other three private media were mostly retweeted by users from like-minded parties. This finding might indicate that public television became more plural than the other three analyzed private media, and pluralism is an effective tool in order to get "people exposed to a range of reasonable competing views" [46].

It is important to note that some steps of the data preparation process may have accentuated the polarization effect: (1) tweet collection criteria were focused on parties/candidates, (2) the graph only comprises edges of interactions that occurred at least 3 times and (3) the community detection algorithm was modified to enforce the robustness of the clusters. Moreover, retweeting has been proven as a common mechanism for endorsement [7] and Twitter itself presents considerable levels of homophily [28]. Therefore, it might be interesting to discuss whether polarization based on retweets is an effect of the own characteristics of this particular microblogging service. An analysis of the 2011 Spanish election on Twitter revealed that polarization measured in retweet networks was more intense than polarization measured in reply networks [2]. This indicates that diffusion/support networks exhibit segregation to a greater extent than discussion networks. Similar findings are also reported in a study of a Swiss political online platform which concluded that "interactions with positive connotation (supports and likes) revealed significant patterns of polarization with respect to party alignment, unlike the comments layer, which has negligible polarization" [16]. Another analysis conducted on Wikipedia, online platform in which users have to collaborate, discuss and reach agreement on editing articles, did not find a strong preference to interact with members of the same political party [37]. In conclusion, polarization in diffusion/support networks (e.g. microblogging) does not imply a segregated society.

Chapter 4

Conclusions

This thesis has proposed a computational framework to examine new forms of political organization in social media. The results focus on the Twitter networks of Barcelona en Comú in comparison to the other parties for the 2015 Barcelona municipal elections. The findings rely on a dataset from Twitter but social networks are only a slice of the structure of political organizations and not every party activist has a Twitter account. Furthermore, some experts are sceptical with the digital forms of activism because of the "loss of coherence, morality or even sustainability" [35] and pointed out the rise of a low commitment and feel-good form of activism. Nevertheless, online platforms are playing a key role in political discussion and campaigning, and social media data are leveraging the capacity of revealing patterns of individual and group behaviours [29, 21]. Recent studies about the communication dynamics in social media for collective action have demonstrated that "relatively low commitment participants are potentially very important as a collective" [3] and, therefore, Twitter might be seen as an informative and valuable data source to examine collective behaviour and self-organization in social and political contexts.

The computational framework defined in this thesis focuses on (1) community detection and (2) cluster characterization. The fuzzy membership of some nodes in certain communities (e.g. media accounts) motivated the modification of a standard community detection algorithm (Louvain method) by setting a confidence interval to enhance the robustness of the clusters. In comparison to the standard Louvain method and another community detection algorithm for overlapping communities (Clique Percolation Method), the evaluation proved that the new algorithm identified the political networks in a more stable way. Cluster characterization was inspired by the metrics proposed in [16] to compare political party networks. The original dimensions of this framework were hierarchical structure, information efficiency, and social resilience. The redefinition of these three dimensions and the inclusion of new metrics in the computational framework constitute an improvement of the characterization of political networks:

- *Hierarchical structure*. In-degree centralization [15] was originally applied in [16] to measure the hierarchical structure of a network. This metric is based on (1) how the centrality of the most central node exceeds the centrality of all other nodes and (2) the comparison to a star network. Maximum and average in-degree have commonly differences of several orders of magnitude in social graphs. Therefore, in-degree centralization is approximately equal to the ratio between the maximum in-degree and the number of nodes for social networks with a heavy tailed in-degree distribution. In other words, the in-degree centralization is not a good metric to capture hierarchical structure for social diffusion graphs, and the Gini coefficient for in-degree inequality represents a more reliable measure of the hierarchical structure of a network.

- *Information efficiency*. Information efficiency in social networks is closely related to the *small-world phenomenon*. This thesis proposes the average path length, as the previous framework does [16], and the clustering coefficient to better characterize efficiency in social networks.

- *Social Resilience*. Previous studies indicated the suitability of the *k*-core decomposition to measure the resilience of social networks [17]. This framework recommends the term *coreness* which represents a more precise definition of this metric. In addition, showing the distribution of nodes along *k*-cores does capture resilience better than maximum *k*-core as done in [16].

4.1 Future work

The results showed that the tension between the decentralization of networked movements and the centralization of political parties results into a movement-party structure: both paradigms co-exist in two well-defined clusters. From this result, future work should investigate the origin of this particular structure:

- Did the online structure of Barcelona en Comú result from the confluence of minor parties and the 15M activists?
- Instead of evolving into a centralized organization, did the 15M networked movement implement a party interface over its decentralized structure?

The computational framework could be refined by adding longitudinal analyses of the formation of the clusters in order to provide answer to these open questions.

Finally, it is interesting to note that city council elections were held in every Spanish city in May 2015 and candidacies similar to Barcelona en Comú were built. Indeed, similar organizations (e.g. Ahora Madrid, Zaragoza en Común) obtained the Government of many of the largest Spanish cities. For this reason, future work should apply this framework to examine whether the characteristics observed in Barcelona en Comú are also present in these other grassroots movement-parties.

References

[1] Adamic, L. A. and Glance, N. (2005). The political blogosphere and the 2004 us election: Divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, pages 36–43. ACM.

[2] Aragón, P., Kappler, K. E., Kaltenbrunner, A., Laniado, D., and Volkovich, Y. (2013). Communication dynamics in twitter during political campaigns: The case of the 2011 spanish national election. *Policy & Internet*, 5(2):183–206.

[3] Barberá, P., Wang, N., Bonneau, R., Jost, J. T., Nagler, J., Tucker, J., and González-Bailón, S. (2015). The critical periphery in the growth of social protests. *PLOS ONE*, 10(11).

[4] Bennett, W. L. and Segerberg, A. (2012). The logic of connective action: Digital media and the personalization of contentious politics. *Information, Communication & Society*, 15(5):739–768.

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

[6] Borondo, J., Morales, A., Losada, J., and Benito, R. (2012). Characterizing and modeling an electoral campaign in the context of twitter: 2011 spanish presidential election as a case study. *Chaos*, 22(2):3138.

[7] boyd, d., Golder, S., and Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In *System Sciences (HICSS)*, pages 1–10. IEEE.

[8] Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.*, 30(1-7):107–117.

[9] Castells, M. (2013). *Networks of outrage and hope: Social movements in the Internet age*. J. Wiley & Sons.

[10] Congosto, M. L., Fernández, M., and Moro, E. (2011). Twitter y política: Información, opinión y¿ predicción? *Cuadernos de Comunicación Evoca*, 4:11–15.

[11] Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., and Flammini, A. (2011). Political polarization on twitter. In *ICWSM*.

[12] Edelstein, J. D. and Warner, M. (1979). *Comparative union democracy: Organisation and opposition in British and American unions*. Transaction Publishers.

[13] Feller, A., Kuhnert, M., Sprenger, T. O., and Welpe, I. M. (2011). Divided they tweet: The network structure of political microbloggers and discussion topics. In *ICWSM*.

[14] Fortunato, S. (2010). Community detection in graphs. *Physics reports*, 486(3):75–174.

[15] Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social networks*, 1(3):215–239.

[16] Garcia, D., Abisheva, A., Schweighofer, S., Serdult, U., and Schweitzer, F. (2015). Ideological and temporal components of network polarization in online political participatory media. *Policy & Internet*, 7(1):46–79.

[17] Garcia, D., Mavrodiev, P., and Schweitzer, F. (2013). Social resilience in online communities: The autopsy of friendster. In *Proceedings of the first ACM conference on Online social networks*, pages 39–50. ACM.

[18] Garimella, K., Morales, G. D. F., Gionis, A., and Mathioudakis, M. (2015). Quantifying controversy in social media. *arXiv preprint arXiv:1507.05224*.

[19] Gerbaudo, P. (2012). *Tweets and the streets: Social media and contemporary activism*. Pluto Press.

[20] Gini, C. (1912). Variabilità e mutabilità. *Reprinted in Memorie di metodologica statistica (Ed. Pizetti E, Salvemini, T). Rome*, 1.

[21] Golder, S. A. and Macy, M. W. (2014). Digital footprints: opportunities and challenges for online social research. *Sociology*, 40(1):129.

[22] González-Bailón, S., Borge-Holthoefer, J., Rivero, A., and Moreno, Y. (2011). The dynamics of protest recruitment through an online network. *Scientific reports*, 1.

[23] Granovetter, M. S. (1973). The strength of weak ties. *American journal of sociology*, pages 1360–1380.

[24] Gruzd, A. and Roy, J. (2014). Investigating political polarization on twitter: A canadian perspective. *Policy & Internet*, 6(1):28–45.

[25] Jacomy, M. (2011). Force atlas 2 layout.

[26] Karp, R. M. (1972). *Reducibility among combinatorial problems*. Springer.

[27] Katz, R. S. and Mair, P. (1995). Changing models of party organization and party democracy the emergence of the cartel party. *Party politics*, 1(1):5–28.

[28] Kwak, H., Lee, C., Park, H., and Moon, S. (2010). What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pages 591–600. ACM.

[29] Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., et al. (2009). Life in the network: the coming age of computational social science. *Science (New York, NY)*, 323(5915):721.

[30] Lietz, H., Wagner, C., Bleier, A., and Strohmaier, M. (2014). When politicians talk: Assessing online conversational practices of political parties on twitter. *arXiv preprint arXiv:1405.6824*.

[31] Lipset, S. M., Trow, M. A., Coleman, J. S., and Kerr, C. (1956). *Union democracy: The internal politics of the International Typographical Union*. Free Press Glencoe.

[32] Margetts, H. (2001). The cyber party. *ECPR Joint Sessions. London*.

[33] Michels, R. (1915). *Political parties: A sociological study of the oligarchical tendencies of modern democracy*. Hearst's International Library Company.

[34] Mills, C. W. (1999). *The power elite*. Oxford U. Press.

[35] Morozov, E. (2012). *The net delusion: The dark side of Internet freedom*. PublicAffairs.

[36] Mosca, G. (1939). The ruling class: elementi di scienza politica.

[37] Neff, J. J., Laniado, D., Kappler, K. E., Volkovich, Y., Aragón, P., and Kaltenbrunner, A. (2013). Jointly they edit: Examining the impact of community identification on political interaction in wikipedia. *PLOS ONE*, 8(4):e60584.

[38] Newman, M. E. (2004). Analysis of weighted networks. *Physical Review E*, 70(5):056131.

[39] Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23):8577–8582.

[40] Newman, M. E. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113.

[41] Palla, G., Derényi, I., Farkas, I., and Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818.

[42] Pareto, V., Livingston, A., Bongiorno, A., Rogers, J. H., et al. (1935). Mind and society.

[43] Peña-López, I., Congosto, M., and Aragón, P. (2014). Spanish indignados and the evolution of the 15m movement on twitter: towards networked para-institutions. *Journal of Spanish Cultural Studies*, 15(1-2):189–216.

[44] Rothschild-Whitt, J. (1976). Conditions facilitating participatory-democratic organizations*. *Sociological Inquiry*, 46(2):75–86.

[45] Sandiumenge, L. (2015). La guerrilla digital de Colau. http://districte15.info/la-guerrilla-digital-de-colau/. [Online; accessed 22-July-2015].

[46] Sunstein, C. R. (1999). The law of group polarization. *University of Chicago Law School, John M. Olin Law & Economics Working Paper*, (91).

[47] Sunstein, C. R. (2009). *Republic. com 2.0.* Princeton University Press.

[48] Toret, J. (2015). Una mirada tecnopolítica al primer año de podemos. seis hipótesis. *Teknokultura*, 12(1):121–135.

[49] Toret, J., Calleja-López, A., Marín, O., Aragón, P., Aguilera, M., Barandiaran, X., and Monterde, A. (2015). Tecnopolítica y 15m: La potencia de las multitudes conectadas.

[50] Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *nature*, 393(6684):440–442.