Master thesis on Intelligent and Interactive Systems
Universitat Pompeu Fabra

Computational comparative analysis of the Twitter networks of the 2015 and 2016 Spanish National elections

Helena Gallego Gamo

Supervisor: Andreas Kaltenbrunner, Pablo Aragón, Vicenç Gómez

July 2017

Universitat Pompeu Fabra
Barcelona

Master thesis on Intelligent and Interactive Systems
Universitat Pompeu Fabra

Computational comparative analysis of the Twitter networks of the 2015 and 2016 Spanish National elections

Helena Gallego Gamo

Supervisor: Andreas Kaltenbrunner, Pablo Aragón, Vicenç Gómez

July 2017

Contents

Acknowledgement

This is strange. I don't really want this 'thanks' part to sound like a template, or something robotic, because it shouldn't be and I am not really like that. But on the other hand, I also don't want this to sound cheesy and although I don't want to, it will probably sound like it and I will be very embarrassed the day I will present this to the people who I am writing this for. Anyway, here we go:

First, my supervisors, Pablo, Andreas and Vicenç. Isn't it obvious? Because they wrote me when I did that small little project in class and I thought, oh God, they don't really know what they are doing. And then we met and my interior voice was telling me 'Oh no, you can't do that' and, then I started in Eurecat and my little voice said 'Oh no, you can't do that' again. But somehow, we managed to write a journal, a paper and this report. I just want to thank you for that. Helena wouldn't have gotten far without them (yes, this is a sentence from LoTR), specially Pablo, which has not only been a supervisor and a colleague but a friend.

Also to all the digital humanities team (because of the cookies), David, Matteo, Ludo. David, for the laughs, seriously, you are so damn funny. Because you have listened to half of my dramas and well, also helped me a lot with my work, to be honest. Matteo, Ludo, to make me feel so well in the office all the time. I have really felt very welcome in Eurecat.

To all Eurecat colleagues that have contributed more to my 'futbolin' learning than SNA.

To Eurecat as a company, which hired to me do my thesis and has given me the opportunity to work with such a lovely amount of people and discover a lot of cool stuff.

To my family, Pati, Tedi, Imma, Albert and co. Always supporting unconditionally.

To the second family, for the laughs and the support. To Xènia, always best friend and lately, flatmate, to be the most amazing gypsy friend that someone can have in this world.

To Víctor, because even not being in my life anymore has made me the person who I am and has suffered more than anyone in the beginnings of this project and all this master program.

To Iván, because of reasons.

Well, seems like I managed to make my acknowledge sound more cheesy than a last will.

Abstract

In the last years, Spanish politics have transitioned from bipartidism to multipartidism. This change led to an unstable situation which finally evolved to the rare scenario of two general elections in the period of six months. The two elections had a main difference: the two biggest left-wing parties formed a coalition in the second election while they had run separately in the first one. In the second election and after merging, the coalition lost around one million votes contradicting opinion polls. In this study, community analysis in the retweet networks of the two online campaigns is performed in order to assess whether activity in Twitter reflects the outcome or parts of the outcomes of both elections. The results show that the left-wing parties lost more online supporters than the other parties. Furthermore, an inspection of the Twitter activity of the supporters unveils a decrease in engagement especially marked for the smaller party in the coalition, in line with post-electoral traditional polls. The clusters obtained with the community detection method are also used to situate in the ideological spectrum a set of Spanish media sources and to understand their audiences and behavioral differences when replying or retweeting them.

Keywords: Twitter; Politics; Political Parties; Spanish Elections; Online Campaigning; Political Coalition; Engagement; Political Participation; Mass Media; Media Sources; Journalism; Political Spectrum; Clusters; Community Detection; Volatile Electorate; Temporal Networks; Social Networks.

Chapter 1

Introduction

As social media are playing a key role in shaping public debate in political contexts, as a kind of new public sphere [1], it is increasingly important to understand their usage during political campaigns. One the one hand, as social media have a strong impact on voters' perceptions and decision making, it is important to understand their dynamics and influence [2], and their usage by politicians [3, 4]. On the other hand, social media can be observed as a mirror of trends underlying society [5]. Although translating signals from the online to the offline world is not always straightforward, and previous studies aimed at predicting election results through the analysis of Twitter [6] received many criticisms [7, 8, 9], it is undoubted that the analysis of social media as emerging political battleground can unveil important aspects of electoral campaigns. Indeed, a growing amount of research is devoted in particular to investigating multiple aspects of the usage of Twitter during elections, as illustrated in the systematic literature review presented in [10].

1.1 Motivation

This study is focused on the Spanish general elections of 2015 and 2016, and the activity on Twitter is compared during the two consecutive campaigns to assess whether and how it reflects changes in the engagement of the supporters of different parties. This case study is of special interest for several reasons. The first one is that the 2015 general elections marked the end of 40 years of Spanish bipartidism. After the country was shaken by the economic crisis of 2008 and by the 15M (or *Indignados*) movement of 2011 with massive protests against the two major parties [11], the elections in December 2015 were held in a very different scenario with respect to all previous elections [12]. The emergence of new political forces and the resulting fragmented parliament with no clear majority led, after six months of negotiation, to new elections in June 2016 [13]. The fact that two elections were celebrated within such a short time interval constitutes another element of interest, which motivates us to analyze and compare the two corresponding online campaigns. The main parties involved in the elections and having a presence in the whole country (sorted by electoral result) are the following:

- Partido Popular (PP) [1] - Traditional conservative party located in the center-right or political right.
- Partido Socialista Obrero Español (PSOE) [2] - Traditional social-democratic party located in the center-left political spectrum.
- Podemos (Pod) [3] - Left-wing political party founded in the aftermath of the 15-M Movement protests.
- Ciudadanos (CS) [4] - Liberal party created in 2005
- Izquierda Unida (IU) [5] - Traditional left-wing party

It is also important to mention the organizations Compromís (Valencia), En Marea (Galicia) and En Comú Podem (Catalunya), regional confluences that included local bottom-up forces in a coalition with Podemos.

From the results in Table 1, it is seen that the participation drops notably from the first to the second elections, suggesting a decrease in the motivation of the electorate. Also, they show that PP increased its votes in 2016; this, combined with

[1] https://en.wikipedia.org/wiki/People%27s_Party_(Spain))
[2] https://en.wikipedia.org/wiki/Spanish_Socialist_Workers%27_Party
[3] https://en.wikipedia.org/wiki/Podemos_(Spanish_political_party))
[4] https://en.wikipedia.org/wiki/Citizens_(Spanish_political_party)
[5] https://en.wikipedia.org/wiki/United_Left_(Spain))

Table 1: Participation, percentage of obtained votes and parliament seats per party for the 2015 and 2016 elections. Pod+ stands for the sum of Podemos, En Comú Podem, En Marea, and Compromís. In 2016 IU is added to this sum as well.

Election	Participation	PP	PSOE	Pod+	IU	CS	Other
2015	69.67%	28.71%	22.01%	20.4%	3.68%	13.94%	11.26%
		123	90	69	2	40	26
2016	66.48%	33.01%	22.63%	20.79%		13.05%	10.52%
		137	85	71		32	25

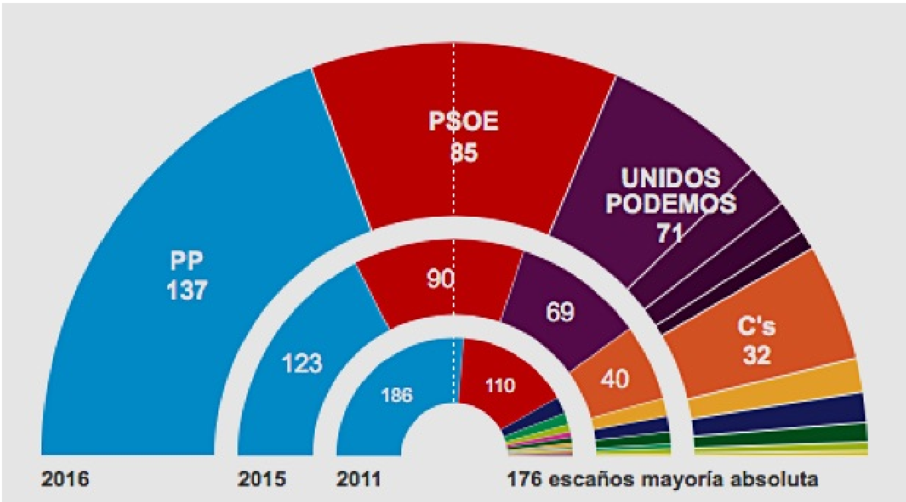


Figure 1: Representatives obtained by each party in 2016 and 2015 Spanish National elections. Source: ABC.com [14]

the participation drop, led to a higher amount of representatives for the party. In the 2015 election, some of the main left parties were presented in a coalition formed by Podemos, En Comú Podem, Compromís and En Marea. After some negotiations with Podemos, Izquierda Unida declined the offer to join as well the coalition. However, in the 2016 election, the two parties agreed and Izquierda Unida joined the coalition of 2015 which was re-named Unidos Podemos. The current Spanish electoral law, which penalizes small forces and gave IU only two representatives in the 2015 Congress after achieving almost one million votes (in Figure 1, they are not even noticeable), triggered the decision of Izquierda Unida to join the coalition in 2016. The results in Table 1 for these parties are shown separately in 2015 and together in 2016. It can be see that, although the sum of representatives is the same in 2016 than in 2015, the amount of votes dropped significantly (around 1 million votes) contradicting several polls[6].

1.2 Research questions

While the election results clearly indicate the increase or decrease in votes of each party between the two elections, they do not explicitly indicate voter migration between parties (or between parties and abstention-ism), which is left to opinion polls. In the 2015 election, there were around one million people who voted Podemos and IU who did not vote Unidos Podemos in 2016. Moreover, the winner party, Partido Popular increased its votes in 2016, and it can be hypothesized that it received votes of people who have voted for other parties in 2015. Several studies using post and pre-electoral polls tried to determine the voter transfers from one election to the other. The study [15] shows that probably only 73% from UP repeated their vote. Their voters did not vote PSOE, the other left party, they did not go to the schools as 15% of the former voters of Podemos and IU recognize that they did not vote on 26-J. In addition, [16] estimated that the coalition managed to retain 74% of Podemos voters (almost four million) but only six out of ten from IU (60%, around half a million). This means that there are also differences between where the voters went within the coalition electorate. To complement opinion polls about voter migrations between parties with evidence of social media activity, the following research question is presented:

- ***RQ1***: *It is observed from Twitter activity a migration of supporters between parties from the first to the second election?*

To answer this question, the retweet network is considered to perform community analysis to identify clusters of political parties and characterize their structure following the methodology of [17]. As retweets generally represent endorsement, they have been shown useful in previous literature to detect clusters corresponding to political parties, both in the context of Spain [3, 17] and of other countries [18]. The clusters obtained are used to study the migration of users between them. As known

[6]See for example `http://datos.cis.es/pdf/Es3141mar_A.pdf`

from the electoral results, the parties who constituted Unidos Podemos lost more than 1 million voters from the first to the second election. The hypothesis for this question is that a drop in the users clustered around the accounts of these parties will be observed. We further expect the analysis to indicate which of the parties in the coalition lost most supporters on Twitter and whether such lost users started supporting other parties. Several studies have examined the correlation between social media use and political engagement. Holt et al [19] report that both, political social media usage and attention to political news in traditional media, increase political engagement over time, and suggest that frequent social media use among citizens can function as a leveler in terms of motivating political participation. Findings from [20] reveal that a variety of Internet uses are positively related to different forms of political participation, whereas the relationship between most uses of traditional media and participation is weak. Finally, Dimitrova et al [21] demonstrate that there are only weak effects of digital media use on political learning, but that the use of some digital media forms has appreciable effects on political participation. From 2015 to 2016, participation dropped significantly showing a general demotivation or tiredness in the electorate. Given that Twitter activity can be related with the political engagement and there has been a motivation decrease between the two campaigns, the second question of this study is:

- ***RQ2:*** *Is the demotivation of the electorate reflected in their Twitter activity/engagement?*

The volume of activity per user in the two campaigns is analyzed to answer this research question and determining if there are notable differences between them. We will look separately at users supporting different parties, with a special attention towards Podemos and IU, the parties that lost more votes.

Meanwhile the Spanish citizenship lost the confidence in the traditional parties and showed a general in-conformism with the politico-social picture that led to the 15M movement, something similar happened with mass media. A lately study from Reuters and the Oxford University [22] concludes that the Spanish media are the least credible of the eleven countries consulted in Europe and the second least credible of the twelve studied around the world. Furthermore, mass media is known to be following the propaganda model [23] to manipulate populations. In Spain specifically, mass media in Spain is being widely criticized to be positioned towards a party [24] and it is interesting to understand which sources are related to which parties or where are the media sources situated in the ideological spectrum. Provided the current criticism to Spanish mass media and their possibility of them to be related to different ideologies or parties, the third question of this study is:

- ***RQ3**: Can we situate in the political spectrum the media sources by their online audience?*

To answer the research question the clusters obtained previously will be used to determine the distribution of party supporters retweeting or replying the different media sources. Having analyzed who retweets and replies the media sources, the audience will be used in order to compute an ideological indicator to situate them in the political spectrum. The study is compared with another one [25] which pursues the same goal using another methodology.

Undecided voters have been always under the spotlight in Politics sciences as they are a large and volatile group with the potential to determine the election result [26]. Several studies have tried to determine their profile and their decision-making processes [27]. Several studies [28] claim that even a few months before the election, rates of 20% or more undecided voters are not uncommon and also, data indicate that most of these undecided respondents come to a decision only a few days before the vote, if not the very same day of the election. For the Spanish National 2016 elections, the CIS study held one month before the celebration [7] reported that almost a 32,4% were still undecided. That is, during campaign time, a huge amount of undecided citizens would have had the ability to change the party they voted in the 2015 election.

[7] `http://datos.cis.es/pdf/Es3141mar_A.pdf`

The post-electoral study [15] done by Metroscopia, a Spanish team in poll studies, revealed that PP was the party with more faithful voters for the 2016 elections: an 86% of those who supported the party in 2015 did it again in 2016. PSOE and Podemos, had also a very faithful electorate, with a 76% and a 73% of loyal voters. Ciudadanos and IU, on the other hand, are shown as the parties most likely to have been 'betrayed' by their December voters, since only 59% of who supported them confess that they did do it again in the second election. That is, PP, PSOE and Podemos, counted with a less volatile electorate than the other parties and thus, their electorate has been less variable than in the other parties.
Given that we know the percentage of faithful voters for the parties in the second election provided by post-electoral polls and also that a high amount of undecided voters will be deciding their vote during the campaign and would be maybe changing their votes during this time frame, our last research question is:

- ***RQ4***: *Can we observe from Twitter data from the second election which are the parties with the most loyal electorate and which have the more volatile electorate?*

In order to answer this research question, we present a methodology that computes the party clusters volatility over the 2016 campaign. The methodology extracts a daily re-tweet network, applies the methodology used previously to obtain the party clusters, matches them between days and, using the user's changes across days, computes the overall volatility of a party. Finally, we will compare if the party clusters volatility obtained for the campaign is in line with the loyalty rates given in the post-electoral polls. A high party cluster volatility would indicate less loyalty from their electorate as this would be more changeable and less stable during the campaign timeframe.

1.3 Structure of the report

After motivating and introducing the research questions of the study, the structure of the report is explained. First, related work on the topics related to the research is detailed in order to give some background to the field. Second, the dataset used for the study is presented. The methodology applied in order to answer the research questions is explained in the Methods section followed by the results obtained. The results are sectioned in a chapter for each research question and are discussed in the Discussion section. Finally, the overall conclusion of the study is expressed in the Conclusion section with suggestions on a further work for the study in the last section.

Chapter 2

Contributions of this work

The section is divided in two, considering the improvements and contributions in relation to the fields of the thesis: social and computer science.

From the social network analysis field, three further improvements in methodologies can be detached. First, the methodology presented in [17] is improved resulting in the N-Louvain method from this study. The improvement includes the usage of the Jaccard coefficient in order to match the communities across iterations providing robustness and automation to the method. Second, a method to analyze cluster dynamics between two networks has been developed to understand the migration of users belonging to communities. Finally, a new methodology to track community evolution and stability based on the N-Louvain method following the two-step approach, detect communities and match them across networks, has been implemented.

On the social and politics field, several conclusions have been extracted in the Spanish context. First, it has been concluded that, although the results were not as explicit as expected, Twitter activity can be seen as an indicator of political engagement: the results provided by the methodology applied to the retweet networks was in line with the post-electoral polls. Also, several Spanish media outlets have been analyzed and situated in the political spectrum using Twitter data. The study allowed us to determine that the left parties in Spain tend to be more critical with the media outlets online unlike the right ones. Finally, it is worth to mention that this study has been submitted to Socinfo2017 conference (9th International Conference on Social Informatics) and it has been accepted to be presented in its poster session.

Chapter 3

Related work

Previous literature is found in relation to almost all topics covered in the study. A review on previous research is presented in this section and will be used as background to develop the methodology to answer the research questions.

We go further from the studies [2, 3, 4, 5, 6, 7, 8, 9, 10] cited in the Introduction, where the importance on understanding the social online world has been highlighted presenting strong evidence that both worlds can be strongly related. To do that, we present other studies: [29] finds that the participation in online political groups is strongly correlated with offline political participation, as a potential function of engaging members online. Furthermore, in [30] distinct submodes of e-participation, comparable to those occurring offline, can be identified and their results suggest that the online environment may be fostering a new social-media-based type of expressive political behavior. To answer our first research question, we will use the extend the methodology of [17] of considering the retweet network and perform a community analysis to identify clusters of political parties and characterize their structure. Performing community analysis in retweet networks has been shown useful in previous literature to detect clusters corresponding to political parties, both in the context of Spain [3, 17] and of other countries [18].

There is an extensive literature questioning the relation between media usage and political engagement developed during the last years. Mostly all studies are based on analyzing different social network datasets in other to establish their deliberative practices and their relations with the offline world. In [31], the evidence analyzed across fifteen cases from varied countries suggested that engagement with digital environments influences users' political orientations and that contextual features play a significant role in shaping digital politics. Also [32] provide a strong evidence against the Internet having a negative effect on engagement and, even that the data did not establish that Internet use will have a substantial impact on engagement, the effects of their use on engagement seem to increase across time. Furthermore, the book [33], which provides linkages to established theories of media and politics, political communication, governance, deliberative democracy and social movements, also proves that the Internet usage does contribute to the heterogeneity of political discussions. Contrariwise, in [34] is commented that the proliferation of virtual communities over the net, in and of itself, is not an indicator of political revitalization, however, deliberative practices of citizens could be an integral element to regenerate civic political life. When comparing on-line and off-line environments, it is important to consider the profiles of the citizens involved in order to understand that profiling differentiation can derive to misconceptions in the analysis or predictions. For example, In Spain, results [35] show that, while online participation is mainly associated with internet-related skills, there is a significant gender gap. Also, the unemployed tend to engage socially and politically online more than the rest of the population. Those are topics which are worth to consider when trying to compare both real and online worlds.

A lot of literature has been written in the social science field explaining how propaganda and systemic biases function in mass media. In [23], a model that seeks to explain how populations are manipulated and how consent for economic, social, and political policies is "manufactured" in the public mind due to this propaganda is presented. In Spain specifically, media sources have always been known to be positioned towards an ideology or explicitly to a party and business groups [24]. A recent study from a researcher in Universidad de Navarra [25], uses Twitter data from the debate held before the elections in 2015 to determine the situation in the political spectrum of each media source. Our study differs from this one for several reasons: One the one hand, the data collected for their study was only taken during one day and focused only in the debate hashtag. On the other hand, their study is based on polls within Twitter provided by different media sources while this one computes the indicator based on the distribution of retweets by party supporters given by a community detection method. Finally, their study only analyzes the polls presented in Twitter after the debate to situate them in the spectrum while in this study different kinds of behavior at retweet and reply level are also analyzed.

Different solutions for tracking dynamic communities and its stability in dynamic networks have been developed in the past. Several models consider the two step approach for the purpose: communities are first detected for each time stamp, and then compared to determine correspondences. The methods differ on the matching between communities. [36], for example, bases their algorithm on clique percolation and allows to investigate the time dependence of overlapping communities on a large scale and uncover basic relationships characterising community evolution. [37] presents two methods: one that consists of statistical analyses and visualizations for an interactive analysis of subgroup evolutions in communities that exhibit a rather membership structure and, another focused in communities in an environment with highly fluctuating members. [38] tracks the dynamic communities using the Jaccard coefficient, as done in our study and [39] uses a similar method to study dynamics in networks of Face-to-Face Human Interactions. Other researchers go further from the two step approach and develop alternative methodologies with the same finality. However, their main goal is mostly to provide a better performance in large networks, which is not our focus in the study. [40] proves that the baseline approach by enumerating all communities in each graph and comparing all pairs of communities between consecutive graphs is infeasible and impractical and propose an efficient method by introducing graph representatives and community representatives to avoid generating redundant communities and limit the search space. Also in [41], the researchers argue the two step approach is inappropriate in applications with noisy data and they propose a method for analyzing communities and their evolutions through a robust unified process: communities not only generate evolutions, they also are regularized by the temporal smoothness of evolutions. [42] proposed a method that relies on the principle of Minimum Description Length (MDL), to extract the communities, and to find good cut-points in time when communities change abruptly. Finally, [43] proposes a method that given a sequence of snapshots of an evolving graph, discovers rules describing the local changes occurring in it. Adopting a definition of support based on minimum image they study the problem of extracting patterns whose frequency is larger than a minimum support threshold.

Chapter 4

Dataset

The study is based on two different datasets collected from Twitter in relation to the electoral campaigns of the 2015 and 2016 Spanish national elections (collected during December 4-20 2015 and June 10-26 2016 respectively). The data collection was based on party official accounts and party candidate accounts. For each election, we collected all tweets that either: (a) were created by, (b) retweeted or (c) mentioned one of these accounts. The list of all party candidate and official party accounts considered for data collection is detailed in Table 3.

To detect the Twitter organization of political parties, we build directed weighted graphs of users (nodes) and retweets (edges). Each edge indicates that the source user retweeted a message posted by the target user. We filter edges with weight lower than 3 to exclude anecdotal interactions as done in [17]. The two resulting networks for 2015 and 2016 have the following characteristics presented in Table 2.

Table 2: Retweet network stats for 2015 and 2016: number of retweets for the whole election (# tweets), number of nodes (N) and edges (E) in the network, clustering coefficient (cl) and average path length (l).

Elections of	# tweets	N	E	cl	l
2015	3 196 677	57 575	164 411	0.004	7.18
2016	1 602 528	72 269	168 135	0.0015	6.215

During the 2016 election, data in relation to several Spanish media sources were collected in addition to the candidates and parties data. The list of Spanish media sources in scope and their related accounts are found in Table 4 in the Appendix. The data has been collected in the same way as done for the political parties and candidates but using the media sources official accounts.

Table 3: Twitter accounts of the selected political parties and candidates which were used to retrieve the datasets.

Party	Party account	Candidate account
PP	@PPopular	@marianorajoy
PSOE	@PSOE	@sanchezcastejon
Podemos	@ahorapodemos	@Pablo_Iglesias_
IU	@iunida	@agarzon
C's	@CiudadanosCs	@Albert_Rivera
En Comú Podem	@EnComu_Podem	@XavierDomenechs
Compromís	@compromis	@joanbaldovi
Equo	@Equo	@juralde
Marea-Anova-EU	@En_Marea	@tone_corunha
ERC-CATSÍ	@Esquerra_ERC	@gabrielrufian
DL	@ConvergenciaCAT	@franceschoms
EAJ-PNV	@eajpnv	@MikelLegarda
Bildu	@ehbildu	@ikerurbina1
CCa-PNC	@gnacionalista	@PabloRodriguezV

Table 4: Twitter accounts of the selected Spanish media sources.

Media source	Description	Twitter account
324	Regional (Cat) television channel	@324cat
ABC	National newspaper	@abc_es
Antena 3	National television channel	@antena3com
La brújula	National radio program	@brujulaondacero
betevé	Regional (BCN) television channel	@btvnoticies
COPE	National radio channel	@cope_es
Al rojo vivo	National television program	@debatalrojovivo
El món a RAC1	Regional (CAT) radio program	@elmonarac1
La SER	National radio channel	@la_ser
La RAZÓN	National newspaper	@larazon_es
La Sexta	National television channel	@lasextatv
Las Mañanas de RNE	National radio program	@lasmananas_rne
Las Provincias	Regional (VAL) newspaper	@lasprovincias
La voz de Galicia	Regional (GAL) newspaper	@lavozdegalicia
Levante-EMV	Regional (VAL) newspaper	@levante_emv
Libertad Digital	Online national newspaper	@libertaddigital
NOTICIAS en Cuatro.com	National television program	@noticias_cuatro
El objetivo	National television program	@objetivolasexta
Onda Cero	National radio channel	@ondacero_es
Publico	National newspaper	@publico_es
Radio 5, RNE	National television program	@radio5_rne
La Sexta Noche	sextanochetv	@sextanochetv
Telecinco	National television channel	@telecincoes
TVE	National television channel	@tve_tve

Chapter 5

Methodology

5.1 N-Louvain method

Many previous studies have relied on the Louvain method [44] because of its high performance in terms of accuracy, and its efficiency. However, the usage of this algorithm for detecting clusters corresponding to political parties raises some issues. Given that the algorithm has a random component, every execution may typically produce different partitions for the same network. To obtain robust results, and classify only nodes who reliably fall into a given cluster, we follow the method introduced in [17], based on the idea of executing multiple times the Louvain algorithm, and classifying only nodes that fall most of the times into the same cluster.

To identify each cluster across executions, we improve the previous method by applying the Jaccard index [45] to every pair of clusters c_i and c_j across different executions:

$$J(c_i, c_j) = \frac{|c_i \cap c_j|}{|c_i \cup c_j|}.$$

Thus, clusters across executions are matched if they are the most similar ones. This allows us to assess the proportion of times a node falls within the same cluster. 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, ε represents the sensibility level of the algorithm ($\varepsilon = 0.05$ in this study). This procedure allows

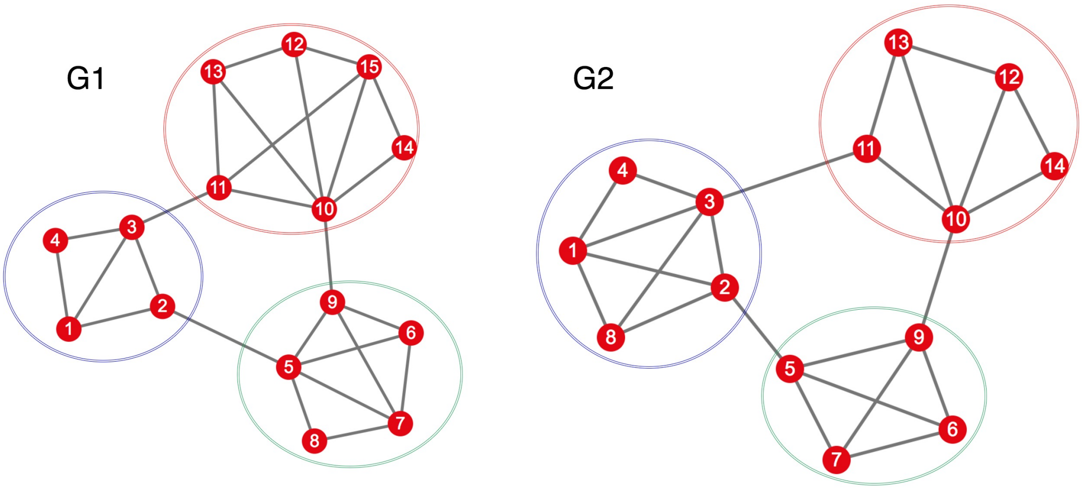


Figure 2: Example that shows two clustered graphs, G1 and G2. 8 and 15 nodes belong to Other cluster and None categories, respectively, while all the rest belong to the Same cluster category.

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 a given 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.

5.2 Cluster changes between networks

To characterize how users change between two consecutive networks, G_1 and G_2, we consider five possible categories, depending on how a user i that belongs to a cluster in G_1 is related to the clustering in G_2. Let $c_1(i)$ and $c_2(i)$ denote the cluster to which i belongs in G_1 and G_2, respectively. There are three main possible scenarios, either the user belongs to the same cluster in both networks, $c_1(i) = c_2(i)$ (**Same cluster**), it belongs to different clusters, $c_1(i) \neq c_2(i)$ (**Other cluster**), or i does not fall robustly in any cluster of G_2. In the last case, we can still assign a cluster to i depending on whether i retweeted users belonging to the same cluster $c_1(i)$ (we call this category **Associated with same cluster**) or retweeted users belonging to another cluster (**Associated with other cluster**). Finally, if the level of activity of i does not reach the threshold to be included in G_2 (we only include interactions that occur at least three times), we assign i to the category **None**.

5.3 Cluster volatility in a time frame

The **cluster volatility** is a metric that characterizes whether the clusters are very changeable during a certain period of time nor not. While the size of the cluster in two consecutive time stamps can be the same, the nodes within them could have changed drastically. The volatility is the cluster average balance, the amount of new users less the amount of lost users in the clusters normalized by the cluster size in the previous day. The balance allows us to determine which percentage of the cluster is different respect the previous day, either negative (more lost than new users) or positive (more new than lost users). In order to compute the cluster volatility, we apply the following method:

First, a network is generated for each time stamp t building a directed graph with all retweets comprised between t until $t + w$. In our case, we generate a network for each day from the election campaign and we will have, for each day, a network that comprises retweets starting that day until w days later. Second, we apply the N-Louvain method with N $=$ 100 and $\varepsilon = 5$ to each time stamp network in order to obtain the final clusters for each time stamp. Third, we match with the Jaccard coefficient all clusters in each time stamp with the obtained ones in the previous time stamp in order to compute the balance between stamps. Given two time stamps $t1$ and $t2$, two clusters (c_1 and c_2) which are matched within both stamps, that means, c_2 is the evolution of c_1, and have different sizes, s_1 and s_2, for each time stamp, the balance between them ($bal_t 1 - t2$) is difference between new nodes ($\#$ of nodes present in the cluster in $t2$ but not in $t1$) and lost nodes ($\#$ of nodes present in the cluster in $t1$ but not in $t2$). Note that s_1 plus bal_{t1-t2} is equal to s_2.

$$b(c)_{t_1 - t_2} = \frac{new(c)_{t_1 - t_2} - lost(c)_{t_1 - t_2}}{s_1}$$

Finally, we obtain the cluster variability for the time frame computing the average balance for all time stamps for each cluster.

Chapter 6

Results

We start showing some general results about our community discovery analysis for both election campaigns. We then analyze how the found clusters change between the two elections, follow with a quantification of the change in political engagement and end this section with an analysis of the Spanish mass media online audience.

6.1 Community detection

Table 5 shows the clustering results obtained using the N-Louvain method in both networks. For clarity, we only show the largest clusters.

Table 5: Number of nodes (N_{2015} and N_{2016}) and edges (E_{2015} and E_{2016}) for the intra-network of each cluster in the retweet networks of 2015 and 2016.

Cluster	N_{2015}	E_{2015}	N_{2016}	E_{2016}
Podemos	16 114	33 488	9 771	12 818
IU	10 439	22 422	10 314	12 304
PP	8 345	28 677	5 614	11 682
PSOE	7 538	25 119	5 541	10 174
CS	7 200	24 110	5 458	9 501
ECP	1 412	2 925	1 791	2 868

We observe that, out of the four parties that formed the coalition –Podemos, En Comú Podem (ECP), En Marea and Compromís–, only two clusters are identified, the ones corresponding to Podemos and ECP. Whereas En Marea and Compromís

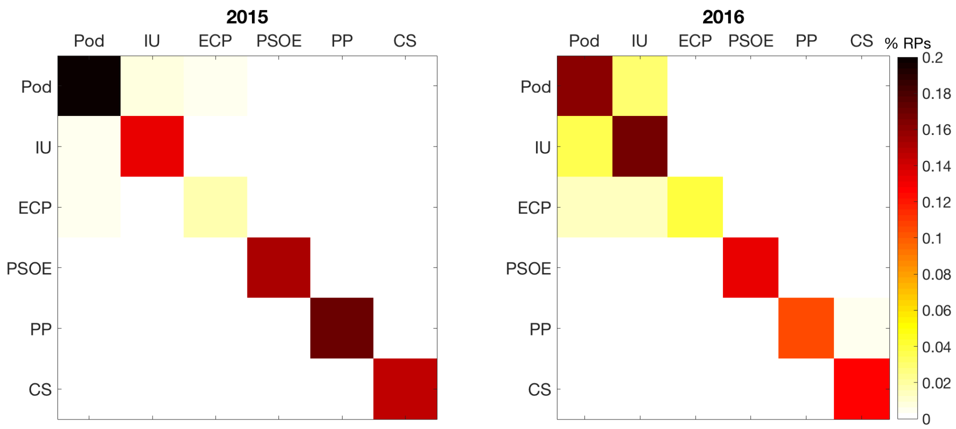


Figure 3: Normalized weighted adjacency matrices of the 2015 (left) and 2016 (right) retweet networks aggregating nodes by party clusters.

are effectively integrated into Podemos, the party of ECP, in contrast, forms a separate cluster. This can be explained due to the different language and the specificities of the debate about Catalan independence, which creates stronger intra-party interaction. In addition, we also observe that the IU party keeps its own cluster in both elections, despite merging with Podemos in the second election. The IU cluster is slightly bigger than the one of Podemos in the second election. This is noteworthy, since Podemos had by far a larger amount of votes in the first election, and one might expect the opposite effect in the network. In general, we can conclude that the formation of the coalition for the 2016 election *is not captured* by the observed communities since Podemos, IU, and ECP are identified in different clusters. Moreover, there is no simple obvious relation between the size of the identified communities and the electoral results, in terms of votes.

We now analyze the inter-cluster and intra-cluster density of edges. This will provide a measure of how strongly connected are the clusters within the same party and between the different parties in the different campaigns. Table 5 (second and fourth columns) shows that the amount of intra-cluster edges is smaller in 2016 than in 2015, with a decrease of almost a half, indicating weaker connections in the second elections.

What about the inter-cluster edges? We would expect some of these interactions to increase in the second election, as a consequence of the electoral coalition and of the synergies between the parties. To examine all the interactions between the parties, we consider the interaction matrix A, where A_{ij} is the sum of all retweets that users from cluster i made for the tweets from users of cluster j. Since the clusters have different sizes, we normalize A_{ij} by the sum of all retweets made by the users assigned to cluster i. Figure 3 shows this interaction matrix A for both election campaigns. As expected, both matrices are diagonally dominant, since the vast majority of retweets were made between users from the same cluster in both elections, being in 2015 this behavior more pronounced than in 2016. Comparing the parties involved in the coalition, we clearly observe that their interactions are increased in 2016, as the (yellow) off-diagonal elements indicate. Interestingly, the interaction between ECP and the other members of the coalition is not symmetric. This fact may be explained again by linguistic reasons since ECP users retweet both messages in Spanish and Catalan, but most users in the Podemos and IU clusters only speak Spanish and therefore do not retweet ECP messages in Catalan language. We conclude that despite the coalition is not captured at the clustering level (parties within the coalition do not merge into a single cluster), *it is captured* at the level of the interactions between clusters, that increase remarkably in 2016.

6.2 Cluster dynamics between the two elections (RQ1)

We now analyze how the clusters change between campaigns. Table 6 shows some general indicators. For each party cluster, we report its size in 2015 and 2016, the number of users which are present in the cluster in 2015 but not in 2016 (*lost* column) with the corresponding percentage in parenthesis, the number of users which are present in the cluster in 2016 but were not in 2015 (*new*), and the *balance*, or difference between new and lost users. The bottom row shows the quantities for Unidos Podemos (UP), which corresponds to the sum of the parties involved in the 2016 coalition. Although the coalition did not exist in 2015, we use it as a reference in our analysis.

We observe that all but a single cluster (ECP) shrink in the second campaign (negative balance), indicating a significant decrease in activity and suggesting an overall decrease in motivation. Another important observation is that all clusters lose more than half of the users they had in 2015. The cluster that loses fewer users is PSOE (62%) and the cluster that loses most users is Podemos (nearly 80%). This illustrates the high variability between the users assigned to the clusters in the two campaigns. More precisely, each cluster has a core of no more than 38% of users that are kept in the subsequent elections, while the largest majority of users is lost.

The only cluster with positive balances is ECP, with 379 more supporters in 2016. In contrast, Podemos has the highest negative balance among all clusters, losing 6 324 users. Notice that IU, although being the cluster apparently more stable (with the smallest absolute balance), it is actually the second one that lost more members. This is explained by the fact that IU formed coalition in the second election and, in spite of losing many users, many new ones joined from other clusters. Looking at the joint cluster Unidos Podemos (UP), we see that it suffers the highest loss compared to the other parties not in the coalition (69.6% of UP vs 62%–65% of the others). This means that are not all the users migrate within the parties of the coalition. Since losing a user from the community does not necessarily mean that the user stopped being a political supporter, we now focus our analysis on the lost users of each cluster.

To understand how users migrate between clusters in the two networks, we apply

Table 6: Main clusters per party. In columns: cluster sizes in 2015 and 2016, # of users present in the cluster in 2015 but not in 2016 (*lost*) and the corresponding percentage, # of users present in the cluster in 2016 but not in 2015 (*new*), difference (*balance*) between new and lost users. Last line (UP) is the sum of ECP, Podemos, and IU.

Cluster	size 2015	size 2016	*lost*		*new*	*balance*
CS	7 200	5 458	4 771	(66.3%)	3 029	-1 742
PP	8 345	5 613	5 446	(65.3%)	2 714	-2 732
PSOE	7 538	5 541	4 674	(62.0%)	2 677	-1 997
ECP	1 412	1 791	930	(65.9%)	1 309	379
Podemos	16 113	9 771	12 806	(79.5%)	6 464	-6 342
IU	10 439	10 313	7 792	(74.6%)	7 666	-126
UP	27 964	21 875	19 448	(69.6%)	13 359	**-6 089**

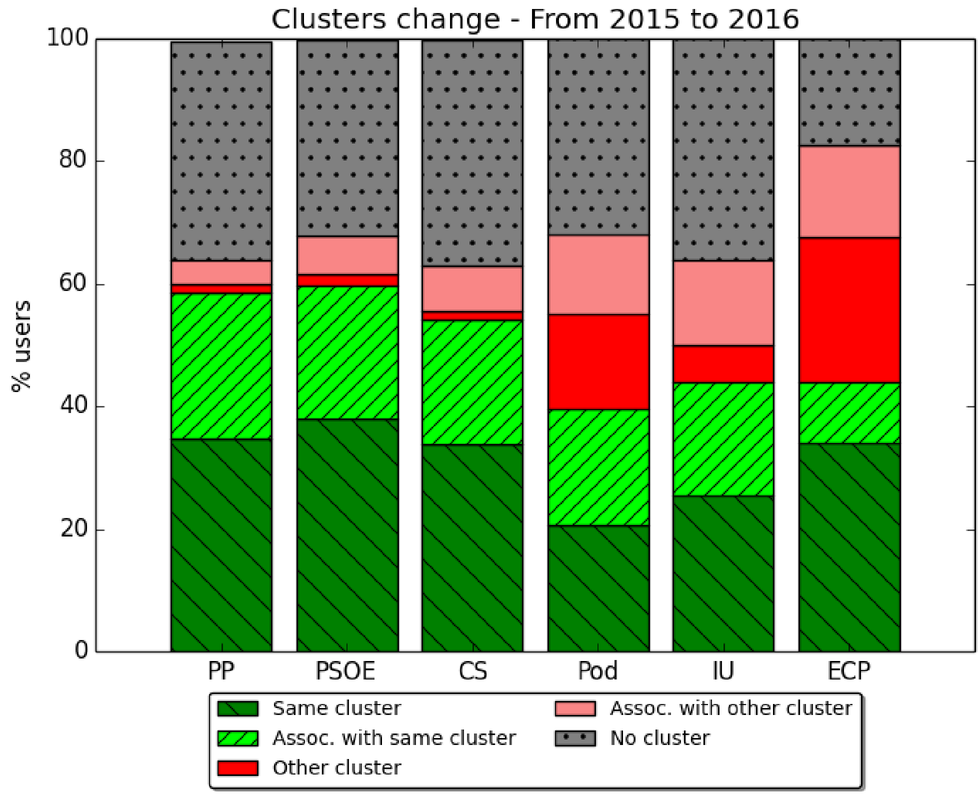


Figure 4: Proportion of users from each cluster in 2015 who: remain in the same cluster in 2016 (*same cluster*); retweet mostly users from the same cluster in 2016 (*associated with same cluster*); lie in another cluster in 2016 (*other cluster*); retweet mostly users from another cluster in 2016 (*associated with other cluster*); are not associated to any cluster in 2016 (*no cluster*).

the methodology described in Section 5.2. Figure 4 shows the distribution of the different categories of users for each party cluster in 2015, providing a more detailed view of the (*lost*) users in Table 6. Note that the values in that column correspond to the regions in the figure that not coloured in dark-green.

We analyze first the distributions of PP, PSOE, and CS. They follow a similar pattern with around 35% of users remaining in the same cluster and around 25% of users associated with the same cluster. Therefore, for these three clusters in 2015, despite losing the majority of users according to our clustering criteria, we can say that approximately 60% of their users do not change their support in 2016. The remaining 40% (approximately) is composed mainly of users who do not have a cluster assigned in 2016 and by a small percentage of users who migrated to other

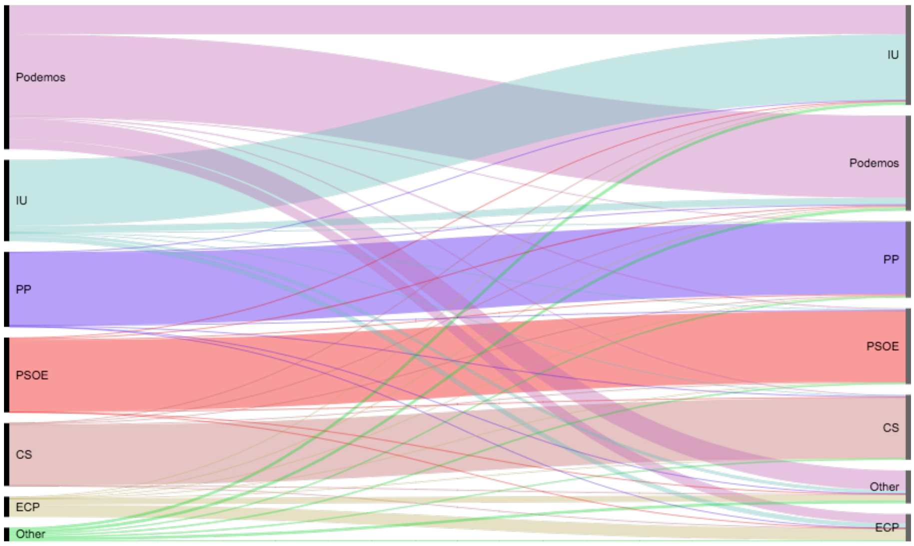


Figure 5: Match of cluster users: amount of users from a 2015 cluster (left) in the 2016 clusters (right).

parties, either falling in the corresponding cluster or just being associated with it. Notice that the latter is very unlikely.

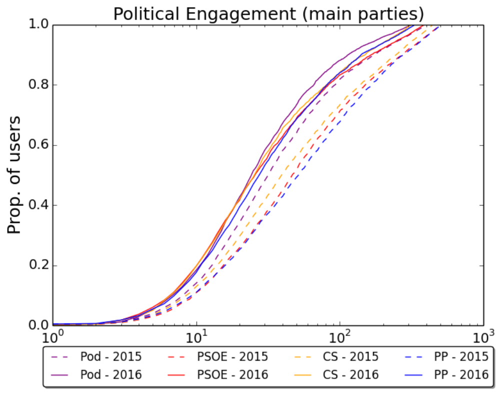
Regarding the clusters corresponding to UP, we observe that Podemos has the smallest number of stable supporters (dark/light green) and that ECP is the one with the smallest proportion of users who do not fall in any cluster in 2016 (gray). The latter observation indicates that ECP users keep a high activity in the 2016 campaign. It is interesting to mention that, when viewed independently, Podemos, IU, and ECP have a smaller proportion of users that stay in the same cluster or are associated to the same cluster (dark and light green in Figure 4) than the other parties do. However, when considered altogether in UP, the proportion increases and becomes comparable to the other parties. This fact suggests that migrations occur mostly within the clusters of the UP coalition parties. This is confirmed in Figure 5, which shows the flow of clustered users between campaigns for the users clustered in both elections (either in the **same cluster** or in an **other cluster**). It is noticeable that most of these users fall in the same party in both elections, indicating a strong political association. Clearly, Podemos is the cluster that suffers more changes, with a considerable amount of users that mostly migrate to IU and, to a lesser extent, to ECP. We do not see the same behavior in IU.

The following conclusions are extracted from the entire analysis on cluster changes: Although the large variability observed initially in the compositions of the clusters, when one considers the *associated* class, there does not seem to be a high amount of migrations between clusters, with exceptions within Unidos Podemos. It seems that users who actively participate in online campaigns on Twitter are usually very positioned towards one party and only retweet from other parties very sporadically. In general, users that retweet the messages of a party tend to either keep supporting the same party, or stop participating actively in the campaign.

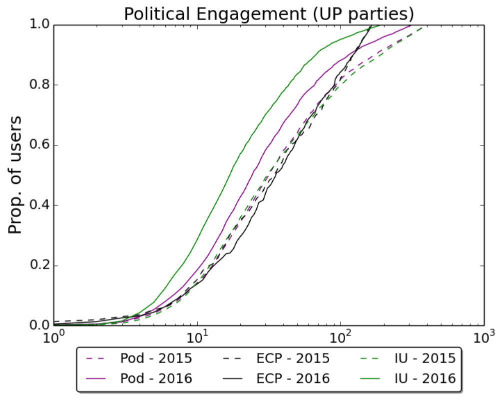
Unidos Podemos is the entity that loses more support from the first to the second election, as Table 6 and Figures 4 and 5 show. The total balance between the two elections is negative and stronger than for the rest of the parties. However, when analyzing the nature of the cluster in 2016 and its changes in relation to 2015, this negative balance is not as high as expected in relation to the other parties from the electoral results and it does not seem to reflect the general demotivation which was interpreted from the electoral results. In Unidos Podemos, we have seen a strong migration of supporters from Podemos to IU which did not happen in the opposite direction. The Spanish electoral law that favours bigger parties may have had an influence, pushing citizens closer to IU to vote and campaign for the bigger party Podemos in 2015.

6.3 Political engagement (RQ2)

Activity in Twitter can be an indicator of the political engagement of the population. To characterize the activity of users in each cluster and in each election, we calculate the cumulative distribution function or probability $P(X \leq x)$ that the number of user retweets X is less than or x, for those users that were present in both campaigns. Results are displayed for PP, PSOE, CS, and Podemos in Figure 6a. The solid curves lie above the dashed ones, indicating a decrease of activity in all parties. These results confirm that political engagement decreased, perhaps due to the user fatigue after a long period of political activity. To analyze the engagement within the different parties that form the UP coalition, we break down



(a) Podemos, PP, PSOE and CS.



(b) UP clusters: Podemos, IU and ECP

Figure 6: Cumulative distribution of the number of tweets per users who fall in the same party cluster in 2015 (dashed) and 2016 (continuous). **(a)** The four major parties have all less active users in the 2016 campaign. **(b)** For parties in the UP coalition, a noticeable larger drop of activity of IU users, while ECP users maintained their level of activity.

the coalition UP and show in Figure 6b the cumulative distribution functions for each UP cluster individually. First, we observe that all curves show a similar profile in 2015. However, in 2016 the picture changes. We observe that IU has much less activity than Podemos. Since our analysis includes the strongest supporters of the party only, a decrease of their activity suggests that those users might have been unhappy with the coalition and were demotivated during the second election. This result is in agreement with previous literature [13] and with the post-electoral study from Metroscopia [16] which reported that the UP coalition retained only three out of four Podemos voters (74%) and only six out of ten IU voters (60%). Moreover, ECP shows the opposite effect compared to the rest of parties (it actually increases its activity in the second election), also in agreement with the electoral results in Catalonia, where Unidos Podemos lost fewer voters from 2015 to 2016. From these results, we can conclude that our proposed methodology of using the clusters and measuring the activity distribution satisfactorily captures the observed behavior with respect to engagement observed in the election.

6.4 Media indicators in the second election (RQ3)

We analyze the Spanish mass media in Twitter in the following section. We try to situate each Spanish media outlet in the ideological political spectrum using Twitter data. First, the audience of each media outlet is analyzed exhaustively and an ideology indicator is computed afterward.

6.4.1 Audience for media outlet based on clustering

The users labeled as party supporters in the previous sections are used to determine the audience of a media outlet. In Figure 7, the audience of each media outlet is observed. Each bar represents the proportion of retweets done by users labeled as party supporters from the largest clusters obtained in the previous section (the users labeled as Others belong to other clusters). The distribution is done normalizing the tweets without considering the users which did not fall into any cluster, which was considerably higher than expected. However, the amount of labeled users is enough to get an idea of the media outlet audience. The media are sorted for a better visual interpretation of the results. The sort is done using the amount of retweets done by Partido Popular supporters, as it is the party with a higher presence retweeting media outlets. We can see that in the left publico_esm, the media outlet with fewer retweets from PP supporters and the media outlet in the right, the one preferred by those.

Due to the fact that the retweeting is considered as an endorsement tool, retweets were more suitable data to determine the audience of each media outlet. However, it is also interesting to show the same results using replies instead of retweets in order to see if there are behavioral differences and, to determine which parties are being more active when replying media outlets. Figure 8 shows the same stacked bar but using replies for the distribution computation.

Interesting conclusions from Figures 7, 8 and Tables 8, 7 are extracted. First, two types of mass media based on their retweet audience are observed: On the one hand, there are media outlets whose audience is almost one party (ie. cope_es, la_razon)

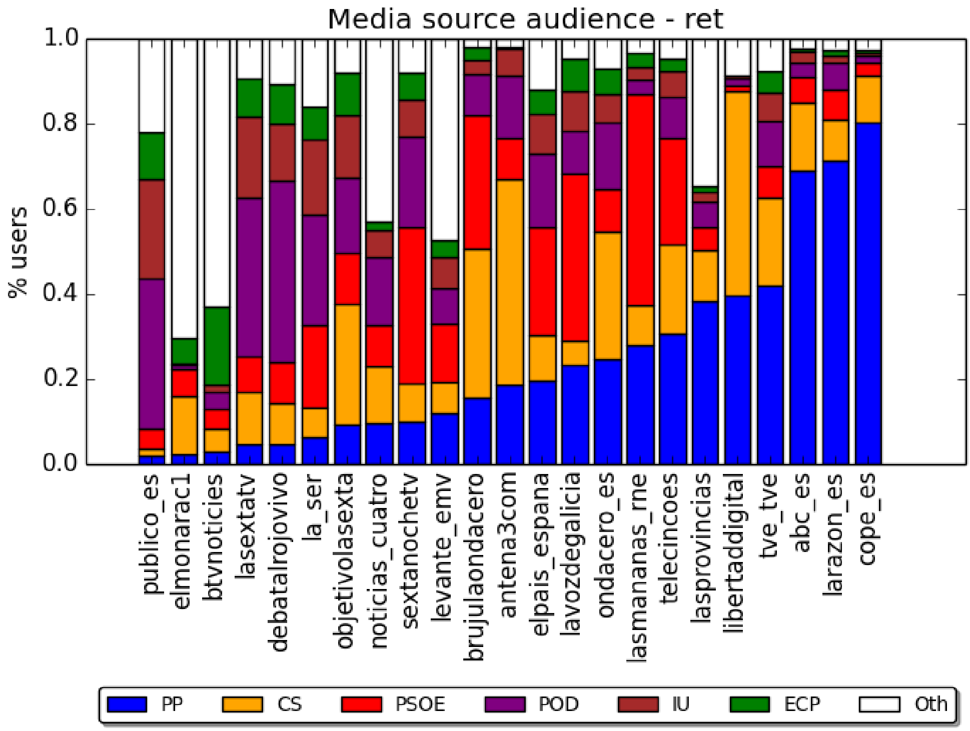


Figure 7: Distribution of replies done by party in each Spanish media. The figure is sorted by amount of retweets done by PP.

Table 7: Average percentage of replies and retweets done to Spanish media outlets by users labeled as party supporters.

Cluster	Retweets	Replies
PP	**24.58 %**	18.67%
CS	16.73%	10.46%
PSOE	15.01%	9.81%
POD	13.89%	**24.61%**
IU	7.42%	14.04%
ECP	5.27%	4.51%
Other	17.07%	17.84 %

and on the other hand, we observe other media outlets which have an heterogeneous audience (ie. elpais_espana). The 3rd column in Table 8 shows the Gini coefficient computed for the retweet distribution for each media outlet. This allows us to determine which are the outlets with heterogeneous audience (elpais_espana = 0.44, la_ser = 0.44, objetivolasexta = 0.46) and which ones with audience from only one party (abc_es = 0.76, cope_es = 0.79).

Table 8: Media indicators: party with highest retweet presence in the online audience of the outlet (Max Ret), party with highest reply presence (Max Rep), Gini coefficient of the outlet retweet (G_ret) and reply (G_rep) audience and ideology of the outlet (Spectrum: 1-10, left-right).

Media outlet	Max ret	Max rep	G ret	G rep	I
abc_es	PP	PP	0.76	0.49	7.31
antena3com	CS	Podemos	0.66	0.52	5.66
brujulaondacero	CS	CS	0.59	0.44	5.38
btvnoticies	Other	Other	0.68	0.64	3.65
cope_es	PP	PP	0.79	0.67	7.74
debatalrojovivo	Podemos	Podemos	0.52	0.55	3.3
elmonarac1	Other	Other	0.74	0.75	5.09
elpais_espana	PSOE	Podemos	0.44	0.44	4.74
la_ser	Podemos	Podemos	0.44	0.52	3.6
larazon_es	PP	PP	0.72	0.58	7.23
lasextatv	Podemos	Podemos	0.53	0.48	3.38
lasmananas_rne	PSOE	Podemos	0.68	0.49	5.49
lasprovincias	PP	Other	0.60	0.44	6.72
lavozdegalicia	PSOE	Podemos	0.57	0.30	4.88
levante_emv	Other	Podemos	0.51	0.37	4.79
libertaddigital	CS	PP	0.71	0.53	7.05
noticias_cuatro	Other	Podemos	0.56	0.41	4.64
objetivolasexta	CS	Podemos	0.46	0.44	4.45
ondacero_es	CS	Podemos	0.52	0.49	5.43
publico_es	Podemos	Podemos	0.56	0.50	2.71
sextanochetv	PSOE	Podemos	0.53	0.56	4.19
telecincoes	PP	Podemos	0.57	0.50	5.67
tve_tve	PP	PP	0.56	0.25	6.1

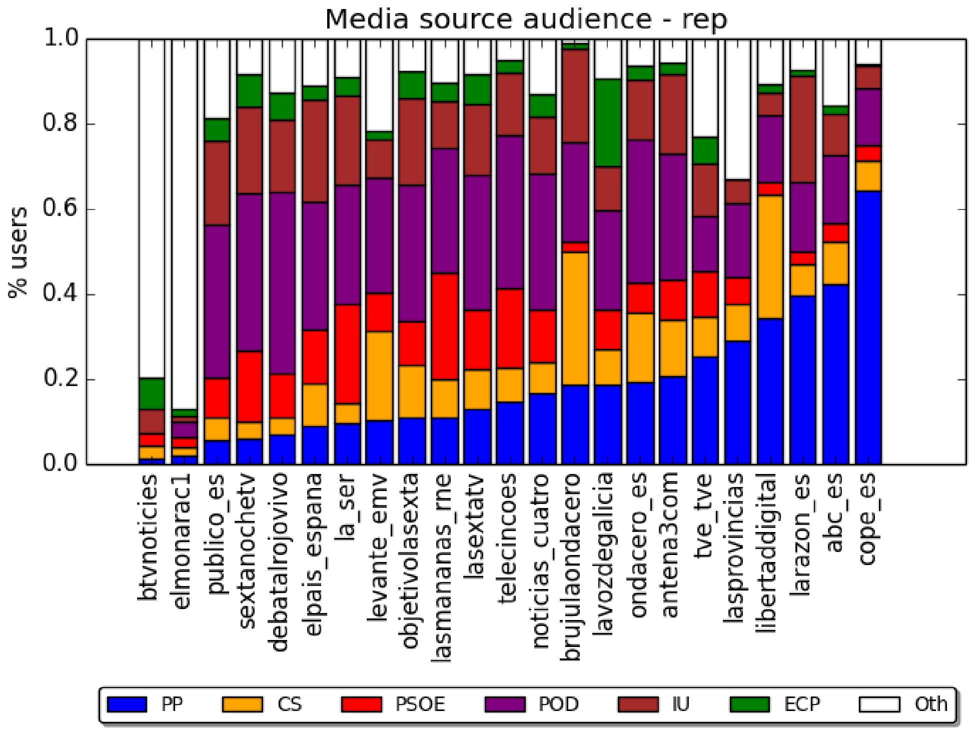


Figure 8: Distribution of replies done by party in each Spanish media outlet. The figure is sorted in the same way as the retweets one for a better comparison.

Second, differences are noticeable depending on the retweeting party. We observe that there are parties which seem to retweet more the mass media accounts than others. Figure 7 shows this effect visually, as at first sight, it is noticeable that there is more blue presence than purple and, more orange than red, for instance. Table 7 evidences this fact, showing in the first column the average percentage of the media audiences distributions shown in the figure. The average distribution shows that PP is the party that tends to retweet more the media outlets followed by CS and PSOE. IU and ECP are the ones retweeting less the media outlets. Finally, the regional media outlets (monrac1, btvnoticies, levante_emv) show a higher presence of Other users, as they are being retweeted by users which fell in regional party clusters (ie. ERC). The regionalism is also noticeable in ECP, which is appearing less in general as it is a regional small party and has a higher presence in regional outlets (ie. btvnoticies).

We then compare the reply distribution of media outlets in Figure 8 with the results obtained from retweets. First, we see a different general distribution visually (the parties with higher reply presence are different than the ones seen before). We can see that the PP (blue) replies mass media accounts less than retweets them and, the opposite happens with Podemos, which increments in replies. The second column in Table 7 shows the average for replies, which is way higher in Podemos and IU than previously. These results suggest that the Podemos and IU supporters may have a more critical position towards the Spanish mass media than PP, PSOE, and CS as they have more predisposition to reply them than to retweet. Analyzing the retweets distribution, it has been seen that the outlets distribution was very uneven and that the party supporters are positioned towards the media outlets. At replying time, the outlets seem to be more equally replied than retweeted. The Gini coefficients for the reply media outlets audience show that the reply audience is, in general, more heterogeneous than at retweeting time. We can observe that those (G rep) are lower than the retweet coefficients (G ret). This fact suggests that the media outlets seem to be replied either by their main party audience (defined by retweets distribution) and also their critics. Different orders for the figures explained previously are applied and displayed in Figures 14, 15, 16 and 16.

6.4.2 Ideology of a media outlet

Understanding the mass media audience is useful to label them with a political ideology. The goal is to find an indicator to situate a media outlet in the ideological political spectrum (left/right) using Twitter data. The results commented in the previous section are used to compute this indicator. The media outlets do not have online activity statistically significant to use it for the indicators. This is why the audience is used, as the media are widely retweeted and replied by the population and knowing which party supporters show what the media show can be significant to position the outlets in the political spectrum.

The ideology indicator is computed by the weighted arithmetic mean of the retweet audience distribution showed previously and the positions of each party in the po-

Table 9: For each party, situation in the political spectrum (extreme left - 1, extreme right - 10) provided by a CIS study from the citizenship.

Party	Ideology
PP	8.26
CS	6.37
PSOE	4.4
IU	2.46
POD	2.3
ECP	2.3

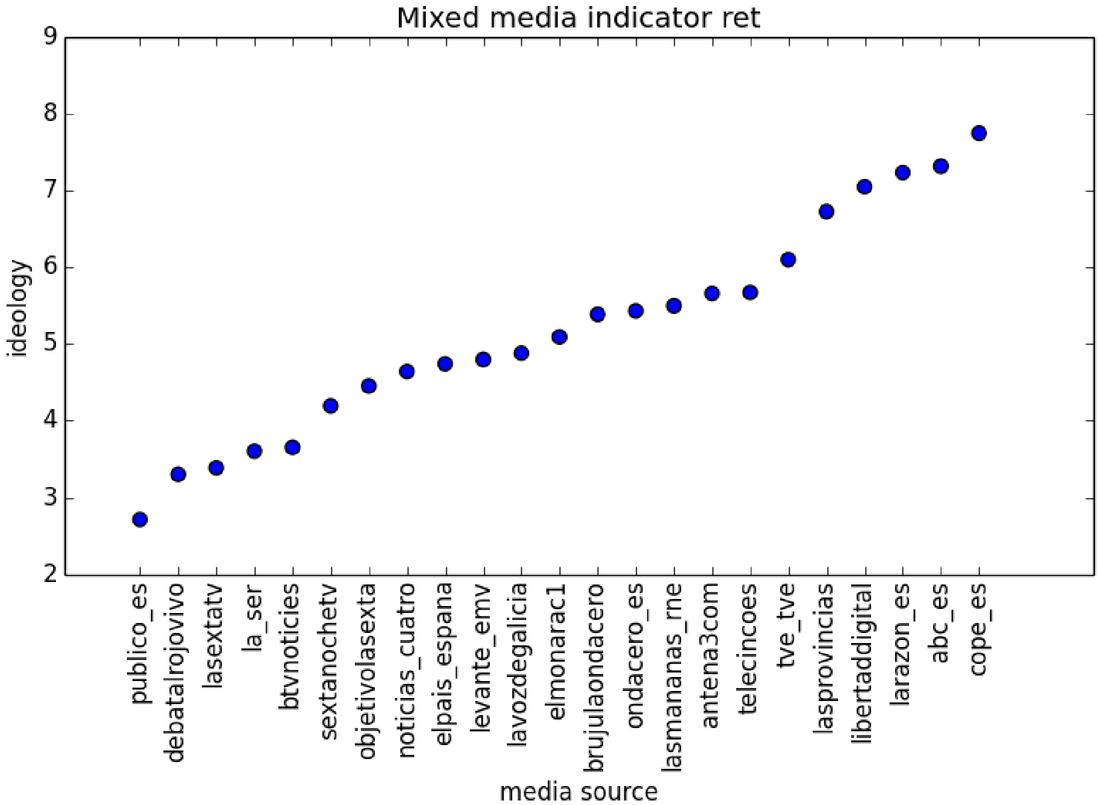


Figure 9: For each media, situation in the political spectrum (left - 1, right - 10) given by the weighted arithmetic mean.

litical spectrum provided in Table 9, which shows the ideological positioning that the citizens associate to each party based on CIS data of November 2015[1]. The valuation oscillates between 1 (extreme left) and 10 (extreme right).

Figure 9 shows all media outlets in the political spectrum sorted from left to right. The y axis belongs to the ideology indicator computed previously which can be also

[1]See http://ep00.epimg.net/descargables/2015/12/03/21679134b4464ad41e54d8042deb43a8.pdf

found in the last column in Table 8. The media outlets more situated to the left by this indicator would be publico_es, debatealrojovivo and lasextatv. On the other extreme of the plot and situated to the right we can find cope_es, abc_es, larazon_es and libertadigital.

The results obtained are comparable with the ones in [25]. In their case, the media whose readers are placed on the ideological right of the average result are Expansión, **abc_es**, **ondacero_es**, La Gaceta, El Confidencial y El Mundo. In their central area are media such as Europa Press, el Español, 20 Minutos o Voz Pópuli. The media whose readers are located on the ideological left are **elpais_espana**, El Periódico, El Huffington Post or El Plural, and in more left-hand positions are the headers of media such as eldiario.es, **Cadena SER**, **Público**, InfoLibre and CTXT. The mass media in bold indicates that those outlets correspond in to the same situation in our study. The study also differs from the outlets analyzed as some of the mentioned in their study are not present in this research.

6.5 Cluster evolution during the campaign (RQ4)

The evolution and cluster volatility during the 2016 campaign are analyzed in this section. Our intention is to determine if the evolution of the party clusters can be seen as an indicator of the loyalty on the parties. The methodology explained in 5.3 is applied to extract the volatility for the cluster parties during the whole 2016 campaign. The experiments have been performed with different values of w (2, 3, 4 and 5), the window size for the parameter applied in the methodology to get a smoother transition between days. However, results and, most important, proportions are kept for the different values and are only shown for w = 3.

Before computing the volatility and analyzing the results, it is important to comment the network and cluster sizes evolution as they play an important role in the

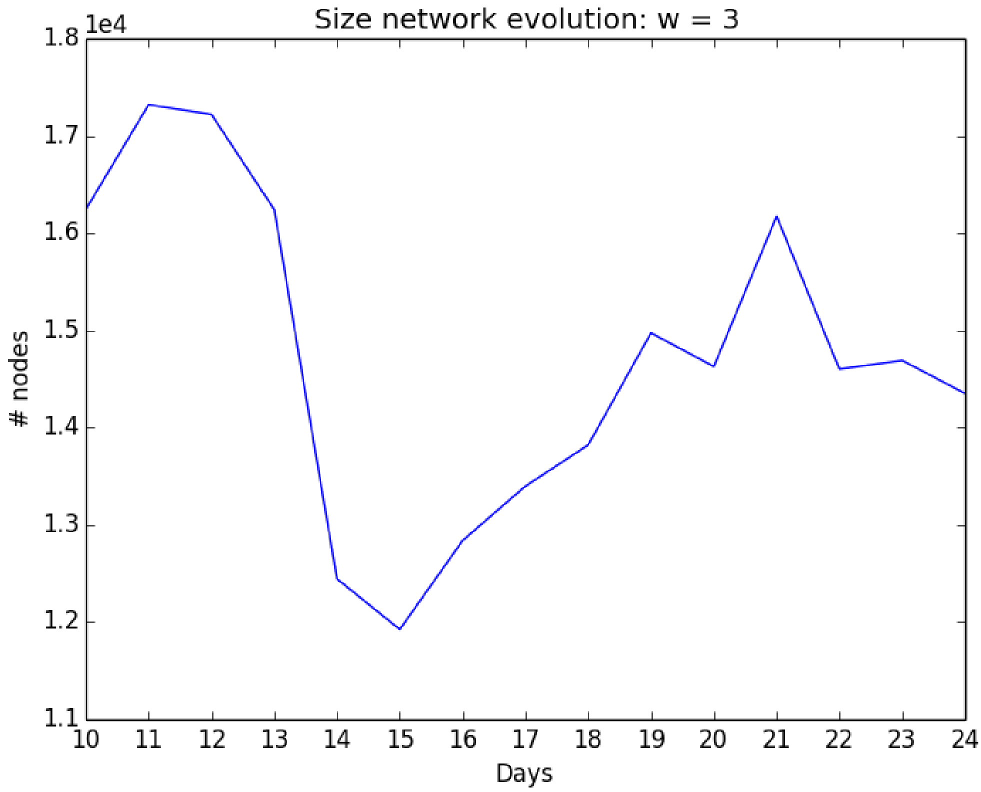


Figure 10: Network evolution with w = 3.

analysis. Figure 10 shows the evolution of the network during the campaign days. The data pattern is explained relatively easy: the first days, the networks grow as the campaign is starting and the people are highly motivated. Also, the national debate is held on the 13th and the data generated is contained days in between the 10th and the 15th in case of w = 3. That is the reason why we notice a big decrease on those days, as the days post debate the Twitter activity is starting to decrease, to finally grow again the days previous to the election day and decrease after it has happened.

As explained in the methodology, after the networks are extracted, the N-Louvain method is applied to obtain the daily clusters. Figure 11 shows, for example, the network and the final clusters obtained for the 15th day for w = 3. The daily clusters are matched across days and their sizes evolution can be extracted. We analyze now their sizes over days, which are showed normalized by the network sizes in Figure 12.

Figure 11: Network generated for day 15 with w = 3. The data comprised are re-tweets from 15 to 18 after applying an edge filtering of weight 3.

We can extract several conclusions on the clusters size evolution during the campaign days. Comparing the evolution with the absolute values obtained for 2016 in the previous section, we observe a similar proportion on sizes as in the overall 2016 clusters shown in Table 5. That is, IU has the highest representation, followed closer by Podemos and PP, PSOE and CS with around 10% - 15% each. Even though the proportions are maintained, the parties follow different patterns during the campaign: IU and PP show a similar pattern, as both seem to be the most affected by the debate. Even though the sizes are normalized by the network size and such a difference between days would not be expected, their clusters shrink

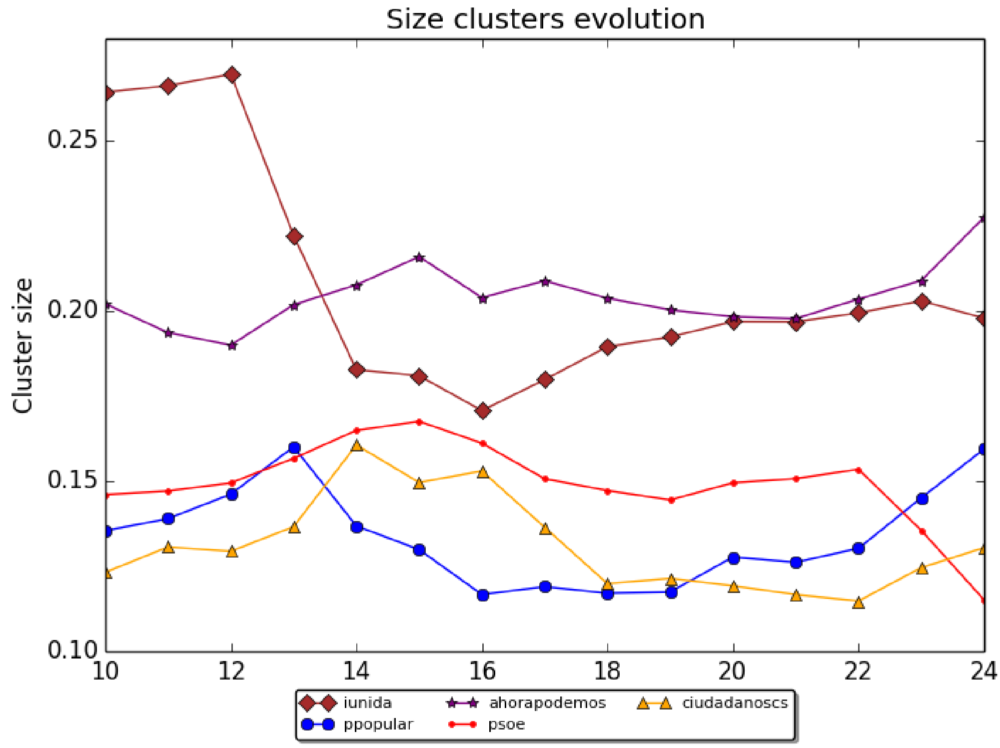


Figure 12: Clusters evolution with w = 3.

after the motivation from the debate and the first days, unlike PP, CS, and Podemos. PP, however, shows a high decrease in the last days from the campaign, which does not happen in IU. IU seems to be a party which started the campaign highly motivated but lost the motivation during the campaign. This is in line with the engagement section 6.3, where it was seen that IU seemed to be the party less motivated in the second campaign. While mostly all curves seem to reflex slightly the behaviors on the party voters (IU lost motivation, Podemos kept their votes, PP has a high increase in the last days in line with its victory), the opposite seems to happen with PSOE. A big loss was expected from the pre-electoral polls on this party, however, there were surprising results on election day as seen in 10 where PSOE shows the second most loyal electorate, only after PP. Although the cluster sizes evolution present tendencies towards the parties direction over time within the campaign, they are not enough to determine whether the clusters are very changeable over days or not. While the size of the cluster in two consecutive time stamps

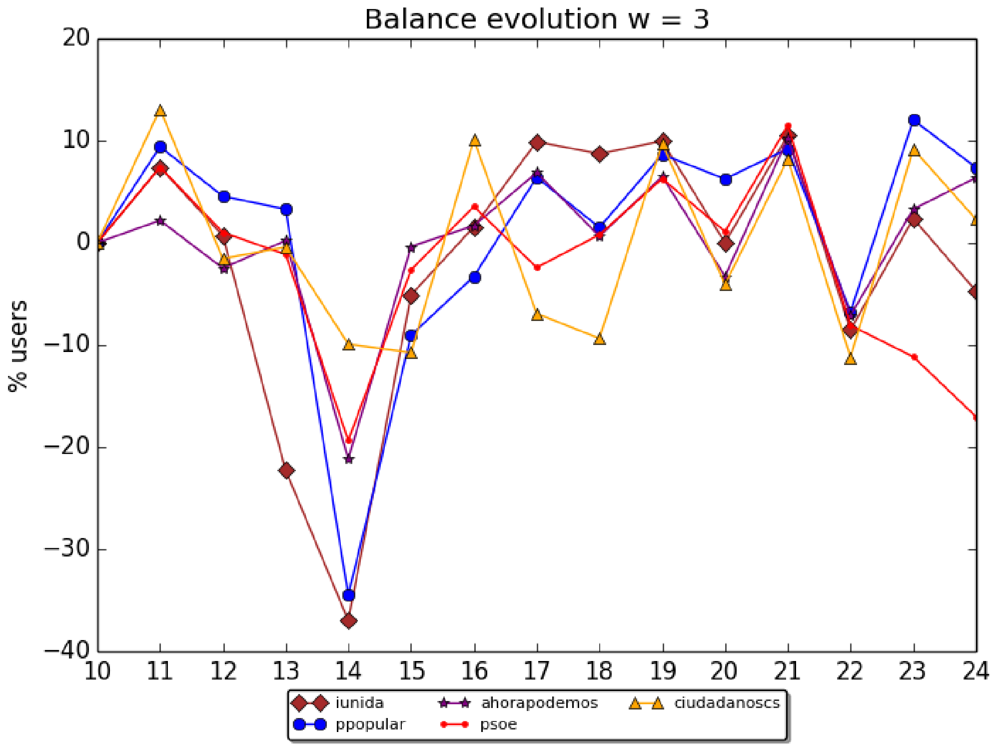


Figure 13: Cluster balance (new - lost) with w = 3.

can be the same, the nodes within them could have changed drastically. Figure 13 shows the daily balance: the amount of new users less the amount of lost users in the clusters normalized by the cluster size in the previous day. The balance allows us to determine which percentage of the cluster is different respect the previous day, either negative (more lost than new users) or positive (more new than lost users). The average percentage for each party is shown in Table 11 allowing us to determine which are the most volatile or changeable clusters.

Figure 13 shows that the balance, despite the days around the debate (13 to 15) where it reaches levels to more than the 30%, stays in between the 10% for all days. As the volatility average could have been affected by this special event, it has also been computing without considering those days, displayed in the second column in 11. The values of V in Table 11 show that PP has the highest positive value followed by Podemos. Although a high value means higher variability, the fact that is positive shows that in each time step there were more new users than lost users. That is, during campaign time, undecided voters have considered more the option to vote those parties and have been more loyal to them. Those results are in line with faithful voters extracted from post electoral polls and shown in Table 10. Contrary to those parties, we find PSOE and IU in the table to have the highest negative volatility. As we have seen in Table 10, PSOE showed more loyalty than was expected from the polls. The value provided by the methodology would not capture the online world behaviors in this case. Table 11 also shows an alternative volatility measure (V*) which has been computed without considering the 13 to 15 days values. As almost all clusters had a considerably negative balance on those days, the average gets increase by the values exclusion. However, the proportions are maintained in almost all parties. The main difference relies on IU, which has a negative value of V but a positive V*. The days after the debate show a superior decrease for this party than for the rest and removing those days has significantly changed the results. It is interesting to mention that Podemos and IU, have a similar value V* but it is positive for the first party in V and negative for the second one. This can be an indication that the debate could have affected very much the decision of IU to not vote the coalition as they were not appealed by the leader of the big party.

Table 10: Percentage of faithful voters by party provided by a post-electoral study done by CIS [15].

Party	% faith. voters
PP	86%
CS	59%
PSOE	76%
IU	59%
POD	73%

Table 11: Clusters volatility for each party. V is the average of the % balance (new - lost users) during the whole campaign. V considers all days and V* does not consider days 13, 14 and 15.

Cluster	V (w = 3)	V * (w = 3)
PP	1.46	5.18
POD	0.64	2.58
CS	0.04	1.81
IU	-2.09	2.74
PSOE	-3.37	-2.29

Chapter 7

Discussion

We have presented a methodology to analyze online Twitter campaigns based on several steps. First, we have used a robust community discovery method and matched automatically the user clusters across multiple executions of the Louvain method using the Jaccard coefficient. Second, we have proposed a characterization of the cluster composition dynamics in consecutive elections to reflect changes in party inclinations. Finally, we have analyzed political engagement by means of the Twitter activity distributions in the different clusters. Our proposed methodology can be seen as an improvement on similar approaches proposed recently for the analysis of online Twitter campaigns [17]. We have applied this methodology to social network data extracted from campaign related user retweeting activity during the 2015 and 2016 Spanish National elections. We find that the parties which joined in a coalition after the 2015 elections kept their separate online structures and did not form a unique online cluster. However, the interactions between the clusters in the coalition grew suggesting that their supporters got closer in the second election.

The analysis of user migration between party clusters (RQ1) reveals that several users have transitioned within the coalition. The results expose an important transfer of users associated to Podemos in the 2015 election to the cluster of Izquierda Unida in 2016. Those users might have been supporting the bigger party in the first election as a matter of utility when it came to getting representatives while actually feeling closer to the smaller party. The results also show a smaller proportion of users who remain in the UP clusters in 2016 compared to other parties, which may reflect the demotivation of its electorate, although this signal is weak compared to the large decrease in votes for UP.

Previous research has indicated how Twitter activity may be thought as an indicator of the political engagement of the users [20, 21, 19]. Our study has also analyzed whether there is a relation between the motivation of the electorate and activity on Twitter (RQ2). Despite our analysis shows a lower activity in 2016 than in 2015 for all mayor parties, in line with the participation fall, the results follow a very similar pattern for all parties although the electoral results were different for all of them. Moreover, the decrease in activity is not significantly higher for the users in the UP clusters, which lost the highest amount of votes. However, our analysis revealed differences within the UP clusters, showing a much larger decay in activity for IU supporters. This may indicate that users strongly associated to IU were less appealed by the coalition, in agreement with existing studies [13, 16].

The research also analyzed a set of Spanish mass media accounts (RQ3) to understand their Twitter (retweet and reply) audiences and determine behavioral differences of the party supporters and situate them in the left-right political spectrum. First, it has been seen two different media sources types when it came to their retweeting audience: singled party sources, which are retweeted by a predominant party (ie. cope_es, abc_es); and heterogeneous sources, which have more varied audience (ie. elpais_espana). The Gini coefficient applied to the audiences allowed to rank the media sources in terms of equality on their audience. Second, behavioral differences are obtained when analyzing the retweet and reply audiences. The results showed a tendency to retweet mass media on the right parties (PP and CS) while the left ones (Podemos and IU) reply them more. This last fact suggests that the left parties are more critics with the Spanish mass media nowadays. Finally, we used the weighted mean to compute an ideology indicator to situate each media source in the Spanish ideological spectrum. Our methodology differed from the proposed in [25] as it used community detection on the retweet network instead of polls within Twitter. However, the results obtained are similar and the position of the media sources does not differ much from the other study.

The last part of the study was focused in determining which parties had the most volatile electorate and which ones counted with the most faithful based on the 2016 campaign data (RQ4). We developed a methodology that uses a sliding window technique to build different daily networks, applies the N-Louvain method to the networks to extract the daily clusters and matched them across days to draw the size evolution over the campaign. Our approach differed from existing studies [38, 37, 36, 39], which also use the two step approach, including the N-Louvain as community detection method, the sliding window for the network generation in order to provide less abrupt changes in the communities and the Jaccard cofficient to match them across days. The clusters evolution showed the motivation of the parties during the first days of the campaign followed by a decrease of size after the debate was held. Finally, some of the parties showed a size increase in line with the final campaign days. The evolution shows also the demotivation of IU, which was also seen in the previous sections. The methodology has also proposed a metric, the cluster volatility (V), in order to relate it with the options that were most likely to have been explored by the undecided voters during campaign time. The results showed that PP has been the most positive value followed by Podemos and that would have meant that, during campaign time, undecided voters have considered more the option to vote those parties and have been more loyal to them. PSOE and IU have the most negative values, being in line with existing studies in the coalition party but not in the first one, which gave a surprise in the election day showing more loyalty from their voters than the expected. The alternative computation of the volatility without considering the after debate days indicated that the event could have had a negative impact on the IU undecided voters to not vote the coalition as they were not appealed by the performance of the leader of the big party in the coalition.

Chapter 8

Conclusion

This study has presented four research questions motivated by the current Spanish political situation and the unique scenario of having two elections in the short period of six months.

First, it questioned whether the Twitter activity reflected the migration of supporters between parties from the first to the second election. The question was answered applying to the retweet networks a new community detection method based on Louvain and comparing the clusters obtained in both elections. The results showed that the methodology applied reflected the results from the elections.

Second, it asked if the Twitter activity was a reflection on the demotivation of the citizenship after a long period of political movement. Our analysis showed a lower activity in 2016 than in 2015 for all mayor parties, in line with the participation fall and revealed a decrease of the motivation in the IU supporters, also in line with post electoral polls.

Third, we situated in the ideological spectrum a set of Spanish media sources using their retweet audiences and the citizen ideological perception of the parties obtaining similar results that a previous study.

Finally, our study questioned the relationship between the clusters volatility over time with the loyalty and volatility of the electorate. A new methodology is presented to extract the evolution and volatility of the clusters over the 2016 campaign.

The results were partially in line with the polls and the election outcome. Although the methodology did not capture as expected the results seen from the off-line world, it is interesting to be applied in larger time-frames to see the evolution from another perspective. The methodology proposed is also valuable as it can be applied, not only in electoral campaigns and political studies, but in several other fields where evolution and variability are the main focus.

Chapter 9

Future work

The study presented can be extended in the following ways: First and most important, from the community detection part, the threshold chosen to filter the retweet networks could have influenced the results from all parts exposed in the report. The experiment should be repeated with different thresholds to see if the there are significant changes. As the clustering is the basis of the report, all the following experiments should be repeated iteratively with the results of the clustering done with all thresholds. The results should be compared to see if the different iterations differ from one to another in a large scale.

The method to understand the dynamics between elections method can be applied to all consecutive elections, that means, it would be interesting to follow this study taking data from the elections which are going to be held in the future. Also, the elections held in 2011 could have been used to complement the study and do a comparison of the clusters obtained previously to the 15M as well an analysis on where did the electorate come prior the emergence of the new parties. However, the dataset resulting from those elections was not big enough to perform a significant comparison and, furthermore, the data was captured in a very different manner than the other two, which would make the comparison not applicable.

The media sources study differed from the initial idea to what was finally done. The lack of activity that the media sources post in Twitter changed the original idea of focusing the research in the media coverages to their audiences. However, it would be still important to find an indicator to determine the pluralism/equality in their political coverage. In this study, the mentions to leaders or political words written by the official media accounts were captured and equality metrics were applied to the data. Nevertheless, the lack of data was not enough to determine any conclusion. A longer time-frame of data capturing could be something worth to try to get more data to analyze. Also, an analysis focused using natural language processing techniques and more content related can be a better approach to this research line.

Temporal network techniques have been applied to the second campaign in order to understand their volatility. A method has been developed for this reason using the days as time-frame. However, this technique can be applied to different datasets, using different window sizes and time-frames to study the evolution of networks, not only in politics but in varied fields where network analysis can be studied. In our study, the method has been applied only to the second campaign data. As seen from the polls and from several previous studies, around a 30% of undecided made the decision on who to vote in campaign time. However, the rest made their choices before the campaign, even one year prior to the election. For further studies, data related to the parties can be gathered in a larger time frame in order to study the volatility and evolution during a longer period of time.

List of Figures

List of Tables

Bibliography

[1] Castells, M. The new public sphere: Global civil society, communication networks, and global governance. *The aNNalS of the american academy of Political and Social Science* **616**, 78–93 (2008).

[2] Metaxas, P. T. & Mustafaraj, E. Social media and the elections. *Science* **338**, 472–473 (2012).

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

[4] Vergeer, M. & Hermans, L. Campaigning on twitter: Microblogging and online social networking as campaign tools in the 2010 general elections in the netherlands. *Journal of Computer-Mediated Communication* **18**, 399–419 (2013).

[5] Caldarelli, G. *et al.* A multi-level geographical study of italian political elections from twitter data. *PloS one* **9**, e95809 (2014).

[6] Tumasjan, A., Sprenger, T. O., Sandner, P. G. & Welpe, I. M. Predicting elections with twitter: What 140 characters reveal about political sentiment. *ICWSM* **10**, 178–185 (2010).

[7] Chung, J. E. & Mustafaraj, E. Can collective sentiment expressed on twitter predict political elections? In *AAAI*, vol. 11, 1770–1771 (2011).

[8] Gayo-Avello, D. No, you cannot predict elections with twitter. *IEEE Internet Computing* **16**, 91–94 (2012).

[9] Jungherr, A., Jürgens, P. & Schoen, H. Why the pirate party won the german election of 2009 or the trouble with predictions: A response to tumasjan, a., sprenger, to, sander, pg, & welpe, im "predicting elections with twitter: What 140 characters reveal about political sentiment". *Social science computer review* **30**, 229–234 (2012).

[10] Junghferr, A. Twitter use in election campaigns: A systematic literature review. *Journal of information technology & politics* **13**, 72–91 (2016).

[11] Wikipedia. Anti-austerity movement in spain — wikipedia, the free encyclopedia (2017). URL `https://en.wikipedia.org/w/index.php?title=Anti-austerity_movement_in_Spain&oldid=784324152`. [Online; accessed 14-June-2017].

[12] Orriols, L. & Cordero, G. The breakdown of the spanish two-party system: the upsurge of podemos and ciudadanos in the 2015 general election. *South European Society and Politics* 1–24 (2016).

[13] Simón, P. The challenges of the new spanish multipartism: government formation failure and the 2016 general election. *South European Society and Politics* **21**, 493–517 (2016).

[14] A.B.C. Rajoy amplía su victoria y reclama su derecho a gobernar (2016). URL `http://www.abc.es/elecciones/elecciones-generales/abci-resultados-elecciones-generales-pp-amplia-victoria-y-psoe-saca-15-escanos-noticia.html`.

[15] Llaneras, K. Qué votantes cambiaron su voto el 26-j (2016). URL `http://politica.elpais.com/politica/2016/07/22/ratio/1469195845_977293.html`. [Online; accessed 10-June-2017].

[16] Ferrándiz, J. P. Fidelidades y fugas para explicar los resultados del 26j (2016). URL `http://metroscopia.org/fidelidades-y-fugas-para-explicar-los-resultados-del-26j/`. [Online; accessed 10-June-2017].

[17] Aragón, P., Volkovich, Y., Laniado, D. & Kaltenbrunner, A. When a Movement Becomes a Party: Computational Assessment of New Forms of Political Organization in Social Media. In *ICWSM-16 - 10th International AAAI Conference on Web and Social Media* (The AAAI Press, 2016).

[18] Conover, M. *et al.* Political polarization on twitter. In *ICWSM* (2011).

[19] Holt, K., Shehata, A., Strömbäck, J. & Ljungberg, E. Age and the effects of news media attention and social media use on political interest and participation: Do social media function as leveller? *European Journal of Communication* **28**, 19–34 (2013).

[20] Bakker, T. P. & De Vreese, C. H. Good news for the future? young people, internet use, and political participation. *Communication research* **38**, 451–470 (2011).

[21] Dimitrova, D. V., Shehata, A., Strömbäck, J. & Nord, L. W. The effects of digital media on political knowledge and participation in election campaigns: Evidence from panel data. *Communication Research* **41**, 95–118 (2014).

[22] Reuters. Reuters institute digital rews report 2015 (2015). URL `https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Reuters%20Institute%20Digital%20News%20Report%202015_Full%20Report.pdf`. [Online; accessed 14-Julio-2017].

[23] Herman, E. S. & Chomsky, N. *Manufacturing consent: The political economy of the mass media* (Random House, 2010).

[24] Anonymous. Quién manda en los medios de comunicación en españa (2011). URL `http://www.solosequenosenada.com/2011/11/18/quien-manda-en-los-medios-de-comunicacion-en-espana/`. [Online; accessed 14-Julio-2017].

[25] Pujol, F. El perfil ideologico de los lectores de prensa. anÁlisis encuestas (2015). URL `https://smreputationmetrics.wordpress.com/2015/12/09/`

el-perfil-ideologico-de-los-lectores-de-prensa-analisis-encuestas-7deldebatedec
[Online; accessed 14-Julio-2017].

[26] Hayes, B. C. & McAllister, I. Marketing politics to voters: late deciders in the 1992 british election. *European Journal of Marketing* **30**, 127–139 (1996).

[27] van der Meer, T. W., van Elsas, E., Lubbe, R. & van der Brug, W. Are volatile voters erratic, whimsical or seriously picky? a panel study of 58 waves into the nature of electoral volatility (the netherlands 2006–2010). *Party Politics* **21**, 100–114 (2015).

[28] Arcuri, L., Castelli, L., Galdi, S., Zogmaister, C. & Amadori, A. Predicting the vote: Implicit attitudes as predictors of the future behavior of decided and undecided voters **29**, 369 – 387 (2008).

[29] Conroy, M., Feezell, J. & Guerrero, M. Facebook and political engagement: A study of online political group membership and offline political engagement **28**, 1535–1546 (2012).

[30] Gibson, R. & Cantijoch, M. Conceptualizing and measuring participation in the age of the internet: Is online political engagement really different to offline? *The Journal of Politics* **75**, 701–716 (2013). URL https://doi.org/10.1017/S0022381613000431. https://doi.org/10.1017/S0022381613000431.

[31] Anduiza, E., Jensen, M. J. & Jorba, L. *Digital Media and Political Engagement Worldwide: A Comparative Study* (Cambridge University Press, New York, NY, USA, 2012), 1st edn.

[32] Boulianne, S. Does internet use affect engagement? a meta-analysis of research. *Political Communication* **26**, 193–211 (2009). URL http://dx.doi.org/10.1080/10584600902854363. http://dx.doi.org/10.1080/10584600902854363.

[33] Chadwick, A. & Howard, P. *Routledge Handbook of Internet Politics*. Routledge handbooks (Routledge, 2010). URL https://books.google.es/books?id=GJdfuGSa1xUC.

[34] Kim, J.-Y. The impact of internet use patterns on political engagement: A focus on online deliberation and virtual social capital. *Info. Pol.* **11**, 35–49 (2006). URL `http://dl.acm.org/citation.cfm?id=1412560.1412565`.

[35] DePaula, N. & Dincelli, E. An empirical analysis of local government social media communication: Models of e-government interactivity and public relations. In *Proceedings of the 17th International Digital Government Research Conference on Digital Government Research*, dg.o '16, 348–356 (ACM, New York, NY, USA, 2016). URL `http://doi.acm.org/10.1145/2912160.2912174`.

[36] Palla, G., Barabási, A.-L. & Vicsek, T. Quantifying social group evolution. *arXiv preprint arXiv:0704.0744* (2007).

[37] Falkowski, T., Bartelheimer, J. & Spiliopoulou, M. Mining and visualizing the evolution of subgroups in social networks. In *2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings)(WI'06)*, 52–58 (2006).

[38] Greene, D., Doyle, D. & Cunningham, P. Tracking the evolution of communities in dynamic social networks. In *2010 International Conference on Advances in Social Networks Analysis and Mining*, 176–183 (2010).

[39] Kibanov, M., Atzmueller, M., Scholz, C. & Stumme, G. Temporal evolution of contacts and communities in networks of face-to-face human interactions. *Science China Information Sciences* **57**, 1–17 (2014). URL `https://doi.org/10.1007/s11432-014-5067-y`.

[40] Chen, Z., Wilson, K. A., Jin, Y., Hendrix, W. & Samatova, N. F. Detecting and tracking community dynamics in evolutionary networks. In *2010 IEEE International Conference on Data Mining Workshops*, 318–327 (2010).

[41] Lin, Y.-R., Chi, Y., Zhu, S., Sundaram, H. & Tseng, B. L. Facetnet: A framework for analyzing communities and their evolutions in dynamic networks. In *Proceedings of the 17th International Conference on World Wide*

Web, WWW '08, 685–694 (ACM, New York, NY, USA, 2008). URL http://doi.acm.org/10.1145/1367497.1367590.

[42] Ferlez, J., Faloutsos, C., Leskovec, J., Mladenic, D. & Grobelnik, M. Monitoring network evolution using mdl. In *2008 IEEE 24th International Conference on Data Engineering*, 1328–1330 (2008).

[43] Berlingerio, M., Bonchi, F., Bringmann, B. & Gionis, A. *Mining Graph Evolution Rules*, 115–130 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2009). URL https://doi.org/10.1007/978-3-642-04180-8_25.

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

[45] Jaccard, P. *Etude comparative de la distribution florale dans une portion des Alpes et du Jura* (Impr. Corbaz, 1901).

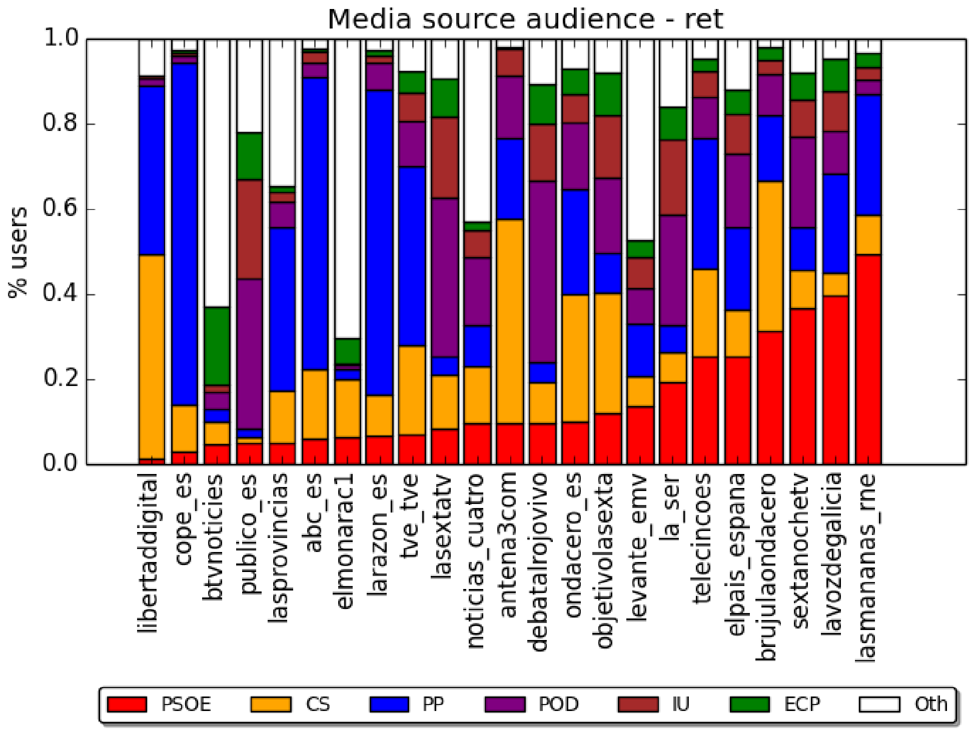


Figure 14: Distribution of retweets done by party in each Spanish media. The figure is sorted by amount of retweets done by PSOE.

Appendix A

First Appendix

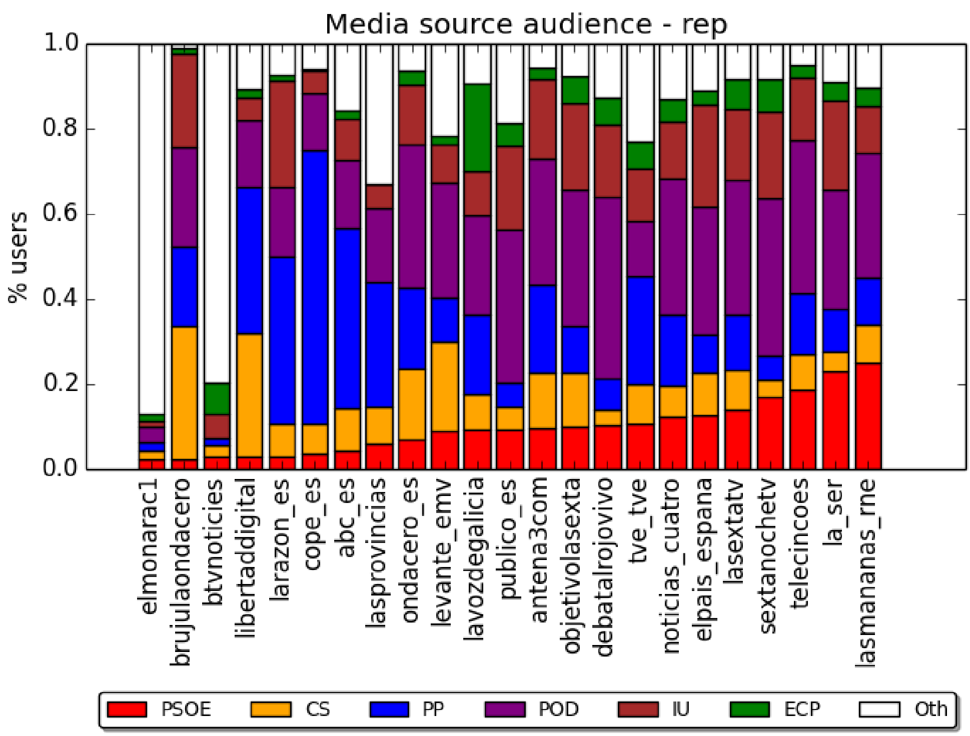


Figure 15: Distribution of replies done by party in each Spanish media. The figure is sorted by amount of replies done by PSOE.

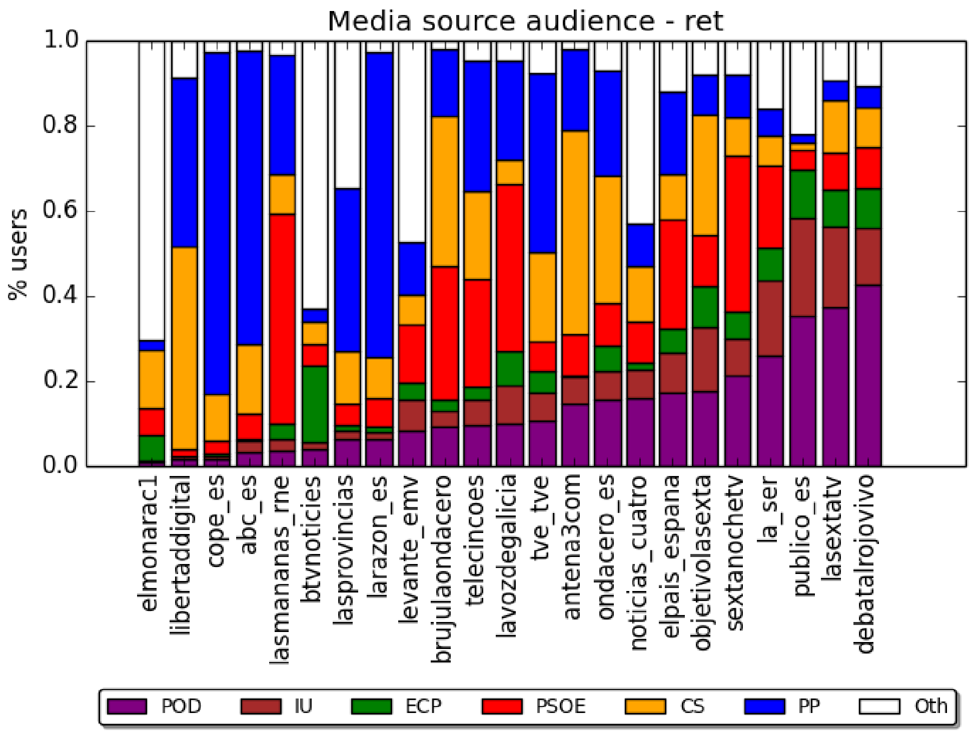


Figure 16: Distribution of retweets done by party in each Spanish media. The figure is sorted by amount of retweets done by Podemos.

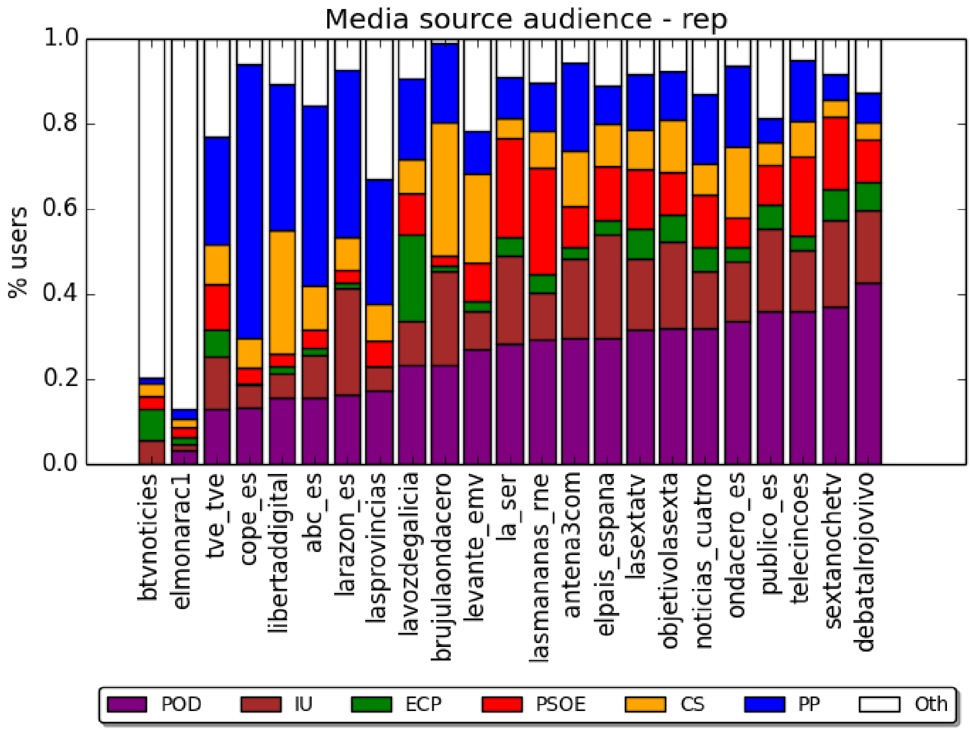


Figure 17: Distribution of replies done by party in each Spanish media. The figure is sorted by amount of replies done by Podemos.