Patterns of online mobilization on Twitter: an analysis of the #OcupaMinC protest

Marcelo Leal

Introduction

Since the beginning of the Internet, social movements have been using it to plan and coordinate protests around the world. E-mails, listservs, forums, and temporary websites increased the connectedness of activists during the first generation of the web and are still used in a daily basis. However, the advent of the Web 2.0 during the 2000s has produced bigger changes for activism and protest. The new focus on user-generated content and collaborative consumption platforms has increased the participation of ordinary people, as well as the rate of communication flows, in online mobilizations. Social networking services (SNS), as Twitter and Facebook, have produced a space where individuals with different identities can interact with each other and collectively act for a common cause. In spite of this trend and the academic interest on the relationship between SNS and online mobilization, social science research has found a variety of explanations about the effects of these platforms on online dynamics of recruitment and diffusion in political action. Thus, the aim of this paper is to make a contribution to the current debate about the implications of the use of social media through the analysis of the #OcupaMinC protest on Twitter.

The #OcupaMinC was an online protest against the dissolution of the Brazilian Ministry of Culture (MinC) made by the vice-president Michel Temer when he assumed the president’s office. The protesters occupied buildings of the Ministry’s regional branches and had a lot of media attention, both because the protest had the participation of many famous artists and because it was a strong reaction against the interim
government after the Brazilian Senate accepted the charges against the former president Dilma Roussef. In a week, the occupations were scattered across the country and Temer was forced to reinstate the Ministry using a presidential decree. But despite the offline occupations, the protesters also mobilized on social media, even before the date of the Ministry’s dissolution. Indeed, when the first information about the political reform of the new government appeared on the news, some days before the official release, some Twitter users created the #FicaMinC (#KeepMinC) hashtag to criticize the supposed extinction of the Ministry. As the offline mobilization took place, the online protest also adapted itself and began to use the #OcupaMinC (#OccupyMinC) hashtag, along with hashtags about regional occupations (usually the original hashtag followed by the state postal code) or about the occupation of specific branches of the Ministry, as #OcupaIphan and #OcupaFunarte (related to, respectively, the occupations of the National Historic and Artistic Heritage Institute and the National Foundation of Art).

The theoretical framework used in our research includes major theories about activism, participation and protest in the Social Sciences, which have emphasized the importance of social networks and shared consciousness in the making of political mobilizations, and recent studies about the Twitter and Facebook “revolutions” of the last years, which have underlined the effects that SNS had in the engagement of individuals and in the construction of collective identities in those movements. According to this scholarship, SNS play a central role in political mobilization because they reduce the costs of participation by providing a space where individuals can easily connect with each other and build common narratives about it, although at the cost of loosening the movement identity (Bennett and Segerberg 2012). Nonetheless, these analyses have
found different dynamics of recruitment of new members and distinct dynamics of diffusion of information among these members when analyzing diverse movements in different countries. Therefore, our research draws on previous studies about dynamics of recruitment and of diffusion on Twitter\(^1\) to test the following hypothesis:

\[
\text{H1: The recruitment of new protesters is dependent on the activation times of central users in the network.}
\]

\[
\text{H2: The diffusion of information is dependent on the participation of central users in chains of information.}
\]

In order to test them, we extracted the related data from Twitter according to the main hashtags\(^2\) used in the protest and used social network analysis to explore it. In the following sections we introduce a brief theoretical background, the methodology used in the data gathering, treatment, and analysis, and the discussion of the results of the data analysis.

**Theoretical background**

Networks are central categories in the study of political behavior, mainly the ones related to social cooperation (Cranmer et al. 2016). Indeed, social networks have been one of the major explanatory variables of political mobilization and protest, as seen in many theories in the social sciences, as the Resource Mobilization Theory and the Political Process Model (McAdam, Tarrow and Tilly 2011; Tarrow 2011; Tilly and


\(^2\) See Appendix 1 for the full list of hashtags used in the search.
Wood 2009). According to these approaches, the existence of an interpersonal network is a necessary (but not sufficient) condition to ordinary people with different identities to act collectively. It is the connective structure provided by these networks that permits the aggregation and coordination of actors, due to the participation opportunities and the solidarity incentives that it provides.

In this scenario, social networking services play a central role in political mobilization by providing a space where individuals can easily connect with each other and build common narratives about the real world, independently of their geographical location. They allow the creation of personalized messages by their users and enable the diffusion of these messages inside the user’s network, which makes them tools for interaction by different groups of people (Bennett and Segerberg 2012; Bimber, Stohl and Flanagan, 2008) and for collective identity construction (Gerbaudo and Treté 2015) inside these groups. And due to these features, some social scientists have stated that organizational and communication constraints are reduced (or even eliminated) with the advent of the Internet (Bennett and Toft 2008; Bimber, Stohl and Flanagan, 2008).

Nonetheless, some researchers have pointed out that the increased use of SNS has not been adequately addressed by the social sciences because of two major limitations: on the one hand, most of the researches have focused on webpages interconnections and on the count of the most used words on them, with few studies about SNS and their uses by social movements (Hanna 2013); on the other hand, among the limited analyses that have SNS as their main object of study, few of them have looked to the internal dynamics of the online networks of political mobilizations (Theocharis 2013). When we look for analysis about the use of Twitter in protests, Theocharis (2013) points out that:
There are few data about how networks of activists connected through Twitter function, or the consequences of removing key actors from the network. That is, we do not fully understand the internal mechanisms and dynamics of activists’ Twitter connections, which allow them to collectively share and distribute important information. Moreover, when these networks are formed, we have little knowledge of how robust they are, which network structures best facilitate the distribution of information to actors, whether networks depend on a subset of actors to distribute information and organize or coordinate protest action, and the consequences of cutting off well-connected actors from the network. (36)

Bennett and Toft (2008) and Poell (2014) also underlie the important of analyzing the digital platforms that provide the mean through where the movement narratives are distributed. Their main argument is that the narrative sharing among individuals is an important part of generating social networks and that the Internet is central in this process because it fosters weak ties through loosely narratives. Bearing this in mind, Bennett and Segerberg (2008) argue that a new social logic is at play: the logic of collective action has become the logic of connective action. The main characteristics of this logic is that networks operate through organizational processes of social media and they do not require strong organizational control or the symbolic construction of a united ‘we’. They state that the communication plays a role as an organizing principle in this logic. Most important: they argue that communication networks become the organizational form of the political action.

Juris (2014) also points out that a new logic is generated by the new SNS: the logic of aggregation. This logic is also centered on user-generated content platforms and it is characterized by viral flows of information and the agglomeration of individuals in concrete spaces, which are vital for the transformation of “smart mobs” in physical embodied movements. This is the result of the characteristic of SNS: they provide a
greater audience reach and are more widely used when compared to first generation web applications, they create a sense of connectedness and solidarity among the members of the movement, and they provide lower barriers of access and participation. Moreover, they have broadcasting features that permit the quick diffusion of messages and the emergence of the viral flows of information. Nonetheless, the aggregation that emerge from this logic are under constant threat of disaggregation due to the differences in the collective subjectivity of the people mobilized, which are not solved under a unified set of demands.

However, critics of the connective action and aggregate logics state that usually these frameworks underestimate the weight of collective identity construction in social movements and protests. According to the critics, the main question around narratives generated in online-based political mobilizations should not be about how loosely or specific they are; instead, it should focus on the process of identity construction and how the characteristics of the Internet shape it (Gerbaudo and Treté 2015). They agree that the diverse technologies and platforms that provide Internet services have enabled a much more dynamic flow of information and that this information is much more transitory due to the fluidity of information in the digital era. Nonetheless, they argue that homophily patterns are the driven-force of aggregation in social media, which results in specific dynamics of information diffusion. As social networks in SNS are fostered by the tendency to associate with like-minded actors, information sharing is partially determined by the number of friends sharing the same content (Brooking and Singer 2016)³.

³This approach produces different results than the one presented by Juris (2014), since his findings seems to support the idea that the reach of shared messages in SNS during political mobilizations goes far beyond a user network, while the homophily patterns approach seems to suggest the contrary.
Besides the number of mobilized friends, Margetts et al (2016) argue that low threshold users – the “starters” of an online political mobilization – are responsible for the diffusion of information about a movement and, consequently, by the engagement of new members that have low costs to participate but higher threshold. According to them, the engagement of new members in an online mobilization has a S-shaped curve, following a Schelling’s threshold approach, and the reduced costs of mobilization results in a reduced need of collective trust and smaller group cohesion. Thus, their findings show that the online mobilizations have their power distributed to the periphery of the network, with the ‘starters’ playing a major role.

In a similar way, Gonzáles-Bailon et al (2011) demonstrate through an analysis of the 15-M movement on Twitter that all users, independently of their position in the network, are important players in the recruitment of new members, but only at the local level. That means that the dynamics of recruitment were not related to any specific topology of the social network. Indeed, small clusters inside the network explain better the engagement of new members in the online mobilization through the interaction of low threshold activists. The information cascading, on the other hand, is related to the centrality of the users in the network, which means that few flows of information reached a huge population in SNPs during the 15-M protests; the ones that succeed in doing so were usually boosted by central members of the network (i.e., people with a great number of followers).

However, the findings of Theocharis (2013), who has analyzed the relationship between the actors of the 2010 United Kingdom school mobilizations on Twitter, point to a different direction. His conclusions show a decentralized, but robust, network, in which
the central positions of the actors do not have a great impact in the information flow. Thus, the suspension of a central account of the network on Twitter would not have affected the communication among the actor in a significant way, as the flow of information would easily adapt to it. Nonetheless, as pointed out by the author (52), this does not mean that every online network, and even every network inside this SNS, will have the same structure. Therefore, social scientists should keep on analyzing different online networks in order to find patterns that help to explain how different types of network structures affects different political mobilizations in the Internet.

Method

We analyzed the Twitter activity related to the protests against the dissolution of the Brazilian Ministry of Culture during the period of May 10 (when the first news about the presidential reform were released) to September 19 (when the presidential decree\(^4\) reinstating the Ministry was approved by the Brazilian Congress). For each tweet, beyond its number of identification and message, the data set contains the date and hour that it was posted. For each account, we have also the list of followers and friends of each one. The user’s information was collected through the Twitter’s API (application programming interface), but the data gathering process of the tweets was a little more complicated because the official API restricts the period of tweets collection to just seven days prior to the day of the search. A way to circumvent this problem was to use the advanced search of the Twitter website and to collect the webpage source code. However, as the advanced search results need to be constantly updated (Twitter uses the infinite

---

\(^4\) According to the Brazilian legislation, presidential decrees have the force of law when they are issued, but they need to be approved by the Congress in 120 days; otherwise, they are revoked.
scrolling feature to show them all), and the size of the source code in the most active days of the protests usually crashed web browsers, we updated an open-source Python script available on GitHub\(^5\). The original script automatized the tweet gathering process, but due to Twitter ability of detecting automated code, we modified it in order to make requests to the webpage in randomized periods, acting more as a human user would do. It was also modified to add a new function to save the data in csv and json format, for posterior manipulation.

After the compilation of the data, we used social network analysis techniques to analyze the relationships and the information flow inside the network. First, we classified each user according to its activation time in the network, i.e., the first time the user sent a message about the protest, and the activation time of his neighbors. When we compare both activation times, we can categorize each user according to its thresholds levels to participate in the online mobilization. In other words, we calculated the proportion of a user’s neighbors who were activated before him in order to say whether the user has a low activation threshold (near 0) or a high threshold (near 1). Second, we compared the centrality of the nodes and compared them to the threshold results in order to test our first hypothesis.

Third, we used chains of messages as a proxy to the information diffusion in the networks. Each chain of messages was calculated through the comparison of the timestamp of a tweet posted by a user and the timestamp of a tweet posted by this user’s neighbors. But as the messages in Twitter have a short life (i.e., they usually reach a large proportion of followers in few minutes and reach half of their organic reach in a few

hours), a message just is considered part of a chain if it is posted in a small time frame – in our research we used 3 hours as a threshold, since this is considered the period that a tweet reaches its mid-life. Finally, we analyzed whether the participation of the most central users in the information cascade is determinant for the existence of a chain.

Results

Our database contains 5,530 tweets posted by 2,971 unique Twitter accounts. Each node of our network represents one of these accounts and the edges represent existence of direct relationship between two nodes, i.e., the ties indicate whether one user is following other one in Twitter. In our initial analysis, we identified a high number of isolates\(^6\) in the network – around 12% of the nodes were not connected to any other node – and ten components. When looking just to the ten components, we found out that 99.2% of the nodes where inside the main components, with a very small percentage of users in the other nine minor components (only 21 users). For the purposes of this work, we decided to delete the isolates and the minor components of our network\(^7\), ending up with a connected graph of 2,583 nodes and 49,545 edges. This network has a very low density (the graph has only around 0.7% of the possible edges) and a low transitivity (the likelihood that the adjacent vertices of a vertex are connected is of approximately 15%), indicating that a high number of the users have a low number of followers, while a small number of users have a high number of followers. The number of dyads may also indicate

\(^6\) We found 367 isolates in the network. This high number may be the result of two different processes: on the one hand, it may indicate that our sample, based on the hashtags, is excluding brokers that connect this isolates to the main component; on the other hand, this number may indicate that something beyond the network structure is producing incentives for the users to join the protest, as traditional media.

\(^7\) The isolates and the nodes of the minor components did not participated in the recruitment or in the chain of information of the main component. Therefore, the information provided by them would not help us to answer any of our hypotheses.
this: over nine thousand of the ties are mutual relationships, while over 31 thousand are asymmetric. These results are in accordance with the broadcasting characteristic of Twitter, which creates a small number of broadcasters, i.e., users who do not follow a lot of accounts, but who have a high number of followers.

Figure 1 shows the cumulative distribution function of activated users by days of protest, with the first day of protest being May 10. As we can see, different from most online mobilizations that have an S-shaped function, the #OcupaMinC protest had a highly engaged number of users in its first days. Approximately 50% of the total number of mobilized users had been activated by the 8th day of protest, while more than 80% of them were activated before the 20th day. Part of the explication for this trend is the small period of time between the first rumors of the extinction of MinC (day 0) and the release of the presidential decree reinstating it (day 11), as well as the fast spread of offline occupations and the media attention it generated – which must have acted as an exogenous variable in the recruitment process in Twitter. But even when we analyze just the activation times during these first days of protest (Figure 2), the proportion of activated users by days is more similar to a straight line than the S-shaped form, which means that the recruitment process was constant during that period and that the online mobilization was not affected by bursts of engagement. Another interesting fact is that the online mobilization continued active after the president reinstated the MinC for a long time, converting approximately 20% of new users. This significant amount of new activists was caused because of the prolongation of the occupations in some buildings, which merged the #OcupaMinC movement with the mobilizations against the
impeachment of Dilma Rouseff and the interim president (a movement that it is still active today and usually use the hashtag #ForaTemer in social media).

**Figure 1** – *Cumulative proportion of activated users by day, from May 10 until September 19.*

**Figure 2** – *Cumulative proportion of activated users by day, from May 10 until May 21*

When we compared the activation times of each user with the activation times of their respective neighbors, we got the user threshold, i.e., the proportion of neighbors that need to be activated for a specific user to be recruited and participate in the online
mobilizations. Figure 3 shows the proportion of users according their threshold. The “starters” (users with threshold equal to 0) represent more than 20% of the nodes, while the hardest users to be recruited (those with a threshold equal to 1) represent around 15%. Both the “starters” and the late users represent the biggest proportion of users of our network, with the next peak of user in the 0.5 threshold. These results shows that a great number of users decided to participate as “starters” for reasons not predicted by the network structure, but also a significant number of the high threshold users waited until all their neighbors where activated to post some message.

So in order to test whether the centrality of the nodes is an important variable in the recruitment of medium (from 0.2 to 0.8) and high threshold users (above than or equal to 0.8), we compared the users’ threshold with the in-degree centrality\(^8\) of the nodes. As shown in Figure 4, we could not find a significant correlation between low threshold users and the most central nodes of the network. Indeed, the chart shows that the users with a high in-degree centrality are dispersed among all the threshold levels. Thus, these results led us to refute our first hypothesis (H1): the recruitment of new protesters is not dependent on the activation times of central users in the network. This means that the topology of the network do not have a significant effect in the dynamics of recruitment, which may be better explained by the influence of local networks (i.e, clusters of nodes) than by central users, in line with the homophily patterns produced by SNS.

\(^8\) For our purposes, the in-degree centrality is a suitable measure for one main reason: since it measures the number of nodes following a user, the in-degree centrality can help us to demonstrate whether broadcasters are influential in the dynamics of recruitment. However, for the sake of the argument, we have compared the user’s threshold with other measures of node centrality (degree, out-degree, betweenness, and eigenvector centrality) and all the measures displayed similar patterns to the in-degree centrality. To see the distributions, look at Appendix 3.
Moving to our second hypothesis (H2), we analyzed cascades of information created during the online protest to explore its dynamics of diffusion. The cascades are
composed of tweets that were posted within an interval of at most 3 hours by different users. For the purposes of this paper, we decided that a message is part of only a cascade, meaning that each cascade has a unique number of tweets. From the 5,503 posted during all the protest, 2,319 were part of one of the 472 information cascades, i.e., around 42% of the messages were diffused and did not died in the moment that they were posted, according to our parameters. In Figure 5A we have the number of the tweets by information cascades, showing that almost all chains of information were composed by a small number of tweets, which means that the vast majority of the cascades did not have a long life. Few of them were able to generate chains of more than 100 messages, with the maximum cascade having 753 tweets. When we look at the number of users by cascade (Figure 5B), we see a similar pattern: few cascades include messages from more than 20 users, with the maximum number of users interacting in a cascade being equal to 128.

Our comparison of the size of the cascades with the in-degree centrality of their participants is shown in Figure 6. According to the chart, there is a positive correlation between these both variables, i.e., the likelihood of a central user being involved in a longer chain of information is high. Once again, we compared the in-degree centrality of the users with the information cascades in order to analyze whether the dynamics of diffusion are dependent on central nodes or not. Thus, as shown in Figure 6, we have found a correlation between the topology of the network and the spread of a message on it, as proposed on our second hypothesis.

---

To calculate the cascades, we identified the first user to post a tweet in our dataset and then looked to all his followers to see whether they have posted a message within 3 hours. If any of the followers have posted a message, we consider his message part of this cascade. Then we look to the followers of the follower that have posted something in order to see if they have posted something within 3 hours of the follower’s post. We continued doing the same procedure until the last posted tweet.
Figure 5 – Size of cascades by number of tweets and by number of users

A - Number of tweets by cascade

B - Number of users by cascade

Figure 6 – Size of cascades by in-degree centrality

A - Size of cascade (tweets) by in-degree centrality

B - Size of cascade (users) by in-degree centrality
Limitations

The empirical evidence emerged from this method is an important first step in order to understand the dynamics of recruitment and diffusion in SNS. However, for the purposes of this research, we based our analysis on some assumptions that may have affected our results. We assumed that the likelihood of an individual mobilizing (i.e., favoriting, retweeting, replying, or creating a message) in an online protest depends more on the number of mobilized neighbors and on his threshold than in other factors, as traditional media influence. Nonetheless, the news about the occupations, which were one of the main headlines during the initial weeks of the protest, may have influenced the online protesters in ways that our analysis is not able to capture. We also assumed that once an individual is mobilized, he continued mobilized until the end of our timeframe and that the network structure was the same for the whole period, since we do not have snapshots of the network evolution.

Another important assumption is related to Twitter’s algorithm change of early 2016, which probably have produced impacts in the dynamics of recruitment and diffusion of political mobilizations. Until March of that year, the tweets of the timeline were showed to a user in a chronological way – the newest ones were displayed first –, but the new algorithm is showing first the tweets of accounts that a user interacts with the most. Now, the tweets are ranked based on their organic engagement and are showed to a user according to three factors: the users interests and interactions, the trending topics of the moment, and the tweets with multiple retweets, likes and comments. Thus, Twitter’s change brings its dynamics closer to the ones promoted by Facebook and increases the power of influential accounts, since it favors influential tweets presenting them first in the
timeline. Nevertheless, in spite of this change, we used timeframes to calculate the diffusion of information based in the metrics of the old algorithm, i.e., we assumed that a tweet achieved its mid-life in 3 hours, which may have changed. However, we still do not have good metrics for the new features, although we infer that nowadays tweets of influential accounts must have a longer mid-life, while tweets of normal accounts must have a shorter one, due to the new exposure timeframe.

Finally, we want to mention some pitfalls of our methodology. First, as the search for hashtags and keywords in tweets ignores messages addressing the protest that do not use the specific keywords (Hanna 2013), our sample may have some missing links. Second, we must be aware that, this methodology is not appropriate to explain individual behavior and motivations, since it does not help to explain how much of the behavior is caused by the platform design or by the user’s psychosocial characteristics. To untangle this is a difficult task, mainly because the system’s algorithm is designed to capture some features of the human behavior, as the tendency to homophily, transitivity, and propinquity (Ruth and Pfeffer 2014).

**Conclusion**

This paper makes a contribution to the current debate about the implications of the use of social media regarding protesters recruitment and information diffusion in political action. As showed above, our main hypothesis about the dynamics of recruitment in the network is false: the recruitment of new members is not dependent on the activation times of central users in the network. Indeed, there is a fairly distribution of low threshold users among the degree centrality range. This result gives support to the
homophily patterns approach and to the assumption that the engagement of protesters in Twitter is more related to the local users than to global, influential users.

The hypothesis about the dynamics of diffusion is instead true, as shown above. The diffusion of messages on Twitter, as in many other SNS, has a very short life and a low organic reach. It is also true that the diffusion of the movement message in a broader environment was only possible because of central users that acted as informational gatekeepers in the network.

The technological architectures behind each SNS pre-determined the possible behaviors of interaction among their users, which affected the dynamics of recruitment and the cascading of information in these online mobilizations. To untangle the power of the algorithm and the business strategies behind it is a major problem to the study of social movements in the digital era. Designs and interfaces of social media platforms create different kinds of communities and information environment, making users more or less conducive to political actions. Thus, more research must be done about SNS uses in activism, participation, and protest in order to better understand how the patterns of recruitment and diffusion in Twitter and in other SNS affects the online political mobilizations and how social movements can use this patterns in their favor.
Appendix 1

List of hashtags searched:

1. CulturaResiste
2. FicaMinc
3. MinCeNosso
4. MinCResiste
5. OcupaCapanema
6. OcupaCultura
7. OcupaFunarte
8. OcupaIphan
9. OcupaMinC
10. OcupaMinCAM
11. OcupaMinCDF
12. OcupaMinCES
13. OcupaMinCMG
14. OcupaMinCMT
15. OcupaMinCRJ
16. OcupaMinCSC
17. OMinCeNosso
Appendix 2

Network graph
Appendix 3

Different measures of centrality

User's threshold by degree centrality

User's threshold by out-degree centrality

User's threshold by betweenness centrality

User's threshold by eigenvector centrality
References


