Chapter 78
Characterizing and Modeling Collective Behavior in Complex Events on Twitter

A.J. Morales, J. Borondo, J.C. Losada, and R.M. Benito

Abstract All around the world people are increasingly using Internet and online social networks to relate among each other. This fact is bringing a unprecedented amount of user generated data, which is certainly attracting research on several fields. In this work we analyze the user interactions in Twitter around two politically motivated events, like a Venezuelan protest and the 2011 Spanish Presidential electoral campaign. We found that users participated quite heterogeneously, as a tiny fraction of them concentrates much of the activity or collective attention. This heterogeneity gives place to critical features, like interaction networks with power law distributions and modular structure. Although online social networks appear to be a pure social environment, we found traditional agents, such as well known politicians and media hold loads of influence among the participants.

Over the past years, new technologies and specially online social networks have penetrated into the world's population at an accelerated pace. An important feature of these communication tools is that they provide a large amount of user generated content, useful for research on political activism [1, 2], marketing techniques [3] and social influence dynamics [4]. In this study, we use data available from Twitter, to unveil and analyze the structural and dynamical patterns from the user interactions, in order to characterize the emergent collective behavior. We have focused our study around two relevant events: a Venezuelan political protest, that took place exclusively online, and the 2011 Spanish Presidential electoral process. On these events, users posted messages identified with special keywords, such as #SOSInternetVE for the Venezuelan protest and #20N for the electoral campaign. These special keywords identified the topics which messages we downloaded, using the Twitter API. The properties of these datasets are presented in Table 78.1.

A.J. Morales (✉) · J. Borondo · J.C. Losada · R.M. Benito
Grupo de Sistemas Complejos, Universidad Politécnica de Madrid, ETSI Agrónomos, 28040 Madrid, Spain
e-mail: alfredo.moralesg@alumnos.upm.es

R.M. Benito
e-mail: rosamaria.benito@upm.es

T. Gilbert et al. (eds.), *Proceedings of the European Conference on Complex Systems 2012*, Springer Proceedings in Complexity, DOI 10.1007/978-3-319-00395-5_78,

Table 78.1 Datasets properties

Topic	Period	Messages	Participants
20N	Nov. 5–20, 2011	370000	100000
SOSInternetVE	Dec. 14–19, 2010	420000	77700

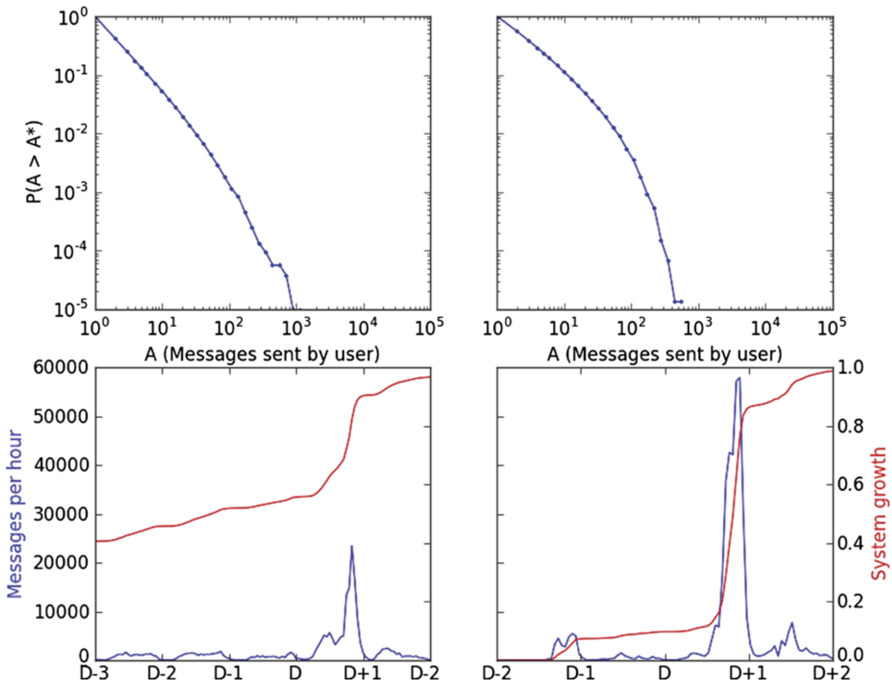

Fig. 78.1 Complementary cumulative distribution of messages sent by user (*top*) and Message rate through time (*bottom*). *The left* and *right panels* correspond to 20N and SOSInternetVE topics respectively

The distribution of the user activity, in terms of messages sent by user, as well as the message ratio through time, are shown in Fig. 78.1. The results indicate that users participated in an extremely heterogeneous way. In both cases, we found that the large majority of users (over 90 %) posted only a few messages each, while half of the messages were posted by less than 10 % of the participant population. This fact implies that the conversations were actually fed by a small portion of very active participants. Besides, both topics grew in a bursty manner, as can be seen in the bottom of Fig. 78.1, where the activity is heterogeneously concentrated in time. It can be noticed that a single conversation may grow up to 60 % of its final size in less than 8 hours.

However, not everybody's messages (or activity) have the same impact on the development of the event, since it remarkably depends on the source's connectivity inside the social substratum. To analyze this matter, we have constructed networks

Table 78.2 Followers, Retweet and Mention networks degree Pearson correlation (r)

Topic	$r_{F,R}$	$r_{F,M}$	$r_{M,R}$	$r_{A,R}$
20N	0.55	0.44	0.35	0.30
SOSInternetVE	0.57	0.70	0.85	0.15

based on "who follows who", which are subgraphs of the Twitter's global followers network, made with the events participants. On Twitter, when a user posts a message, this is instantaneously delivered to his/her own followers. Therefore, these networks represent the social substratum and available channels through which the information may flow along an event. In Fig. 78.2 we present the in and out degree distributions, which illustrate the heterogeneous connectivity found among the participants. In fact, while the large majority of users are followed by less than 20 users each, half of the social links are targeted to less than 2 % of the users. This means that the messages written by these hubs are delivered (and probably read) by half of the participants.

This heterogeneous connectivity gives place to an heterogeneous collective attention. To study so, we have also built other networks, linking the participants according to "who retweeted (retransmitted) who" and "who mentioned who". These interaction mechanisms display effective links where messages were propagated and selectively delivered, respectively. These networks have directed and weighted edges, and the in and out strength distributions are also presented in Fig. 78.2. It can be appreciated that both mechanisms are scale-free at the incoming links, which are the result of the aggregation of individual efforts, reflected in the out strength distribution. In fact, while the large majority is hardly mentioned or retweeted, less than 1 % of the participants, concentrate half of the mentions and retweets. Such an exclusive elite concentrates the largest part of the collective attention in both mechanisms.

Influence in Twitter has been considered to depend not only on the user's topological features in the followers graph, but also on the user's topological features in the retweet and mention graphs [4]. In the two considered cases, these measures are remarkably correlated, as may be seen in Table 78.2, where we present the Pearson correlation for the in degree and in strength values across the three networks, which resulted to be positive at all cases. We detected that the small fraction of hubs (who are influencers among the participants) act like information producers, posting messages widely delivered and retransmitted throughout the network. On the other hand, we found that the large majority of users act like information consumers, either actively or passively. Nevertheless, in order to gain influence, the regular users must play an active part in the conversation, as we demonstrated in a previous study [1], where we detected several cases of regular users who equaled the retransmission levels gained by popular accounts, by means of increasing their activity several order above. This is also supported by the positive Pearson coefficient between the retweets in strength and the user activity, also presented in Table 78.2.

In order to unveil how such heterogeneous users interacted with each other, we calculated the assortativity by degree coefficient [5, 6] for all networks. The results presented in Table 78.3, show that the emergent networks from Twitter are disassor-

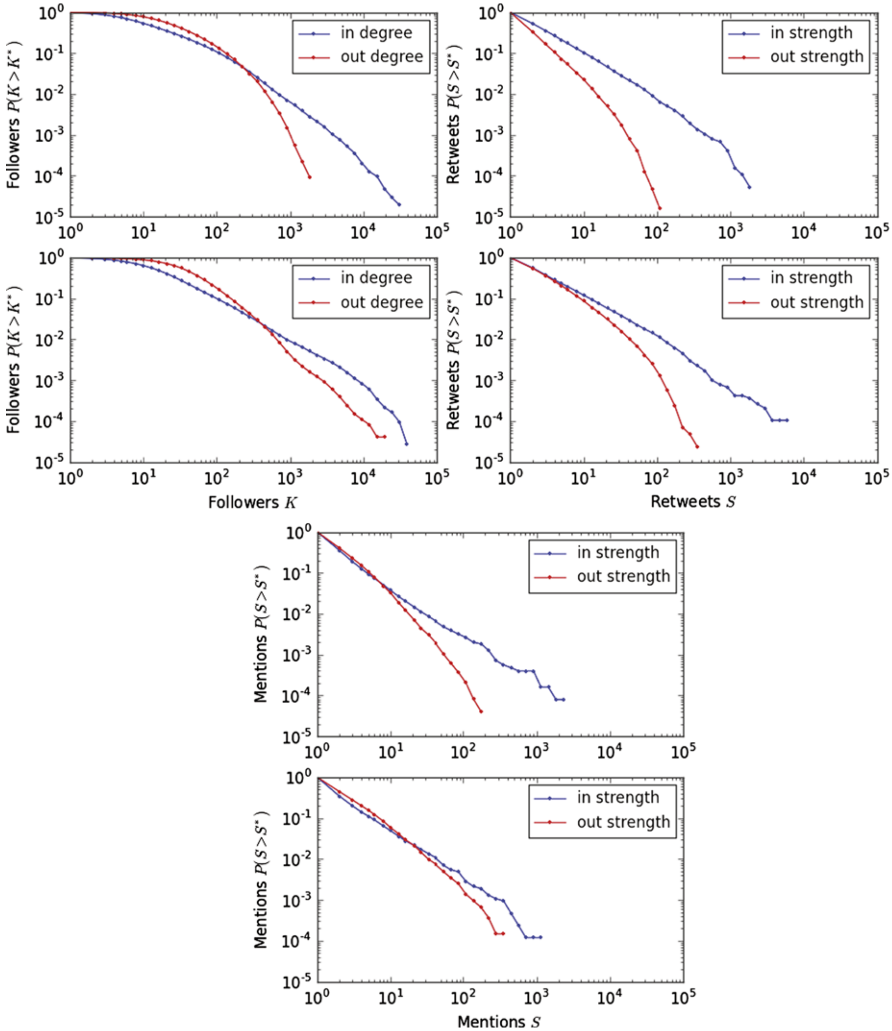

Fig. 78.2 Degree distribution of the Follower Network (*top left*), strength distribution of the Retweet Network (*top right*) and strength distribution of the Mention Network (*bottom*). *The top* and *bottom panels* correspond to 20N and SOSInternetVE topics respectively

tative. This result displays the asymmetric shape of these networks, where the hubs that concentrate much of the incoming links, are often targeted by regular users, who neither mention nor retweet too much, and receive few of the collective attention. Previous works on network assortativity [5], state that social networks tend to be assortative, as popular people want to be friend with popular people, and regular people are usually friends among the regular people. However our measures indicate something different. Hu and Wang [7] reported that other online social networks are also disassortative. The reason for this observations, relies on the difference between

Table 78.3 Assortativity by degree of the followers, retweet and mention networks

Topic	Followers	Retweets	Mentions
20N	−0.09	−0.06	−0.09
SOSInternetVE	−0.10	−0.15	−0.14

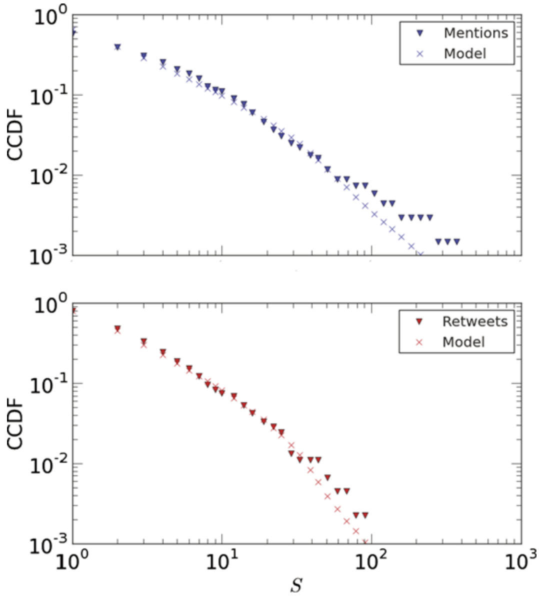

Fig. 78.3 Strength distributions of the political filtered Mention (*top*) and Retweet (*bottom*) networks and model results after 500 realizations

the online and offline world. For example, in Twitter regular people are now able to relate and communicate with popular accounts, either by following, mentioning or retweeting their messages. These new kinds of interactions are responsible for the changes in the structural and dynamical patterns previously reported on social networks.

Such different profiles also give place to the emergence of community structures, as the information consumers usually participate around their preferred information producers. In order to unveil such structures, we have performed community detection analysis based on modularity optimization [8] and random walks [9]. We have found that the retransmission and mention graphs present a higher modular structure than the followers one, being the retransmission graph even more segregative than the mentions map. Such structural differences reinforce the idea that the retransmissions and mentions channels are a substructure of the social substratum that endorses the individual preferences, and also indicate that people are more selective when taking action, than when just receiving the information [1].

On top of this, we have also found that the information producers, at the core of each community, are usually related to mainstream, celebrities or politicians accounts. This lead us to state that even though online social networks appear to be a pure social environment, traditional media agents hold loads of influence inside the network, that they use to boost their messages. However, according to the nature of the event and the interaction mechanism, some collectives may play a more influential role than others among the users. For example, in the 2011 Spanish electoral process [2], mentions are mostly targeted to politicians, while retransmissions are dominated by mainstream, since the first mechanism is used to send personal opinions and the other one is used to rapidly propagate information like news.

Finally, we have been able to model the modular and segregative structure of the mention and retransmission graphs, based on the formalism of heterogeneous preferential attachment [10], by designing connection rules for both micro and mesoscale. The idea behind this model is that the probability of a node i interacting with a node j not only depends on their respective degree, but also on an affinity value between them. This affinity value comes from a function that allow us to tune the mesoscale, independently from the microscale connectivity rules.

We tested this model with the mention and retweet networks of the Spanish electoral process, filtered by official politicians accounts. We found these subgraphs to be highly segregative, since the Pearson coefficient across parties are very close to 1 ($r_M = 0.905$ and $r_R = 0.990$), indicating a considerable lack of debate between the politicians. To model the mesoscale, we first calculated the affinity value across political parties, as the relative flux of interactions among them. In Fig. 78.3 we present the real and modeled strength function for both networks. It can be noticed that the model reproduces very well these distributions, as well as the Pearson coefficient across parties ($r_M = 0.86 \pm 0.03$ and $r_R = 0.989 \pm 0.005$).

In summary, our study reveals the complexity behind the interactions among users and the information diffusion process during particular events on Twitter. These interactions allow us to characterize and model the user's individual and collective behavior. We found that these topics were fed by a small portion of very active participants and driven by a smaller portion of very noticed influencers. These influncers are mostly related to main stream and celebrities, who use the social network to boost the importance of their messages. However, we found that influence might always be boosted by any participant when the activity is remarkably increased. The results obtained bring new insights into how people relate with each other in these communication tools and may serve as frameworks for professionals who use them, in order to maximize the network's potential.

References

1. Morales AJ, Losada JC, Benito RM (2012) Users structure and behavior on an online social network during a political protest. Physica A 391:5244–5253. doi:10.1016/j.physa.2012.05.015
2. Borondo J, Morales AJ, Losada JC, Benito RM (2012) Characterizing and modeling an electoral campaign in the context of Twitter: 2011 Spanish presidential election as a case study. Chaos 22:023138. doi:10.1063/1.4729139
3. Wu S, Hoffman JM, Mason WA, Watts DJ (2011) Who says what to whom on Twitter. In: 20th annual world wide web conference. ACM, New York

4. Cha M, Haddadi H, Benevenuto F, Gummad KP (2010) Measuring user influence on twitter: the million follower fallacy. In: Proceedings of the fourth international AAAI conference on weblogs and social media, Washington
5. Newman MEJ (2003) Mixing patterns in networks. Phys Rev E 67:026126
6. Foster JG, Foster DV, Grassberger P, Paczuski M (2010) Edge direction and the structure of networks. Proc Natl Acad Sci USA 107:10815
7. Hu HB, Wang XF (2009) Disassortative mixing in online social networks. Europhys Lett 86:18003
8. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech 2008:P10008
9. Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. Proc Natl Acad Sci USA 105:1118
10. Santiago A, Benito RM (2008) An extended formalism for preferential attachment in heterogeneous complex networks. Europhys Lett 82:58004