The Spanish “Indignados” Movement: Time Dynamics, Geographical Distribution, and Recruitment Mechanisms

Javier Borge-Holthoefer, Sandra González-Bailón, Alejandro Rivero and Yamir Moreno

Abstract Online social networks have an enormous impact on opinions and cultural trends. Also, these platforms have revealed as a fundamental organizing mechanism in country-wide social movements. Recent events in the Middle East and North Africa (the wave of protests in the Arab world), across Europe (in the form of anti-cuts demonstrations or riots) and United States have generated much discussion on how digital media is connected to the diffusion of protests. In this chapter we investigate, from a complex network perspective, the mechanisms driving the emergence, development and stabilization of the “Indignados” movement in Spain analyzing data from the period between April 25 and May 26, 2011. Using 70 keywords related to the movement, we analyze 581,749 Twitter messages coming from 87,569 users. The online trace of the 15M protests provides a unique opportunity to tackle central issues in the social network literature, like recruitment patterns or information cascades. These findings shed light on the connection between online networks and social movements, and offer an empirical test to elusive sociological questions about collective action.

Javier Borge-Holthoefer
Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), Universidad de Zaragoza, Zaragoza, Spain, e-mail: borge.holthoefer@gmail.com

Sandra González-Bailón
Oxford Internet Institute, University of Oxford, 1 St. Giles OX1 3JS, Oxford, UK, e-mail: sandra. gonzalezbailon@oii. ox.ac.uk

Alejandro Rivero
Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), Universidad de Zaragoza, Spain, e-mail: arivero@bifi.es

Yamir Moreno
Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), Universidad de Zaragoza, Zaragoza, Spain
Department of Theoretical Physics, Faculty of Sciences, Universidad de Zaragoza, Zaragoza 50009, Spain, e-mail: yamir.moreno@gmail.com

1 Introduction

Modern online socio-technological systems are changing the way we communicate with each other, revolutionizing the methods researchers have at hand to face old sociological questions. The networked nature of online communication platforms, plus the inherent complexity of the activity within them, places modern network theory [11] as an outstanding tool to provide deeper insight into questions such as the emergence of authoritative or privileged nodes [24, 39], the importance of the strength and range of connections [35], or time patterns of human activity [8]. These platforms generate an enormous amount of time-stamped data, making it possible for the first time to study the fast dynamics associated to different spreading processes at a global scale. These novel and rich data sources allow testing different social dynamics and models that would otherwise be highly elusive with traditional data-gathering methods. Additionally, the availability of data enables the study of phenomena that take place on time scales ranging from a few minutes or hours to years.

This chapter focuses on how social networking sites (SNSs) contribute to the organization of collective action. Online social media provide efficient and fast means to group together many actors around a common cause, as exemplified by the wave of social unrest staged around the world in 2011. Protests have aroused in the aftermath of financial and political crises in the form of pacific civil movements (as with the Spanish "Indignados" in May or "Occupy Wall-Street" in the United States, which culminated in global marches in October 15); economy-related demonstrations with some violent episodes (as in the case of the protests in Greece, Italy, and UK); and regime changing (and often violent) political uprisings (as in the "Arab spring"). Although these protests respond to very different circumstances, they all made use of SNSs to help protesters self-organize.

We focus on one of these examples: the Spanish "Indignados" movement (also known as the "15M" movement), which is still active as of January 2012. After the original mass demonstrations on May 15 (day *D* from now on), hundreds of participants decided to continue the protests camping in the main squares of several cities (Puerta del Sol in Madrid, Plaça de Catalunya in Barcelona) until May 22, the following Sunday and the date for regional and local elections. During that week, protesters created committees to coordinate the logistics of camp sites and organized around open popular assemblies. The media, which had not covered the movement until the day of the first big demonstrations, started covering the protests on a daily basis, particularly after the authorities tried to evict protesters from the squares by force, and the Electoral Committee declared the protests illegal. Despite the prohibition, the camps remained in place, receiving increasing popular support and staging daily demonstrations. As many of the adherents were online social media users, the growth and stabilization of the movement can be analyzed using time-stamped data of Twitter messages, which we have collected and analyzed as described below (Sections 3 and 4).

A social phenomenon like the 15M movement is an excellent opportunity to understand network formation processes and online spreading dynamics. We characterize the structural patterns of the network of users who sent or received tweets containing keywords related to the 15M movement. We find that this network displays the typical features of other networks in nature such as scale-free degree distributions, a community structure at the mesoscale and high structural robustness [11, 17]. Beyond this general characterization –which has deep implications regarding the dynamical processes occurring on top of the structure–, we also analyze two key sociological aspects: first, we attempt to understand how recruitment takes place in a socio-political movement such as 15M, providing novel empirical evidence of the mechanisms behind the decision to join a protest and enroll in collective action. Models of collective action have identified important network mechanisms behind the decision to join a protest, but they suffer from lack of empirical calibration and external validity. Online networks, and the role that SNSs play in articulating the growth of protests, offer a great opportunity to explore recruitment mechanisms in an empirical setting. Second, we analyze the dynamical patterns that characterize the spreading of information over the 15M network. We show that diffusion is hampered by what we called *information sinks*: a great part of the traffic is delivered to a few users that do not pass the information along. We also find that, because of the firewalls created by these *information sinks*, very few messages generate global cascades. This falls in line with related research, which has shown that information cascades in online networks occur only rarely [7, 28, 42], with the implication that even online it is difficult to reach and mobilize a high number of people.

Details of these analyses follow, capitalizing from and synthesizing previous work [13, 25, 14] where partial stances of the phenomenon were offered. Prior to that, the following section presents a succinct summary of the principles of modern network theory.

2 Network theory in a nutshell

Historically, the study of networks has been mainly the domain of a branch of discrete mathematics known as graph theory. Since its birth in 1736, when the Swiss mathematician Leonhard Euler published the solution to the Königsberg bridge problem (consisting in finding a round trip that traversed each of the bridges of the Prussian city of Königsberg exactly once), graph theory has witnessed many exciting developments and has provided answers to a series of practical questions. In addition to the developments in mathematical graph theory, the study of networks has seen important achievements in some specialized contexts, as for instance in the social sciences. Social network analysis started to develop in the early 1920s and focuses on relationships among social entities, as communication between members of a group, trades among nations, or economic transactions between corporations. It was on the second half of the 20th century, however, when networks began to produce fruitful models, as for instance in the works by Rapoport [38, and successive] or by Granovetter [26], and some methodological advances were made [23].

Although the concept of small-world was already well known by sociologists [30, 43], it was in 1998 that Watts and Strogatz formalized the properties of small-world networks [44], a model that became the seed for the modern theory of complex networks. The final leap to this emerging subfield was encouraged by increasing data availability and computing capacity, which were made widely available in the 1990s. From then network analysis methods have been used to model many complex real-world phenomena. Examples are numerous, ranging from the Internet –a network of routers or domains– to the global economy –a network of national economies, which are themselves networks of markets– or energy, which is distributed through transportation networks, both in living organisms, man-made infrastructures, and in many physical systems such as the power grids. All of them stand as examples of research under a network modeling approach.

There are excellent reviews devoted to the structural characterization and evolution of complex networks [2, 11, 17, 34]. Here we introduce only those network descriptors that are mentioned along the chapter (see, for example, Table 1).

The mathematical abstraction of a complex network is a graph G comprising a set of N nodes (or vertices) connected by a set of E links (or edges), being k_i the degree (number of links) of node i. This graph is represented by the adjacency matrix A, with entries $a_{ij} = 1$ if a directed link from i to j exists, and 0 otherwise. In the more general case of a weighted network, the graph is characterized by a matrix W, with entries w_{ij}, representing the strength (or weight) of the link from i to j. A connected network is an undirected network such that there exists a path between all pairs of vertices. If the network is directed, and there exists a path from each vertex to every other vertex, then it is a strongly connected network. If the network is not connected, i.e. it is made up of broken fragments, an interesting question is how large the giant component (or the largest connected subgraph) is with respect to N.

Much of the work in network theory deals with cumulative degree distributions, $P(k)$. A plot of $P(k)$ for any given network is built through a cumulative histogram of the degrees of vertices, and this is the type of plot used throughout this article (and often referred to just as "degree distribution"). Although the degree of a vertex is a local quantity, a cumulative degree distribution often determines some important global characteristics of networks. This notion can be extended to the weighted scheme, see Figure 2. The first classification of complex networks is related to the degree distribution $P(k)$. The differentiation between homogeneous and heterogeneous networks with respect to their degree is in general associated to the tail of the distribution. If it decays exponentially fast with the degree we refer to as homogeneous networks, the most representative example being the Erdös-Rényi (ER) random graph [21]. On the contrary, when the tail is heavy one can say that the network is heterogeneous. In particular, scale free networks are the class of networks where the distribution follows a power-law, $P(k) \sim k^{-\gamma}$, the Barabási-Albert model [9] being the paradigmatic model of this type of graph.

From $P(k)$ we can calculate the moments of the distribution. The n-moment of $P(k)$ is defined as

$$\left\langle k^n \right\rangle = \sum_k^N k^n p(k)$$

The first moment $<k>$ is the average degree of the network.

Furthermore, several other measures help to qualify further this categorization. Examples are the average shortest path length $L = <d_{ij}>$, where d_{ij} is the length of the shortest path between node i and node j –very small in complex networks if compared to N–, the clustering coefficient C that accounts for the fraction of actual triangles (three vertices forming a loop) over possible triangles in the graph –typically large in social networks, if compared to C in random graphs–, and assortativity r [33], obtained considering the Pearson correlation coefficient of the degrees at both ends of the edges (assortative behavior implies that topologically similar nodes link each other, which has interesting interpretations in terms of social structure).

Finally, researchers have not obviated other levels of analysis: between the global (system-wide) and the microscopic (node) statistical characterization, there exists an intermediate description level: the "meso" scale, in which the relevant analysis unit is a module or community. A modular view of a network offers a coarse-grained perspective in which nodes are classified in subsets on the basis of their topological position and, in particular, the density of connections between and within groups. In social networks, this classification usually overlaps with node attribute data, like gender, ideology or professional occupation. To achieve meaningful partitions of complex networks several algorithms have been designed. A thorough review of these techniques appeared in 2010 by Fortunato [22].

3 Data collection and modeling

We analyze Twitter activity around the 15M protests for the period comprised between April 25, 2011 at 00:03:26 (20 days before the first mass mobilizations) and May 26, 2011 at 23:59:55 (10 days after the first mass mobilizations, and 3 days after the elections). The activity data set follows the posting behavior of 87,569 users and tracks a total of 581,750 protest messages. Messages related to the protests were identified using a list of 70 *#hashtags* (see Appendix for a tag cloud with the most important ones, and Reference [25] for a complete list). The collection of messages is restricted to Spanish language and to users connected from Spain, and it was archived by a local start-up company, Cierzo Development Ltd using the SMMART Platform. Other SNSs had an important influence on the development of the whole movement. Activity on such sites –as well as blogs, web-pages, offline interaction– is blind to this analysis. Admittedly, this work can only account for a partial view of the whole process, a fact that should be kept in mind when reading and interpreting our findings.

An interesting subset of data is obtained when considering only directed messages, i.e. those including mentions, which contain the symbol @. Mentions allow the construction of a *directed* and *weighted* structure, which offers a partial view of the "activity" or "dynamical" network of communication between users. The direction of links indicates a source-target message emission, and link weight stands for the number of times that a user explicitly mentioned another user. Restricting data to messages that contain mentions is useful to characterize direct interactions between pairs of users, and the pace at which they get involved with the protests, although 2/3 of the information is left aside, i.e. messages that do *not* contain mentions are removed; this was the approach in [13]. A movie of this dynamic exchange is available online (see http://15m.bifi.es/index_en.php).

In addition to the dynamic network of active mentions, we also reconstructed the network of followers. This offers an almost-static view of the relationships between users, although that network also changes over time, its rate of change is significantly smaller than in the dynamic network of mentions (i.e. changes occur in the scale of weeks and months). Data for all the users that were active in the exchange of protest messages were scrapped using the Twitter API and a cloud of 128 machines. The scrap was successful for the 87,569 users, identified as active in message exchange, and we obtained their official list of followers. The static network, however, was restricted to those who had some participation in the protests. The resulting structure is a *directed unweighted* network, where direction indicates who follows who (see also [14] and [25]). This network exhibits a high level of reciprocity: a typical user holds many reciprocal relationships (with other users that are likely to be known personally), plus a few unreciprocated nodes which typically point at hubs, the so-called network "authorities". Any emitted message from a node *i* is immediately available to anyone following this user, in addition to those mentioned after the @ symbol. This particularity of Twitter is crucial to understand the concept of information cascade in the next sections.

-- table 1 around here –

Table 1 summarizes the main topological features of these two views of the data (the static follower network and the activity or dynamical network). Both the static and the dynamic levels reproduce well the "small world" features [44], i.e., low L and high C.

4 Methods and findings

We paid attention to three levels of analysis when studying the growth and stabilization of the protests: the global dynamics of message exchange over time; the spatial distribution of users at the community level; and the recruitment dynamics at the individual level. These three levels provide complementary views of the patterns observed in the data. Details on each of them follow.

4.1 Time Dynamics: Movement Growth and Saturation

The analyses in this section answer the following question: does a collective mobilization of thousands of actors demand a slow, progressive growth over time or do SNSs enable their abrupt emergence? Figure 1 illustrates the way in which the activity network evolved by gaining adherents. The red squares in the figure represent the proportion of active nodes at time t (with a resolution of 2 hours) relative to the total number of users in the network at the end of the growth process; blue circles represent the number of messages produced relative to the total (12 hour resolution). Both sets show a similar tendency: the formation of the network and its later increase in size does not proceed in a gradual proportional manner, rather in a sequence of bursts concentrated in just a few days (from day D to day D+7). The process is driven by the offline events surrounding the movement: data does not control for exposure to offline media, which is likely to have interacted with social influence, or to other sources of information that might have also contributed to the system's growth (like, for instance, offline discussion networks). The number of active users saturates after D+7: in May 21 (D+6), the day preceding local and regional elections, almost the 90% of the network was already formed.

-- figure 1 around here –

A key aspect to understand these time dynamics lies at the distribution of degree (static network) and strength (dynamic network). The degree of a node i, k_i, is the number of neighbors it has. The strength s_i of a given node i is defined by the sum of the weights of its links. These magnitudes can further be divided in two measures: on the one hand, we have the in-degree (in-strength) derived from the links incident to the node, k_{in} (s_{in}). Conversely, k_{out} (s_{out}) represents the out-degree (out-strength) generated at a node. From these quantities we can derive $P(k_{in})$ ($P(s_{in})$) and $P(k_{out})$ ($P(s_{out})$), the cumulative probability distributions. $P(k_{in})$ and $P(k_{out})$ are fixed, be
cause we assume that the follower network is static; whereas strength distributions can be measured at different instants t of the evolving network.

Figure 2 shows the cumulative distributions of these quantities for several time aggregation windows. As can be seen in the left panels, even before the occurrence of the events that triggered public protests on day D, both $P(s_{in})$ and $P(s_{out})$ follow power-laws $P(s) \sim s^{-\gamma}$, but with different exponents (γ_{in} = 1.1 and γ_{out} = 2.3, respectively, as measured at D + 10). Plots for the degree of the nodes in the follower network (right panels) exhibit similar behavior, implying the existence of some rare nodes acting as hubs [9]. It is well known that the statistical properties of these variables in other technological, social and natural systems are also heterogeneously distributed. Nonetheless, the fact that the 15M network is scale-free has deep consequences regarding a number of relevant issues, including its origin, complexity, robustness and, from a dynamical point of view, the way in which information flows over the system. As the network obtained comes from the activity of the nodes, the heavy-tailed distribution of both the degrees and strengths of nodes suggests that its dynamics lack any typical or characteristic scale.

-- figure 2 around here –

Focusing on the left panels of Figure 2, the dynamic asymmetry between incoming and outgoing degrees or strengths is not surprising either. Indeed, individual behavior, which ultimately determines the resulting dynamics, is an intended social action, but the emergent properties of the collective behavior of agents are unintended [41]. Essentially, subjects decide when and to whom a given message is sent. Therefore, the aggregate behavior of all agents and their popularity (i.e., how many mentions a node receives, or how many followers it has) result from individual choices. This is what the in-and out-distributions reflect. As a matter of fact, the exponent of the power law characterizing the degree probability distribution $p(s)$ lies in the interval [2,3], as usually found in most real-world networks. Interestingly, spreading dynamics such as rumor and disease propagation processes are most efficient for scale-free networks whose exponent is precisely in this range [36, 32, 31]. Finally, the strength distribution for the tweets sent, $p(s_{out})$, also resembles a power law function with an exponent larger than 3, although in this case the distribution exhibits an exponential cut-off. This might be due to the fact that sending messages has an associated cost in terms of bandwidth availability, the cognitive capacity to produce different messages and ultimately an unavoidable physical limitation to type them [18, 19, 24]. Interestingly, the exponential cut-off is mirrored in $P(k_{in})$ (right panel): the capacity to follow other agents in Twitter is cognitively limited, thus the number of users following more than a hundred people on Twitter decays very fast.

-- figure 3 around here –

One of the main consequences of the functional form of the strength distributions is presented in Figure 3. The emergence of hubs, which is the signaling feature of scale-free networks, leads to a predictable oligopoly in the way information is spread. In Figure 3, we observe that the number of tweets sent grows with the number of active users of the network. The curves corresponding to different days (i.e., instances of the dynamic network) nearly collapse into a single one. This means that as users join the network, the traffic generated scales accordingly. The figure indicates that roughly 10% of active users generated 52% of the total traffic (which is another indication of the dynamical robustness of the network to random failures but, at the same time, of its fragility to attacks directed towards that 10% of users [3, 15]). These patterns, however, are in sharp contrast with the activity patterns that correspond to received tweets. In this latter case the number of in-strength hubs decreases with time. As shown in the figure, by $D + 10$, less than 1% of users receive more than 50% of the information. These nodes are identified as main receptors (government) or as potential spreaders (mass media) of messages. However, what at a priori seems to be a good choice, turns out to be harmful for the process of information spreading. These hubs, which we call *information sinks*, do receive a lot of messages but rarely ever act as spreaders, which means they do not pass those messages along to their followers. These asymmetries in the information flow critically shape other global processes like information cascades which, as will be shown below, are very difficult to trigger.

4.2 Spatial Dynamics: Communities and Geography

A clustered or modular structure is pervasive in many natural, social and technological networks. Modules are islands of highly connected nodes separated by a relatively small number of links and, in social networks, they are created by the fact that actors tend to gather with those who share similar attributes like cultural traits or professional interests [6, 29, 10, 22]. In political networks homophily clusters actors around ideologies or opinions [1, 16].

-- figure 4 around here –

We analyzed the community structure of the 15M activity network at $t = D+10$ (that is, as its size is stabilized). We applied a random walk-based clustering algorithm that optimizes a map equation on a network structure [40]. Although alternative community detection algorithms are at hand, we choose the previous strategy because it is suited for networks in which the dynamics of information flow is relevant [37, 40], as is our case. The output of the algorithm is a partition made up of 6,388 modules. Most of these communities have less than ten nodes, so we focus our analysis on the 30 most important modules from a dynamic perspective, i.e. those which concentrate most of the random walker's activity. These modules do not necessarily coincide with the first 30 communities ranked according to their size, but all of them contain over 100 nodes, accounting for $\sim$15% of the users. Figure 4 shows these 30 communities in a collapsed view (each node represents a community) [20]. Each community is assigned a tag, which corresponds to the most central node in that community. These nodes play an outstanding role in the dynamics of information: modules are highly hierarchical, and nodes that are central in their communities (i.e. the local hubs) are hubs in a global scale as well.

This community structure sheds new light into the social dimension of the protests. First, tags identifying the 30 largest communities are highly heterogeneous. Six of these modules correspond to important mass media (newspapers and television), suggesting that users rely on these agents to amplify their opinion. The same can be said of 3 modules corresponding to famous journalists. More interestingly, seven modules correspond to on-line activists and/or veteran bloggers. These agents are unknown to most people, but they are present in the network from its birth and enjoy a solid reputation that facilitates their being considered a reference in the movement. Remarkably, seven modules are formed by camps in 7 different cities. Madrid is of course the main one, as the movement began there (*#acampadasol*, which comprehends over 3,000 nodes). Other cities are Barcelona, Granada, Zaragoza, Valencia, Seville and Pamplona.

-- table 2 around here –

The fact that communities are heterogeneously dominated (either by a person, an organization or a place) suggests some interesting conclusions. First, unlike previous works [1, 16], modules are not homogeneously defined (for example, groups corresponding to political tendencies). Instead, the activity network breaks down into groups that are representative of different actors in the process of mobilization (mass media, opinion groups, journalists, etc.). Second, communities reflect the relative autonomy of each of the assemblies throughout the Spanish geography: each of these modules hardly connects to any other, indicating a low communication between them. Fourth, Madrid, the capital, poses one exception to this pattern: minor camps hold a strong communication interchange with the community represented by the capital hashtag *#acampadasol*. In conclusion, these analyses show that the movement is highly centralized because most peripheral communities are only influenced by Madrid.

From a geographical point of view, the analyses suggest that in spite of the potential that web 2.0 technologies have to bridge geographical distance, they are mostly used to communicate with geographically close people. In other words, the network is *global*, but communication is mainly *local*. This is further verified in Table 2, where we have summarized the percentage of people whose public geo-location information (when signing up for a Twitter account) coincides with that of the community represented by the city hashtag. Admittedly, location at sign-up time is not free of error –people might move, or simply be inexact when delivering that information. Nonetheless, data suggest that geolocation is a significant factor in users interaction.

4.3 Recruitment Dynamics: the Activation of Users

The third level of analyses focuses on the mechanisms that prompted users to start sending protest messages. In particular, we analyze processes of recruitment (when and how do users join this instance of collective action?) and of information diffusion (how does the system help share –or filter out– information?).

4.3.1 The Network Position of Recruitment Seeds

The analysis conducted to answer the first question provides deeper insights into why horizontal organizations (like the online-born platform coordinating the protest, see Appendix) are so successful at mobilizing people through SNSs: their decentralized structure plants activation seeds randomly at the start of the recruitment process, all over the network structure and in the local contexts created by a myriad of users and organizations. Users who start sending messages before anyone else are the leaders of the movement but they are embedded in very different local networks: some are more central, some more peripheral; their diverse positions within the overall network start activation chains that end up percolating to the entire network.

Time-stamped data tell us the exact moment when users start emitting messages, and allow us to distinguish between activists leading the protests and those who reacted in later stages. It also allows us to calculate, for every user, how many of their contacts had already sent protest messages at the time of their activation (k_a/k_{in}).

-- figure 5 around here –

Activists with an intrinsic willingness to participate have a threshold $k_a/k_{in} \approx 0$, whereas those who need a lot of pressure from their local networks before they decide to join are in the opposite extreme $k_a/k_{in} \approx 1$ (see [25] for details).

In order to identify the network position of recruitment seeds, we used the k-shell decomposition [4, 5] to dissect the network in centrality layers, where we then locate the leaders of the movement. The k-shell decomposition assigns a shell index k_S or "core" to each user by pruning the network down to users with more than k neighbors. The process starts removing all nodes with degree k = 1, which are classified (together with their links) in a shell with index $k_S = 1$. Nodes in the next shell, with degree k = 2, are then removed and assigned to $k_S = 2$, and so forth until all nodes are removed (and all users are classified). Shells are layers of centrality in the network: users classified in shells with higher indexes are located at the core, whereas users with lower indexes define the periphery of the network. *ks* in the static network ranges from 1 to 141, which means that core users have at least 141 connections each.

Figure 5 combines the activation time with the core of each user in the static network. The overall spread of gray dots and the averages, depicted for each core, clearly show that any node, regardless of the *k*-shell it belongs to, can be activated at any time. This result confirms that recruitment takes place through random and distributed seeding.

4.3.2 The Network Position of Information Spreaders

Information diffusion tells us not how the system attains a critical mass, but how it maintains its activity. In the context of collective action, information diffusion plays a key role to coordinate action and to keep adherents informed and motivated. Understanding the dynamics of such diffusion is important to identify the users that are more likely to transform the emission of a single message into a global information cascade.

An information cascade, starting at a trigger, occurs whenever a piece of information is (more or less unchanged) repeatedly forwarded towards other users. If one of those who "hear" the piece of information decides to forward it, he becomes a spreader, otherwise he remains as a mere listener. The information cascade becomes global if the final number of affected users N_C (including the set of spreaders and listeners, plus the seed) is comparable to the size of the whole system N. Intuitively, the success of an information cascade should depend on whether spreaders have a large set of followers or not (Figure 6). This fact highlights the entanglement between dynamics and the underlying (static) structure.

Our definition of cascade takes time into consideration assuming that, regardless of the exact content of a message, two nodes belong to the same cascade as consecutive spreaders if they are connected (i.e. the latter follows the former) and they show activity within a short time interval, Δt [14]. The probability that exogenous factors are leading activation is in this way minimized. This operationalization of cascades borrows the concept of spike train from neuroscience, i.e. a series of discrete action potentials from a neuron taken as a time series.

We apply the latter definition to explore the occurrence of information cascades in the data. In practice, we take a seed message posted by i at time t0 and mark all of i's followers as listeners. We then check whether any of these listeners showed some activity at time $t_0 + \Delta t$. This is done recursively until no other follower shows activity, Figure 6 gives an illustrative summary of the method we followed to reconstruct cascades. In our scheme, a node can only belong to one cascade. We distinguish information cascades (or just cascades, for short) from spreader-cascades. In information cascades we count any affected user (listeners and spreaders), whereas in spreader-cascades only spreaders are taken into account.

We measure the size distribution of cascades and spreader-cascades for three different scenarios: one in which the information volume is low (slow-growth phase, from D−20 to D−10), one in which activity is very high (explosive phase, D−2 to D+6) and one that considers all available data (which spans a whole month, and includes the two previous scenarios plus the time in-between, D − 20 to D+10). Within the different time periods –slow growth, explosive phase and complete time-span–, different time windows have been set to assess the robustness of our results.

Our proposed scheme relies on the contagious effect of activity; large time windows, i.e., Δt > 24 hours, are not considered.

-- figure 6 around here –

The upper panels (a,b,c) of Figure 7 reflect that an information cascade of the size of the system can be reached in any of the three phases. As expected, these large cascades are always rare events, as the power-law probability distributions point out. This result is in perfect accordance with our analysis at the global scale (see Figures 2 and 3) and robust to different temporal windows up to 24h. In contrast, lower (d,e,f) panels *d*o show significant differences between periods. Specifically, the distribution of involved spreaders in the different scenarios changes radically from the explosive period (Figure 7f) to the slow growth period (Figure 7d); the distribution that considers the whole period of study just reflects that the explosive period (in which most of the activity takes place) dominates the statistics. What these differences suggest is that, to attain similar results (a system-wide cascade) a proportionally much smaller amount of spreaders (users who receive a message and pass it on) is needed in the slow growth period.

-- figure 7 around here –

To identify the network position of these spreaders, we capitalize on previous work suggesting that centrality (measured as the k-core, see previous section) enhances the capacity of a node to be key in disease [27] and rumor [12] spreading processes. In the work by Kitsak and colleagues [27], it is discussed whether the degree of a node (its total number of neighbors, k) or its k-core (a centrality measure) can better predict the spreading capabilities of such node. Note that the k-shell decomposition splits a network in a few levels (over a hundred), while node degrees can range from one or two up to several thousands.

Figure 8 explores the same idea, but in relation to information cascades. In the left, a circular layout of the follower network. Central nodes (high k-core) are placed in the center. Nodes are colored according to their “cascading capacity”, i.e. the size of a cascade triggered in a specific node. The upper central panel of this Figure shows the spreading capabilities as a function of classes of k-cores. Specifically, we take the seed of each particular cascade and save its coreness and the final size of the cascade it triggers. Having done so for each cascade, we can average the success of cascades for a given core number. Remarkably, for every phase under consideration (slow growth, explosive and full), a higher core number yields larger cascades. Exactly the same conclusion (and even more pronounced) can be drawn when considering degree (lower central panel).

These results suggest that both degree and k-core are good cascade size predictors. However, the interest of the histograms in Figure 8 lies in the high-end regions: while there are a few hundred nodes in the high cores (and even over a thousand in the last core), highest degrees account only for a few dozen of nodes. In practice, this means that only extremely high degrees, which are very rare, can produce large cascades. On the contrary, high cascading capabilities are distributed over a wider range of cores, which in turn contain a significant number of nodes.

-- figure 8 around here –

The previous findings show that joint dynamics of recruitment and coordination took place in this mobilization. Early adopters in the first days of the protests acted as recruiters and ultimately reached a critical mass of core users. Users at this core, in turn, contributed to the growth of the movement by generating cascades of messages that triggered new activations. These joint dynamics illustrate why Twitter has played a prominent role in so many recent protests and mobilizations: it combines the global reach of broadcasters with local, more personalized relations; in the light of our data, both features are important to articulate the growth of a movement. These features, however, are necessary, not sufficient, conditions: offline events and offline information diffusion also contributed to the growth of a protest that we only track online.

5 Conclusions

It is generally accepted that social networking sites are dramatically altering the way in which humans communicate. Such platforms offer a public domain where political opinion and social activism emerge, creating in the process data that offer scholars the opportunity to test old theoretical claims of how political and social protests emerge. This unprecedented volume of data, however, demands modern tools and research strategies to tackle its complexity from different levels. This chapter has provided one example of the insights that following such a research avenue can yield.

-- figure 9 around here –

A network perspective of mass mobilizations, as observed from online activity, sheds new light on the old problem of collective action. We analyzed one instance of online mobilization integrating three different levels of analysis: time dynamics, spatial organization, and recruitment mechanisms. We paid close attention to the distinction between dynamic activity (i.e. the exchange of messages) and the underlying structure supporting that activity (i.e. the following/follower connections of Twitter users). The analysis of time dynamics reveals the global characteristics of the "Indignados" movement, allowing us to describe the sudden growth of the protests after a relatively long period of slow brewing. This level of analysis provides also some insights into how information flows and the high influence that a minority of users have on global dynamics. The modular structure of the network formed by the exchange of messages reveals that groups of users gather around relatively autonomous entities, like mass media, famous journalists or city-based camps. A somewhat surprising finding is that, in spite of the global reach of SNSs like Twitter, users mainly employ this tool to communicate locally, mostly to coordinate actions that were taking place offline. Finally, the analysis of recruitment patterns and information dynamics, which demands a finer-grained view of the 15M movement, reveals the importance that local contexts have to understand activity at the individual level. For the movement to attain a critical mass of adherents, recruiters planted seeds all through the network; but once the system percolates, only those users who lie at the core are influential enough so as to trigger large information cascades, which make global coordination possible.

These findings provide a comprehensive view of one particular instance of emergent behavior, integrating different levels of analysis to uncover the mechanisms underlying collective action. However, far from exhausting the subject, this research generates new questions. A fundamental problem with online data is that it only offers insight into one side of the story, excluding from the focus of analysis a number of relevant agents (like, for instance, traditional mass media) that also influence the process under study. Political activism, in particular, is a collective endeavor with manifestations at many levels, ranging from public demonstrations and camps to the presence in news and mass media reporting. Thus, the offline events that also contributed to trigger the growth of the movement remain unknown and are beyond the scope of this work. Another important limitation relates to the content of the messages exchanged: What type of content is more likely to be disseminated on a global scale? While theoretical models of information spreading have long been studied and modeled, we are still trying to understand how the attributes of the information exchanged (for instance, its emotional content and appeal) is related to its diffusion. Finally, there are many determinants of geographical propagation that remain unexplored: the spatial location of users contributing to online activity is relevant to understand part of the dynamics, but this does not illuminate the factors that made original protesters appear in certain places and not others. Our study locates the origins of the protest in an online network, but cannot answer the question of why the protest flourished at a time or place.

In spite of all these limitations, however, there is much to be gained from analyzing activity in online networks. Not only is human activity increasingly shifting to that digital realm, demanding new ways to understand social dynamics and behavior; online networks are also encouraging collaboration across disciplines (complexity science, physics, sociology) that were mostly unconnected before, promoting the development of novel methods and theoretical frameworks. Digital data open new windows to analyze long elusive processes like contagion and social diffusion, and the way in which they interact with overlapping network structures (like communication and spatial networks). Better theories and models will emerge from that knowledge, but also better tools for political participation.

Acknowledgements J.B-H is partially supported by the Spanish MICINN through project FIS200801240. S.G-B. is partially supported by the Spanish MICINN projects CSO2009-09890 and CSD2010-00034. Y. M. is supported by the Spanish MICINN through projects FIS2008-01240 and FIS2009-13364-C02-01 and by the Government of Aragon (DGA) through the grant No. PI038/08.

Appendix: Protests Organizers

The "Indignados" movement is a civic initiative with no party or union affiliation that emerged as a reaction to perceived political alienation and to demand better channels for democratic representation. The first mass demonstration, held on Sunday May 15, was conceived as a protest against bipartidism and the management of the economy in the aftermath of the financial crisis. It was organized by the digitally coordinated platform "Democracia Real Ya" ("Real Democracy Now"), born online about three months before the first day of demonstrations. Hundreds of entities joined the platform, from small local associations to territorial delegations of larger groups like ATTAC (an international anti-globalization organization) or "Ecologistas en Acción" ("Ecologists in Action"). Signatories of the original call included student associations, bloggers, defenders of human rights and people from the arts, but also hundreds of individual citizens of different age and ideologies. Under the motto "toma la calle" ("take the streets"), the movement organized peaceful protests that brought tens of thousands of people to the streets of more than fifty cities all over the country, with Madrid and Barcelona leading in numbers.

References

1. Adamic, L., Glance, N.: The political blogosphere and the 2004 us election: divided they blog. In: Proceedings of the 3rd International Workshop on Link Discovery, pp. 36–43. ACM (2005)
2. Albert, R., Barabási, A.: Statistical mechanics of complex networks. Rev. Mod. Phys. 74(1), 47–97 (2002)
3. Albert, R., Jeong, A., Barabási, A.: Error and attack tolerance of complex networks. Nature 406, 378–382 (2000)
4. Alvarez-Hamelin, J., Dall Asta, L., Barrat, A., Vespignani, A.: Large scale networks finger-printing and visualization using the k-core decomposition. Advances in neural information processing systems 18, 41 (2006)
5. Alvarez-Hamelin, J., Dall'Asta, L., Barrat, A., Vespignani, A.: k-core decomposition of Internet graphs: hierarchies, self-similarity and measurement biases. Networks and Heterogeneous Media 3(2), 371–393 (2008)

6. Arenas, A., Danon, L., Díaz-Guilera, A., Gleiser, P.M., Guimerà, R.: Community analysis in social networks. Eur. Phys. J. B 38, 373–380 (2004)
7. Bakshy, E., Hofman, J., Mason, W., Watts, D.: Everyone's an influencer: quantifying influence on twitter. In: Proceedings of the fourth ACM international conference on Web search and data mining, pp. 65–74. ACM (2011)
8. Barabási, A.: The origin of bursts and heavy tails in human dynamics. Nature 435(7063), 1251–1251 (2005)
9. Barabási, A., Albert, R.: Emergence of scaling in random networks. Science 286, 509 (1999)

10. Blondel, V., Guillaume, J., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 2008, P10,008 (2008)
11. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.: Complex networks: structure and dynamics. Phys Rep 424(4-5), 175–308 (2006)
12. Borge-Holthoefer, J., Moreno, Y.: Absence of influential spreaders in rumor dynamics. Physical Review E 85, 026116 (2012)
13. Borge-Holthoefer, J., Rivero, A., García, I., Cauhé, E., Ferrer, A., Ferrer, D., Francos, D., Íñiguez, D., Pérez, M., Ruiz, G., et al.: Structural and dynamical patterns on online social networks: the Spanish May 15th movement as a case study. PloS One 6(8), e23883 (2011)
14. Borge-Holthoefer, J., Rivero, A., Moreno, Y.: Locating privileged spreaders on an online social network. Physical Review E 85, 066123 (2012)
15. Cohen, R., Erez, K., ben Avraham, D., Havlin, S.: Resilience of the internet to random breakdowns. Phys. Rev. Lett. 85(21) (2000)
16. Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Flammini, A., Menczer, F.: Political polarization on twitter. In: Proc. 5th Intl. Conference on Weblogs and Social Media (2011)
17. Dorogovtsev, S., Goltsev, A., Mendes, J.: Critical phenomena in complex networks. Rev. Mod. Phys. 80(4), 1275–1335 (2008)
18. Dunbar, R.: Neocortex size as a constraint on group size in primates. Journal of Human Evolution 22(6), 469–493 (1992)
19. Dunbar, R.: Coevolution of neocortical size, group size and language in humans. Behavioral and brain sciences 16(4), 681–693 (1993)
20. Edler, D., Rosvall, M.: The map generator software package. URL http://www.mapequation.org

21. Erdös, P., Rényi, A.: On random graphs. Publ. Math. (Debrecen) 6, 290–297 (1959).
22. Fortunato, S.: Community detection in graphs. Physics Reports 486(3-5), 75–174 (2010)
23. Freeman, L.: A set of measures of centrality based upon betweenness. Sociometry 40 (1977)
24. Gonçalves, B., Perra, N., Vespignani, A.: Modeling users' activity on twitter networks: Validation of Dunbar's number. PloS One 6(8), e22,656 (2011)
25. Gonzalez-Bailón, S., Borge-Holthoefer, J., Rivero, A., Moreno, Y.: The dynamics of protest recruitment through an online network. Scientific Reports 1, 197 (2011).
26. Granovetter, M.: The strength of weak ties. American Journal of Sociology 78, 1360–80 (1973)
27. Kitsak, M., Gallos, L., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H., Makse, H.: Identification of influential spreaders in complex networks. Nature Physics (2010)
28. Kwak, H., Lee, C., Park, H., Moon, S.: What is twitter, a social network or a news media? In: Proceedings of the 19th international conference on World Wide Web, pp. 591–600. ACM (2010)
29. Lozano, S., Arenas, A., Sánchez, A.: Community connectivity and heterogeneity: clues and insights on cooperation on social networks. Journal of Economic Interaction and Coordination 3(2), 183–199 (2008)
30. Milgram, S.: The small world problem. Psychol Today 2 (1967)
31. Moreno, Y., Nekovee, M., Vespignani, A.: Efficiency and reliability of epidemic data dissemination in complex networks. Physical Review E 69(5), 055,101 (2004)
32. Moreno, Y., Pastor-Satorras, R., Vespignani, A.: Epidemic outbreaks in complex heterogeneous networks. The European Physical Journal B-Condensed Matter and Complex Systems 26(4), 521–529 (2002)
33. Newman, M.: Assortative mixing in networks. Phys. Rev. Lett. 89(20), 208,701 (2002)
34. Newman, M.E.J.: The structure and function of complex networks. SIAM Rev 45, 167–256 (2003)
35. Onnela, J., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J., Barabási, A.: Structure and tie strengths in mobile communication networks. Proceedings of the National Academy of Sciences 104(18), 7332 (2007)
36. Pastor-Satorras, R., Vespignani, A.: Epidemic spreading in scale-free networks. Phys. Rev. Lett. 86(14), 3200–3203 (2001)
37. Pons, P., Latapy, M.: Computing communities in large networks using random walks. Lecture notes in computer science 3733, 284 (2005)
38. Rapoport, A.: Spread of information through a population with socio-structural bias i. assumption of transitivity. Bull. Math. Biophys. 15, 523–533 (1953)
39. Ratkiewicz, J., Fortunato, S., Flammini, A., Menczer, F., Vespignani, A.: Characterizing and modeling the dynamics of online popularity. Physical Review Letters 105(15), 158,701 (2010)
40. Rosvall, M., Bergstrom, C.: Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences 105(4), 1118 (2008)
41. Rybski, D., Buldyrev, S., Havlin, S., Liljeros, F., Makse, H.: Scaling laws of human interaction activity. Proceedings of the National Academy of Sciences 106(31), 12,640–12,645 (2009)
42. Sun, E., Rosenn, I., Marlow, C., Lento, T.: Gesundheit! Modeling contagion through facebook news feed. Proc. ICWSM 9 (2009)
43. Travers, J., Milgram, S.: An experimental study of the small world problem. Sociometry 32(4), 425–443 (1969)
44. Watts, D., Strogatz, S.: Collective dynamics of 'small-world' networks. Nature 393, 440 (1998)

Table 1 Topological descriptors for both the static and dynamic view of the Twitter data. The information in this table for the activity network corresponds to the accumulation of directed (@) messages up to D+10. Both networks are built from the same users, but descriptors diverge largely. Remarkably, although both networks are sparse (low $\langle k \rangle$ with respect to the system's size N), dynamic activity is much more so. Many nodes, however, act as information sinks –they never emit messages, although most messages are mentions to them [13]. The clustering coefficient C is significantly high in both networks, given their density, which, in combination with the low L and D suggests that the network has a "small world" structure. Unlike other reported cases [34], r is negative for the static network, whereas degree appears to be uncorrelated in the case of the activity network. The giant strongly connected component N_{scgc} is comparable to the system's size in the follower network, which means that almost every node in it is mutually reachable; that is not the case for the dynamical network, which has a relatively small strongly connected core. This is not surprising given the biased patterns observed in the emission of directed messages.

		Static network	Activity network
N	Number of nodes	87,569	87,569
E	Number of edges	6,030,459	206,592
$\langle k \rangle$	Average degree	69	2.36
C	Clustering coefficient	0.22	0.034
L	Average path length	3.24	1.7
D	Diameter	11	4
r	Assortativity	-0.14	0.005
SCC	Number of strongly connected components	5,249	73,389
N_{scgc}	Size of giant component	82,253	13,103
$max(k_{in})$	Maximum in-degree	5,773	29,155
$max(k_{out})$	Maximum out-degree	31,798	289

Table 2 Geographic origin of nodes in region-based communities. Region-based modules are mostly formed by nodes whose geographical origin coincides with that of the most central node in the community. This statement is clear for almost all these modules, except in the case of Madrid and Granada. The case of Madrid is not surprising, given that #acampadasol is the reference of the whole movement, thus the community organized around this actant is a more heterogeneous one. Granada is a more intriguing exception.

community tag	area	fraction of users from same area
@acampadasol	Madrid	54%
@acampadabcn	Barcelona	81%
@acampadavlc	Valencia	63%
@acampadazgz	Zaragoza	82%
@acampadagranada	Granada	53%
@acampadasevilla	Seville	83%
@15MPamplona	Pamplona	71%

Fig. 1 Temporal evolution of the structure and the activity in the SNS. In red, the proportion of nodes that had shown some activity at a certain time *t*. In blue, the cumulative proportion of emitted messages as a function of time. Note that the two lines evolve in almost the same way.

Fig. 2 Left panels: Strength distributions for both received (top) and sent (bottom) messages display a power-law behavior as early as at $D - 2$. The fat-tailed distributions indicate that the 15M activity network is scale-free. The exponents that define the power-laws differ significantly between sent and received messages. Right panels: degree distribution of full and symmetric networks, with similar features (see text).

Fig. 3 Information flow: the figure represents the density of tweets received (left) and sent (right) as a function of the cumulative fraction of active users. For each day, data are normalized by the number of active users at that date. As a reference, the horizontal line corresponds to 50% of emitted/received tweets. Note that, on $D + 10$, less than 1% of the nodes receive half of the messages. On the contrary, the pattern of tweets sent hardly evolves from the beginning of the movement: 10% of the active nodes produce 50% of the messages. This asymmetry is coherent with the differences observed for the strength distributions.

Fig. 4 Community structure of the "Indignados" activity network: the figure shows, in a compact view in which each node represents a community, the 30 most important modules. They can be identified by a single node, around which the community is organized. These *local hubs* (labeled in the figure) agglutinate modules and act as information bridges connecting the whole network.

Fig. 5 Activation time as a function of the core of nodes: early adopters, responsible for the recruitment process, are not located at privileged network positions. On the contrary, they are spread all over the topology. The maximum core in the static network is $ks = 141$.

Fig. 6 The figure illustrates the concept of cascade that is used throughout this article. Times at which spreaders emit a message is color-coded. Spreaders are emphasized as larger circles. User 1 emits a message at time t, and all of his followers automatically receive it. Thus, they are already counted as part of the cascade (red circles). One of his followers (user 2), driven by the previous message, decides himself to participate at time $t + \Delta t$, posting a message himself. A second set of followers is included in the cascade. Finally, a third node (user 3) joins in and spreads the cascade further at time $t + 2\Delta t$. A node cannot be counted twice, note for example that user 4 is also following node 3. Many nodes remain unaffected, because they are not connected to any of the spreaders. The final size of the cascade is N_C/N = 22/34; the success of the cascade largely depends on the capacity to contact a "leader" or "privileged spreader", i.e., a hub to whom many people listens and who decides to participate. The interesting point, however, is that the number of spreaders needed to attain such success is very low (3), and over 50% of the cascade is triggered by just one of them. Adapted from [14].

Fig. 7 Upper panels (*a,b,c*): Cascade size probability distributions for the different periods considered. Lower panels (*d,e,f*): Probability distributions of spreaders involved in the cascades for the same periods. The exact periods considered in the analyses are indicated at the top of each panel. Adapted from [14].

Fig. 8 Left panel: nodes in the static network arranged according to their k-core: higher k-cores indicate that nodes are more central, node size accounts for degree centrality, and node color indicates the maximum size of the cascades generated by the user (users generating the largest cascades are depicted in orange). Central upper panel: average spreading capacity (with respect to the system size) of nodes grouped according to their k-core. N_C/N grows with coreness, but the bursty period (red squares) evidences a much less clear tendency, with many fluctuations and a lower overall spreading capacity if compared to the relaxed period (black circles). Central lower panel: The same information is showed as a function of the degree. Again, the slow growth period is the best one at predicting the extent of a cascade. Interestingly, average cascades for highest degrees outperform those triggered by highest k-core nodes by an order of magnitude. See main text for discussion on this aspect. Right panels show the k-core and degree distributions, i.e., how many nodes belong to each class. Note that the highest core contains over 1000 users.

Fig. 9 Word cloud with the main 50 hashtags used to track message activity during the mobilization. Hashtags sizes are proportional to the number of times they appeared in messages, although we have rescaled them to smooth the large difference in their frequencies (which range from 10^5 to 10). A complete list of the 70 hashtags is available in Reference [25].