ingredients, such as genuine complex contagion or self-induced motivation, to participate.

Time-constrained activity cascades. – Before addressing the model itself, we need to understand what is meant to be modeled. There is a general consensus around the concept of cascade, which —in the ICT environment— can be outlined in the following way: the basic criterion to include a node i in the cascade where j belongs to is to guarantee that i) i and j became neighbors at t_1 (the notion of "friend" must be understood broadly here); ii) i received a piece of information from j, who had previously sent it out, at time t_2; and finally iii) the node i sends out a piece of information at time t_3. Typically, no strict time restriction exists besides the fact that $t_1 < t_2 < t_3$: the emphasis is generally placed on whether the *same* content is flowing [8]. This content-based view is useful when considering very specific pieces of information (e-mail chain letters [11] or URL forwarding [12,13], for instance), but renders a scenario in which the only possible transition for a node (user) is from inactive (susceptible) to active (infected).

Such vision oversees the fact that online platforms allow users to share contents, but also (more often than not) to spread behavior. Indeed, discussion over a topic typically happens not by mere information retransmission, but by iterated activity (variable units of information expressed in online text, evolving over time) [14] which influences (motivates) other users to join. A richer, time-constrained representation can then be obtained if conditions i) to iii) above are accepted, except that for i to be included in an avalanche started at j, the piece of information being transmitted may or may not be the same, and $t_3-t_2 \leq \Delta\tau$, where τ is an arbitrary (typically small) time lapse. In other words, the condition for i and j to be included in the same cascade is to exhibit temporally correlated, quasi-synchronized activity [8,15,16]. With this slight modification, not one but multiple cascading events can be measured from an activity data set (for instance, a collection of time-stamped tweets), and a single user may participate many times in the same cascade, in different times.

Empirical (real) cascades hereafter refer to such time-constrained representation.

The threshold model. – Along the lines of content-based cascades, the reputed threshold model [17] (and its networked version [18]) mimics social dynamics, where the pressure to engage a behavior increases as more friends adopt that same behavior. Briefly, the networked threshold model assigns a fixed threshold τ, drawn from a distribution $0 \leq g(\tau) \leq 1$, to each node (individual) in a complex network of size N and an arbitrary degree distribution p_k. Each node is marked as *inactive* except an initial seeding fraction of active nodes, typically $\Phi_0 = 1/N$. Denoting a_i the number of active neighbors, a node i with degree k_i updates its state becoming active whenever the fraction of active neighbors $a_i/k_i > \tau_i$. The simulation of this mechanistic process evolves following this rule until an equilibrium is reached, *i.e.*, no more updates occur. Given this setup, the *cascade condition* in degree-uncorrelated networks can be derived from the growth of the initial fraction of active nodes, who on their turn might induce the one-step-to-activation (vulnerable) nodes. Therefore, large cascades can only occur if the average cluster size of vulnerable nodes diverges. This condition is met at [18,19]

$$F = \sum_k k(k - 1)\rho_k p_k = \langle k \rangle, \qquad (1)$$

where ρ_k is the density of nodes with degree k close to their activation threshold, p_k is the fraction of nodes of degree k and $\langle k \rangle$ is the average degree [18].

For $F < \langle k \rangle$ all the clusters of vulnerable nodes are small, and the initial seed can not spread beyond isolated groups of early adopters; on the contrary, if $F > \langle k \rangle$ then small fraction of disseminators may unleash —with finite probability— large cascades. More recently, the cascade condition has been analytically determined for different initial conditions [19] as well as for modular and correlated networks [20,21], while placing the threshold model in the more general context of critical phenomena and percolation theory [22].

As mentioned, the model has a limited scope since it can account only for one-shot events, for instance the diffusion of a single rumor or the adoption of an innovation. Also, this framework leaves no room for spontaneous initiative: even low-threshold nodes —those with higher propensity to participate in a cascade— will not be recruited unless their neighbors act upon them. Empirical evidence suggests, instead, that once an agent becomes active that behavior will be sustained, and reinforced, over time [23]. This creates a form of enduring activation that will be affected and affect other agents over time in a recursive way. Indeed, events evolve in time —and so do the cascades elicited therein [16], as a consequence of dynamical changes in the states of agents as dynamics progress. Cascades are then events that brew over time in a system that holds some memory of past interactions. Moreover, the propensity to be active in the propagation of information sometimes depends on other factors than raw social influence, *e.g.*, mood, personal implication, opinion, etc.

Integrate-and-fire model: analytical approach. – In this paper, we present a model with self-sustained activity, where system-wide events emerge as microscopical conditions become increasingly correlated. We capitalize on the classical integrate-and-fire oscillator (IFO) model by Mirollo and Strogatz [24]. In this model, each node in a network of size N is characterized by a voltage-like state $m \in [0,1]$ of an oscillator, which monotonically increases with phase ϕ until it reaches $m = 1$, and then it fires or *activates* (emits information to its coupled neighbors), and immediately deactivates (resets its state to $m = 0$). The pulsatile dynamics, makes that each time a node becomes active, the state of its k neighbors is increased by ε. More precisely, m is uniformly distributed at $t = 0$ and evolves

such that

$$m = f(\phi) = \frac{1}{w} \ln(1 + [e^w - 1]\phi) \qquad (2)$$

parametrized by $w > 0$ to guarantee that f is concave down. Whenever $m_i = 1$ then *instantaneously* $m_j = \min(1, m_j + \varepsilon)$, if the edge (i, j) exists. Thus, a node may reach activation either by itself (spontaneously, as its phase comes to an end) or because of its neighbors' action (which may pull it up to action).

Integrate-and-fire models have been extremely useful to assess the bursty behavior and the emergence of cascades in neuronal systems represented in lattices [25] and complex networks [26,27]. We propose to model social systems as a complex network of IFOs representing the time evolving activation of individuals. Within this framework, spontaneous propensity —activation regardless exogenous factors— is guaranteed; while contagion is genuinely complex, *i.e.* the number of necessary external influences (if any) to show activity varies in time. Remarkably, activity is purely periodical only if the oscillators are either isolated (disconnected), dynamically uncoupled ($\varepsilon = 0$) or the dynamics have reached full (irreversible) synchronization. These three scenarios are irrelevant in terms of social modeling. On the other hand, some traces of periodicity in users' activity in socio-technical platforms —like Twitter— have been widely studied at the aggregate level (see, for instance, [28]), and they exist also at the individual level.

The model comprises two free parameters, w and ε, which are closely related. Dissipation w may be interpreted as the willingness or *intrinsic propensity* of agents to participate in a certain diffusion event: the larger w, the shorter it takes for a node to enter the tip-over interval $1 - \varepsilon < m < 1$. Conversely, ε quantifies the amount of influence an agent exerts onto its neighbors when it shows some activity. Larger ε's will be more consequential for agents, forcing them more rapidly into the tip-over region. Both quantities affect the level of *motivation* m of a given agent. Note that ε in the current framework evokes τ in the classical threshold model, in the sense that both determine the width of the tip-over region. Finally, the phase is translated into time steps, and then prescribed as $\phi(t) = t$.

To gain some analytical insight, we use eq. (1) to derive the cascade condition in this new framework. Note that now the distribution of activity is governed by $\rho(t) = 1 - \int_0^{1-\varepsilon} g(m, t) \, dm$ where $g(m, t)$ corresponds to the states' probability distribution at a certain time t. For an initial uniform distribution of motivation m and a fixed ε, the condition for the emergence of cascades reads at time $t = 0$

$$\varepsilon \sum_k k(k-1)p_k = \langle k \rangle \qquad (3)$$

(see inset (a) in fig. 1). And in general for any time

$$\rho(t) \sum_k k(k-1)p_k = \langle k \rangle, \qquad (4)$$
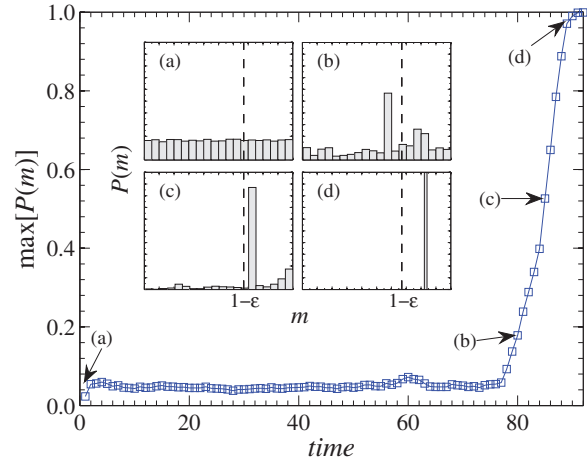
Fig. 1: (Color online) Insets (a)–(d): motivation m probability distributions of four different representative times along the synchronization window. Each snapshot depicts the m-state histogram of the N oscillators. The dynamics begins with a random uniform distribution of m-states (inset (a)) and it progressively narrows during the transition to synchrony (inset (d)). Main: largest fraction of synchronized nodes across time. The path to synchronization evolves steadily at a low level, and eventually suffers an abrupt transition.

which implies that the cascade condition depends on time in our proposed framework. Clearly, in this scenario $\rho(t)$ is not a function of the node degree k, as opposed to ρ_k in Watts' proposal.

As the dynamics evolve in time, the states of the nodes progressively correlate and, consequently, the distribution of states changes dramatically. The evolution of the states distribution is depicted in fig. 1. The initially uniform distribution $g(m, 0)$ (inset (a)) evolves towards a Dirac δ function (inset (d)) as the network approaches global synchronization, *i.e.* global cascade. We have not been able to find a closed analytical expression for the consecutive composition of the function $g(m, t)$ after an arbitrary number of time steps to reveal the evolution of $\rho(t)$, nonetheless it can be solved numerically. Equation (4) reduces to

$$\rho(t)(\langle k^2 \rangle - \langle k \rangle) = \langle k \rangle. \qquad (5)$$

The cascade condition is thus

$$\frac{\rho(t)}{1 + \rho(t)} = \frac{\langle k \rangle}{\langle k^2 \rangle}, \qquad (6)$$

that exactly corresponds to the bond percolation critical point on uncorrelated networks [29–31]. For the case of random Poisson networks $\langle k^2 \rangle \sim \langle k \rangle^2$, then

$$\frac{\rho(t)}{1 + \rho(t)} = \frac{1}{\langle k \rangle}. \qquad (7)$$

It is worth highlighting that eqs. (6) and (7) represent an advance in our understanding of non-linear, pulse-coupled dynamics, regardless of our (social) interpretation.

Integrate-and-fire model: numerical results. –
We can now explore the cascade condition in the $(\varepsilon, \langle k \rangle)$
phase diagram in fig. 2, and compare the analytical pre-
dictions with results from extensive numerical simulations.
Since the time to full synchronization (global cascade) is
different for each $(\varepsilon, \langle k \rangle)$, we introduce *cycles*. One cy-
cle is complete whenever every node in the network has
fired at least once. In this way we bring different time
scales to a common, coarse-grained temporal ground, al-
lowing for comparison. The regions where cascades size
S_c above a prescribed threshold ξ ($0.25N$, in fig. 2) are
possible are color-coded for each cycle $0, 25, 75, 100, \ldots$,
and black is used in regions where cascades do not reach
ξ (labeled as N.C., "no cascades", in fig. 2). Note that if
cascades are possible for a cycle c, they will be possible
also for any $c' \geq c$. This figure renders an interesting sce-
nario: on the one hand, it suggests the existence of critical
ε_c values below which the cascade condition is systemati-
cally frustrated (black area in the phase diagram). On the
other, it establishes how many cycles it takes for a particu-
lar $(\varepsilon, \langle k \rangle)$ pair to attain macroscopical cascades (full syn-
chronization) —which becomes an attractor thereafter, for
undirected connected networks. Given the cumulative dy-
namics of the current framework, in contrast with Watts'
model, the region in which global cascades are possible
grows with $\langle k \rangle$.

Turning to the social sphere, these results open the door
to predicting how long it takes for a given topology, and a
certain level of inter-personal influence, to achieve system-
wide events. Furthermore, the existence of a limiting ε_c
determines whether such events can happen at all.

Additionally, the predictions resulting from eq. (6) are
represented as dashed lines in fig. 2. For the sake of clar-
ity, we only include predictions for $c = 0$ (dashed black),
$c = 25$ (dashed gray) and $c = 150$ (dashed white). Pro-
jections from this equation run close to numerical results
in both homogeneous (fig. 2(a)) and inhomogeneous net-
works (fig. 2(b)) with degree distribution $p(k) = k^{-\gamma}$, al-
though some deviations exist. Noteworthy, eq. (6) clearly
overestimates the existence of macroscopic cascades in the
case of scale-free networks at $c = 0$. Indeed, $\rho(0) = \varepsilon$ does
not yet incorporate the inherent dynamical heterogeneity
of a scale-free topology, thus eq. (6) is a better predictor as
the dynamics loose memory of the hardwired initial con-
ditions. In the general case $c > 0$, deviations are due to
the fact that the analytical approach in the current work
is not developed beyond first order. Second order correc-
tions to this dynamics (including dynamical correlations)
should be incorporated to the analysis in a similar way to
that in [21], however it is beyond the scope of the current
presentation.

According to Mirollo and Strogatz [24], synchronicity
emerges more rapidly when w or ε is large; then the time
taken to synchronize, *i.e.* to observe global cascades, is in-
versely proportional to the product $w\varepsilon$. In our simulations
in the next Section, we use this cooperative effect between
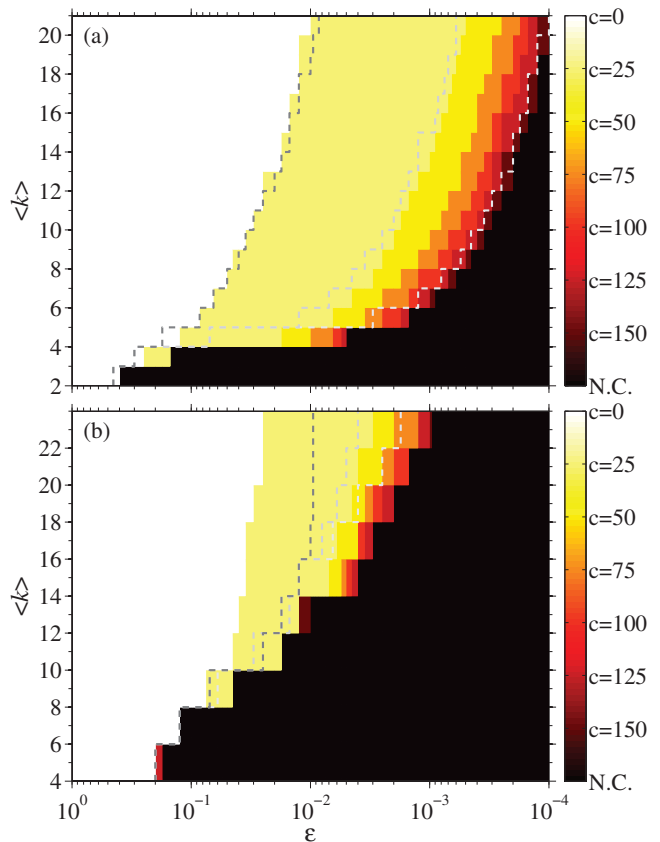coupling and willingness to fix ε, which is set to a set of

Fig. 2: (Color online) $(\varepsilon, \langle k \rangle)$ cascade diagram for different cy-
cles (coded by color), with fixed $w = 3$. Vertical axis and
each dashed line define a confined region in which global cas-
cades might occur according to eq. (6) and for a specific cycle
(here we show only the expected zones for $c = 0$ —dashed
white— and $c = 150$ —dashed gray). Results are obtained for
synthetic Erdös-Rényi (a) and scale-free with $\gamma = 3$ (b) un-
correlated networks of size $N = 10^4$. A cascade is considered
"macroscopical" if the synchronized cluster $S_c \geq 0.25N$. Color
codes indicate the existence of at least one cascade $S > S_c$ in
numerical simulations; analytical predictions are averaged over
200 networks with random initial conditions. Note that the
cascade condition in (a) often underestimates the actual cas-
cade regions because it does not take into account second-order
interactions; the same applies in the lower panel (b), except for
$c = 0$ where the analytical prediction overestimates the results
because the inclusion of the hub into the cascade is improbable
starting from a uniform distribution.

values (slightly above or below) $\varepsilon \simeq \varepsilon_c$, and empirically
estimate w to attain a good matching between observed
cascade distributions and our synthetic results.

Application to real data. – To illustrate the ex-
planatory power of the dynamical threshold model, we
use data from `www.twitter.com`. They comprise a set
of ~ 0.5 million Spanish messages publicly exchanged
through this platform from the 25th of April to the 25th of
May, 2011 [32]. In this period a sequence of civil protests
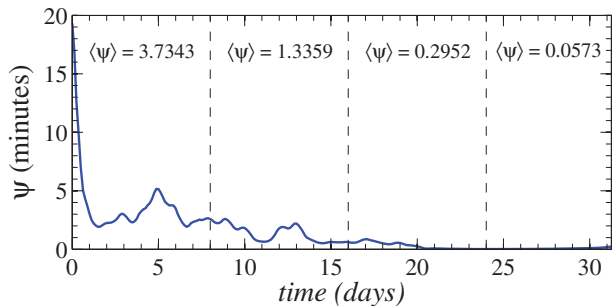and demonstrations took place, including camping events

Fig. 3: (Color online) Average inter-event times Ψ of the whole collection of data. To measure them, an overlapping sliding window scheme has been used (windows span 1 day, the offset between windows is 12 hours). To estimate w for 2 different periods of the protests, we take their corresponding time slices and compute inter-event times grand averages $\langle\Psi\rangle$. Thus, in fig. 4 a $w \sim 10^{-1}$ will be used for the first period and $w \sim 20$ for the last one. The scaling $w \sim 1/\Psi$ is merely an heuristic estimation, and some fine-tune is necessary to achieve a satisfactory matching.
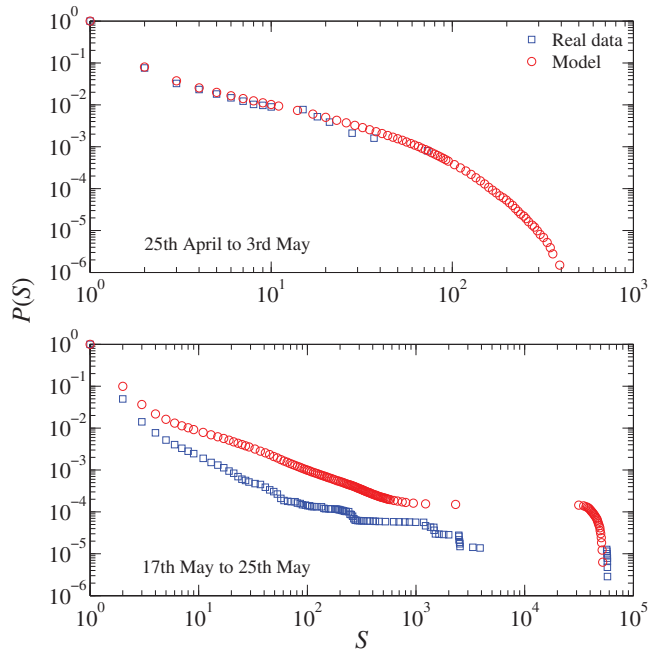

Fig. 4: (Color online) Cascade size cumulative distributions $P(S)$ of real data (blue squares) and the model counterpart (red circles). We have considered two time windows which significantly differ: first eight days (top) for which we have set $w = 0.1$; last eight days (bottom) for which we have $w = 30.0$. Note that $\varepsilon_c \approx 10^{-3}$. The model performs well in both periods, the relative error of the slope in the linear region is $< 1\%$ (see slope values on the figure). Real-data distributions are measured as described in the main text, see also [15,16].

in the main squares of several cities beginning on the 15th of May and growing in the following days. Notably, a pulse-based model suits well with the affordances of this social network, in which any emitted message is instantly broadcasted to the author's immediate neighborhood —its set of *followers*. For the whole sample, we queried for the list of followers for each of the emitting users, discarding those who did not show outgoing activity during the period under consideration. The set of users $N = 87569$ plus their following relations constitute the topological support (directed network) for the dynamical process running on top of it. The average number of followers of this network is $\langle k \rangle = 69$ and its degree distribution scales like $p(k) \sim k^{-1.5}$. Note that, unlike other substrates, friendship networks exhibit a high level of reciprocity. In particular, $r = 0.45$ (as defined in [33]) for this particular case, which implies that many links can effectively be regarded as undirected. This is why the theory —developed for undirected networks— is a reasonable approach for this particular case.

On top of the described network, we measure the empirical *time-constrained activity cascade* (or simply "cascade") size distribution for different periods, as explained previously. To test the proposed model, we run the dynamics on the same topology for some given parameters (w, ε). Since (as mentioned) ε is fixed to the network's particular $\varepsilon_c \approx 10^{-3}$, we only need to determine which values of w can adequately represent the evolving nature of events. Figure 3 represents the average inter-event times Ψ in the data for each 32 days of activity around the 15 M movement. Grand averages $\langle\Psi\rangle$ are computed for 8-day windows, roughly corresponding to the periods for which we offer model fittings (see below). As expected, Ψ drops abruptly in the first days —when the movement is brewing— and smoothly decreases afterwards, until

$\Psi \approx 0$. Intuitively, the willingness of an individual to participate in the protest is proportional to real-world excitation level, *i.e.* it scales like the inverse of Ψ, and thus we envisage that $w \sim 1/\Psi$.

In fig. 4 we compare empirical *vs.* synthetic cascade size distributions for different periods of the protests: the "slow-growth" phase (25th April to 3rd May; blue squares in the upper panel) for the first days, when the protest is limited to some online activists; and the "explosive" phase (19th to 25th May; blue squares in the lower panel), which comprehends the most active interval —the reaction to the Spanish government ban on demonstrations around local elections on the 22nd May. The proposed dynamics is run on the same topology for different w values, with remarkable success (red circles), though the bottom panel does not show so good of an agreement as the top one. In particular, a small range around $w \sim 1/\langle\Psi\rangle$ values was tested in the simulations, seeking a minimization of the relative error of the slope in the linear region of $P(S)$, *i.e.* the cascade size distribution.

Moreover, as seen in fig. 4, our model is able to capture different regimes. Admittedly, the gap observed in the data (fig. 4, bottom panel) is reminiscent of a supercritical regime, in which one either has small-size events or system-wide cascades. The change between regimes,

namely, from the subcritical one represented in fig. 4, top panel, and the super-critical phase needs the activity of the system to be long-lived, as this is a transition that takes place at different time windows. Note that this is a quite relevant feature of the model here introduced, since existing threshold models do not allow that individuals engage in more than one cascade (recall that once one individual is active remains so forever) and therefore cannot lead to the same kind of temporal transition observed in the real data.

Conclusions. – Summarizing, we have proposed a time-dependent continuous self-sustained model of social activity. The model can be analyzed in the context of previous cascade models, and encompasses novel phenomenology as the time dependence of the critical value of the emergence of cascades. We interpret it under a social perspective, where collective behavior is seen as an evolving phenomenon resulting from inter-personal influence, contagion and memory —while, collaterally, it sheds new light to the behavior of pulse-coupled oscillator dynamics. In a general perspective, our modeling framework offers an alternative approach to the analysis of interdependent decision making and social influence. It complements threshold models and complex contagion taking into account time dynamics and recursive activation, and also splits motivation into two components: intrinsic propensity and strength of social influence. We also anticipate that the exploration of the whole parametric space would lead to new insights about the effects of social influence and interdependence in social collective phenomena.

* * *

REFERENCES

[1] Watts D. J., *Annu. Rev. Sociol.*, **30** (2004) 243.

[2] Lazer D., Pentland A. S., Adamic L., Aral S., Barabási A. L., Brewer D., Christakis N., Contractor N., Fowler J. and Gutmann M. *et al.*, *Science*, **323** (2009) 721.

[3] Conte R., Gilbert N., Cioffi-Revilla C., Deffuant G., Kertesz J., Loreto V., Moat S., Nadal J.-P., Sanchez A., Nowak A. *et al.*, *Eur. Phys. J. ST*, **214** (2012) 325.

[4] Giles J., *Nature*, **488** (2012) 448.

[5] Rapoport A., *Bull. Math. Biophys.*, **15** (1953) 523.

[6] Daley D. J. and Kendall D. G., *Nature*, **204** (1964) 1118.

[7] Goffman W., *Nature*, **204** (1964) 225.

[8] Borge-Holthoefer J., Baños R. A., González-Bailón S. and Moreno Y., *J. Complex Netw.*, **1** (2013) 3, doi: 10.1093/cmnet/cnt006.

[9] Gleeson J. P., Cellai D., Onnela J.-P., Porter M. A. and Reed-Tsochas F., arXiv:1305.7440 (2013).

[10] Hogg Tad, Lerman Kristina and Smith Laura M., *Hum. J.*, **2** (2013) 25, arXiv:1308.2705 (2013).

[11] Liben-Nowell D. and Kleinberg J., *Proc. Natl. Acad. Sci. U.S.A.*, **105** (2008) 4633.

[12] Bakshy E., Hofman J., Mason W. and Watts D., *Everyone's an influencer: quantifying influence on twitter*, in *Proceedings of the 4th ACM International Conference on Web Search and Data Mining* (ACM) 2011, pp. 65–74.

[13] Lerman K., Ghosh R. and Surachawala T., arXiv:1202.3162 (2012).

[14] Leskovec J., Backstrom L. and Kleinberg J., *Meme-tracking and the dynamics of the news cycle*, in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM) 2009, pp. 497–506.

[15] González-Bailón S., Borge-Holthoefer J., Rivero A. and Moreno Y., *Sci. Rep.*, **1** (2011) 197.

[16] Borge-Holthoefer J., Rivero A. and Moreno Y., *Phys. Rev. E*, **85** (2012) 066123.

[17] Granovetter M., *Am. J. Sociol.* (1978) 1420.

[18] Watts D., *Proc. Natl. Acad. Sci. U.S.A.*, **99** (2002) 5766.

[19] Gleeson J. and Cahalane D., *Phys. Rev. E*, **75** (2007) 056103.

[20] Galstyan A. and Cohen P., *Phys. Rev. E*, **75** (2007) 036109.

[21] Gleeson J., *Phys. Rev. E*, **77** (2008) 046117.

[22] Hackett A., Melnik S. and Gleeson J., *Phys. Rev. E*, **83** (2011) 056107.

[23] Borge-Holthoefer J., Rivero A., García I., Cauhé E., Ferrer A., Ferrer D., Francos D., Iñiguez D., Pérez M., Ruiz G. *et al.*, *PLoS ONE*, **6** (2011) e23883.

[24] Mirollo R. and Strogatz S., *SIAM J. Appl. Math.*, **50** (1990) 1645.

[25] Corral A., Pérez-Vicente, Díaz-Guilera and Arenas A., *Phys. Rev. Lett.*, **74** (1995) 118.

[26] Timme M., Wolf F. and Geisel T., *Phys. Rev. Lett.*, **89** (2002) 258701.

[27] Roxin A., Riecke H. and Solla S. A., *Phys. Rev. Lett.*, **92** (2004) 198101.

[28] Golder S. A. and Macy M. W., *Science*, **333** (2011) 1878.

[29] Molloy M. and Reed B., *Random Struct. Algorithms*, **6** (1995) 161.

[30] Molloy M. and Reed B., *Combin. Probab. Comput.*, **7** (1998) 295.

[31] Dorogovtsev S., Goltsev A. and Mendes J., *Rev. Mod. Phys.*, **80** (2008) 1275.

[32] http://cosnet.bifi.es/research-lines/online-social-systems/15m-dataset.

[33] Garlaschelli D. and Loffredo M. I., *Phys. Rev. Lett.*, **93** (2004) 268701.