Competition-Induced Criticality in a Model of Meme Popularity

James P. Gleeson,[1*] Jonathan A. Ward,[2] Kevin P. O'Sullivan,[1] and William T. Lee[1]
[1]MACSI, Department of Mathematics and Statistics, University of Limerick, Limerick, Ireland
[2]Centre for the Mathematics of Human Behaviour, Department of Mathematics and Statistics, University of Reading, Whiteknights RG6 6AH, United Kingdom
(Received 31 May 2013; revised manuscript received 17 November 2013; published 30 January 2014)

Heavy-tailed distributions of meme popularity occur naturally in a model of meme diffusion on social networks. Competition between multiple memes for the limited resource of user attention is identified as the mechanism that poises the system at criticality. The popularity growth of each meme is described by a critical branching process, and asymptotic analysis predicts power-law distributions of popularity with very heavy tails (exponent $\alpha < 2$, unlike preferential-attachment models), similar to those seen in empirical data.

DOI: 10.1103/PhysRevLett.112.048701 PACS numbers: 89.65.-s, 05.65.+b, 89.75.Fb, 89.75.Hc

When people select from multiple items of roughly equal value, some items quickly become extremely popular, while other items are chosen by relatively few people [1]. The probability $P_n(t)$ that a random item has been selected n times by time t is often observed to have a heavy-tailed distribution (n is called the popularity of the item at time t). In examples where the items are baby names [2], apps on Facebook [3], retweeted URLs or hashtags on Twitter [4–6], or video views on YouTube [7], the popularity distribution is found to scale approximately as a power law $P_n \sim n^{-\alpha}$ over several decades. The exponent α in all these examples is less than two, and typically has a value close to 1.5. This range of α values is notably distinct from those obtainable from cumulative-advantage or preferential-attachment models of the Yule-Simon type—as used to describe power-law degree distributions of networks, for example [8–11]—which give $\alpha \geq 2$. Interestingly, the value $\alpha = 1.5$ is also found for the power-law distribution of avalanche sizes in self-organized criticality (SOC) models [12,13], suggesting the possibility that the heavy-tailed distributions of popularity in the examples above are due to the systems being somehow poised at criticality.

In this Letter, we present an analytically tractable model of selection behavior, based on simplifying the model of Weng *et al.* [14] for the spreading of memes on a social network. We show that, in certain limits, the system is automatically poised at criticality—in the sense that meme popularities are described by a critical branching process [15]—and that the criticality can be ascribed to the competition between memes for the limited resource of user attention. We dub this mechanism "competition-induced criticality" (CIC) and investigate the impact of the social network topology (degree distribution) and the age of the memes upon the distribution of meme popularities. We show that CIC gives rise to heavy-tailed distributions very similar to the distributions of avalanche sizes in SOC models [16,17], even though our competition mechanism is quite different from the sandpile paradigm of SOC. This Letter may, therefore, be of interest in other areas where SOC-like critical phenomena have been observed in experiments or simulations, such as economic models of competing firms [18,19], the evolution and extinction of competing species [20–22], and neural activity in the brain [23,24].

For clarity, we will phrase the model in terms of meme diffusion as in [14], but the same understanding of the basic mechanism—and the analytical techniques for time-dependent distributions—can also be applied to other models, including the random-copying popularity models of [2,25,26]. The role of competition among items for limited resources has been examined from many viewpoints: see, for example, [27–29] and also related work on competing diseases [30–32]. The distribution of popularity increments (number of selections of an item in a small time interval) in Moran-type models has been obtained analytically [26]; however, our focus is on the (time-dependent) distributions of popularity accumulated over long time scales.

We consider a model of a directed social network, like Twitter, where nodes represent users; there are N nodes and we will take the limit $N \to \infty$ in our analysis. A randomly chosen user has k followers (i.e., out-degree k, note we use the convention that network edges are directed from nodes to their followers) with probability p_k. Each node has a screen, which holds the meme of current interest to that node (see Fig. 1). For simplicity, we assume, here, that each screen has capacity for only one meme, though this case is easily extended [33]. During each time step (with time increment $\Delta t = 1/N$), one node is chosen at random. With probability μ, the selected node innovates, i.e., generates a brand-new meme, that appears on its screen, and is tweeted (broadcast) to all the node's followers. Otherwise (with probability $1 - \mu$), the selected node (re)tweets the meme currently on its screen (if there is one) to all its followers, and the screen is unchanged. If there is no meme on the node's screen, nothing happens. When a meme m is tweeted, the popularity of meme m is incremented by 1 and the memes currently on the followers' screens are overwritten by meme m.

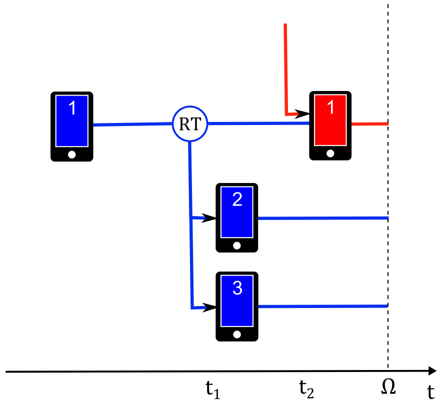

FIG. 1 (color online). Schematic of the model. Time runs horizontally and nodes of the network are listed vertically; the screen color of each node indicates the meme it currently holds. At time t_1, node 1 retweets (RT) the blue meme to its followers (nodes 2 and 3). At time t_2, node 1's screen is overwritten by the red meme, which was tweeted by one of the nodes followed by node 1.

Two memes; initially no competition.—As a first examination of the model's dynamics, we consider just two memes (called red and blue), each of which is initially present on a small number of screens, with every other screen being empty, and with no innovation ($\mu = 0$). A simple mean-field analysis of this two-meme case gives some useful insight. We assume all nodes have z followers, and follow z others, where z is the mean out-degree $\sum_k k p_k$ of the network. Let $r(t)$ be the fraction of screens occupied by the red meme at time t, with $b(t)$ the corresponding fraction of blue-meme screens. Since nodes are selected at random to tweet, the expected popularity (i.e., the cumulative number of tweets up to time t) for the red meme, $n_r(t)$, is related to $r(t)$ by $dn_r/dt = r(t)$, with a similar relation for the blue meme. Under the mean-field assumptions, a deterministic approximation for $r(t)$ and $b(t)$ is given by the solution of the pair of equations

$$\frac{dr}{dt} = -zbr + zr(1-r), \qquad \frac{db}{dt} = -zbr + zb(1-b). \quad (1)$$

The first term on the right-hand side of the first equation, for example, accounts for a decrease in the number of red-meme screens due to memes being overwritten by blue-tweeting nodes. This occurs when a blue meme is tweeted [with probability $b(t)$ in a given time step], and affects a fraction $r(t)$ of the z followers of the tweeting node, giving the term $-zbr$. The second term describes the growth of red memes due to a red meme tweeting [with probability $r(t)$] to nonred followers, the expected number of which is $z[1-r(t)]$.

Equations (1) can be solved analytically: the fraction of nonempty screens is $i(t) = r(t) + b(t)$, with $i(0) \ll 1$, and its dynamics obey the logistic differential equation $di/dt = zi(1-i)$, which is precisely the mean-field approximation for the infected population fraction in a susceptible-infected epidemic model. When $r(t)$ and $b(t)$ are both very small, the solutions show exponential growth in screen occupation and, hence, in the accumulated tweets (i.e., popularities) $n_r(t)$ and $n_b(t)$—see Fig. 2(a)—similar to early-stage growth of independent diseases [34]. The exponential growth continues until $i(t)$ is of order 1, by which time most screens show either the red or the blue meme. When $r(t) + b(t) = 1$, the right-hand sides of Eqs. (1) are both zero. This means that—under the mean-field assumptions that give this deterministic limit—the numbers of screens showing each meme remain constant thereafter, and so the popularities $n_r(t)$ and $n_b(t)$ grow linearly in time, as in Fig. 2(b). This balance is a dynamic one, as the two memes continue to compete for the resource of screen space, but the rate of growth for each meme is precisely equal to the rate of loss due to being overwritten by the other meme. Thus, the linear growth in popularity is induced by the competition between memes, in contrast to the exponential growth at earlier times [Fig. 2(a)] when the memes were not competing for the same resources [7,29].

The mean-field approximation used above ignores finite-N effects, which cause stochastic fluctuations in the number of screens about the mean values $r(t)$ and $b(t)$. In the long-time limit, it is these fluctuations that eventually lead to one meme becoming extinct, with the other filling

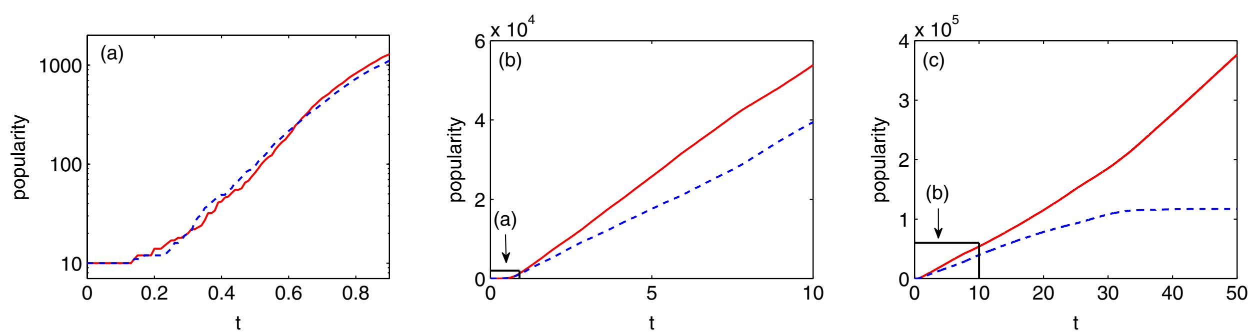

FIG. 2 (color online). Popularities $n_r(t)$ and $n_b(t)$ of red and blue memes in a single realization, starting from ten screens each. Note the different time scales in each figure, and the fact that the vertical scale in (a) is logarithmic. The inset boxes in (b) and (c) show the area of the previous figures.

all screens [as in Fig. 2(c)]. Stochastic fluctuations are also important at early times, when there are only very few screens showing either meme. In order to model the important role of stochastic fluctuations, and also to examine how the results presented here extend to cases with very many memes, we next consider a heavily competitive environment containing multiple memes.

Multiple competing memes.—Now, suppose that there are no empty screens in the network—so we are in the highly competitive regime—and the innovation probability μ may be nonzero. Competition between memes for the limited resource of user attention (i.e., screen space) leads naturally to some memes becoming extremely popular, while others are only moderately popular, or are ignored. We show that the model produces fat-tailed distributions of popularity, which are power-law distributions in the limit $\mu \to 0$. This is explained using a branching process description of the model, where the competitive environment causes each meme to follow a critical branching process (for which power-law distributions are expected [16,27]).

The branching process description is strictly valid only when the number of screens occupied by a single meme is a small fraction of N, but we note that this is the case for long epochs of time in a competitive environment with many memes. We assume here that all nodes follow (approximately) z other nodes, so the in-degree distribution is homogeneous, but we consider heterogeneous distributions of out-degrees. Before examining the details of the branching process, it is worth highlighting the source of criticality in the model when $\mu = 0$. In a single time step Δt, a tweeting node creates (or "gives birth to") an average of $z\Delta t$ new copies of the meme on its screen by overwriting the screens of its followers. However, each screen can be overwritten by another meme (causing "death" of the overwritten meme) with probability $z\Delta t$, and so the birth and death rates of memes are, on average, exactly balanced, giving a critical branching process. This balance between births and deaths remains critical when the model is enhanced in several ways, including modifying the rules so that nodes retweet any given meme at most once, see Sec. S5 of [33].

Next, we give details of the branching process description of the model. We denote the distribution of popularities at age a by $q_n(a)$: this is the probability that a meme has been tweeted n times when its age is a (i.e., at a time $t_b + a$, where t_b is the birth time of the meme). This distribution can be represented via its probability generating function (PGF) [35,36] $H(a, x)$, defined by $H(a, x) \equiv \sum_{n=1}^{\infty} q_n(a)x^n$. The network topology is described by the PGF for the out-degree distribution: $f(x) \equiv \sum_{k=0}^{\infty} p_k x^k$. The mean degree is $z = f'(1)$ and we assume all nodes have in-degree z.

To calculate $q_n(a)$, we first find $H(a, x)$ and then employ an inversion technique based on fast Fourier transforms (FFTs) [31,37,38]. It proves convenient to introduce $G(a, x)$, defined as the PGF for the excess popularity distribution at age a of memes that originate from a single randomly chosen screen (the root of the tree). The tweet event that creates the root is not counted by G: this event increases the popularity of the meme by 1, and places the meme upon the root screen and the screens of all followers of the root node. Consequently, the PGF for the popularity of age-a memes is given by $H(a, x) = xG(a, x)f(G(a, x))$. In Sec. S1 of [33], we derive the following ordinary differential equation for $G(a, x)$, parametrized by x:

$$\frac{\partial G}{\partial a} = z + \mu - (z + 1)G + (1 - \mu)xGf(G). \qquad (2)$$

This equation is easily solved using standard numerical methods, starting from the initial condition $G(0, x) = 1$. Some analysis is also possible [33]: the mean popularity $\partial H/\partial x(a, 1)$, for example, grows linearly with age until a is of the order $1/\mu(z + 1)$; thereafter, it saturates at a value of $1/\mu$. By expanding $G(a, x)$ as a Taylor series about $x = 0$, the probabilities $q_n(a)$ for low n may be determined explicitly. The popularity distribution for larger values of n are determined in a computationally efficient manner using FFTs [31,37,38]: our implementation (Sec. S2 of [33]) determines probabilities q_n for n values up to several thousand, shown as black curves in Fig. 3 [39]. The colored symbols are the results of stochastic simulations of the

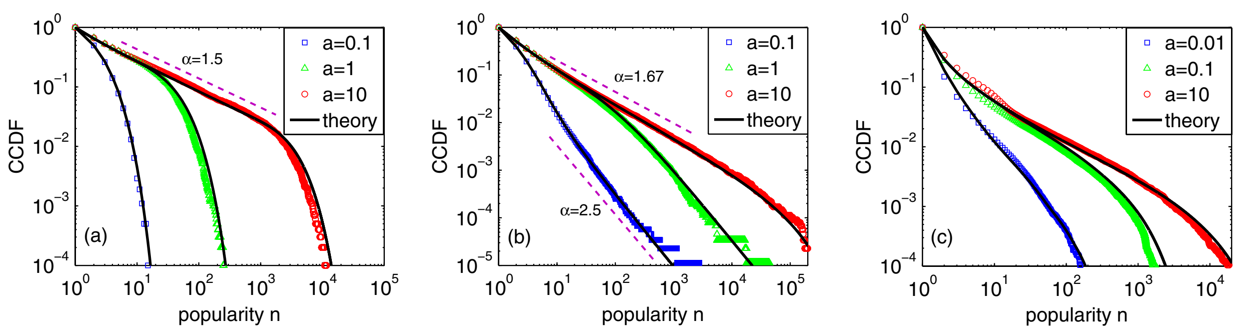

FIG. 3 (color online). Complementary cumulative distribution functions (CCDFs)—the fraction of memes with popularity $\geq n$—for numerical simulations, compared with the theory of Eq. (2). Dashed lines correspond to CCDFs for power law popularity distributions $P_n \propto n^{-\alpha}$. (a) $p_k = \delta_{k,10}$, $N = 10^5$, $\mu = 0$. (b) $p_k \propto k^{-\gamma}$ for $k \geq 4$ with $\gamma = 2.5$ (mean degree $z = 10.6$), $N = 10^6$, $\mu = 0.01$. (c) Twitter network of the Spanish 15M movement [40,41], $N = 87, 559$, $z = 69$, $\mu = 0.05$.

model, giving popularity distributions for memes at various ages a. The match between theory and simulation is very good. Figure 3(a) shows the popularity distributions on networks where each node has exactly $z = 10$ followers, while Fig. 3(b) is for a network where the number of followers (out-degree of a node) has a power-law distribution: $p_k \propto k^{-\gamma}$ for $k \geq 4$, with $\gamma = 2.5$ (and $p_k = 0$ for $k < 4$). In both cases, the k followers of a given node are assigned at random, so the in-degree distributions are Poisson.

The observed power-law popularity distributions can be understood using the long-time (or "old-age", $a \to \infty$) asymptotics of the branching process, found by analyzing the limiting solutions of Eq. (2) in the complex x plane (Sec. S3 of [33]); we summarize the main results as follows. If the out-degree distribution p_k has a finite second moment [i.e., if $f''(1) < \infty$], then the $a \to \infty$ limit of the popularity distribution has the asymptotic form $q_n(\infty) \sim A n^{-(3/2)} e^{-(n/\kappa)}$ as $n \to \infty$, where $\kappa = 2[f''(1) + 2z]/\mu^2(z+1)^2$ and $A = (z+1)[2\pi\{f''(1) + 2z\}]^{-(1/2)}$. This formula shows that the popularity distribution is of power-law form $n^{-3/2}$, up to an exponential cutoff at $n \approx \kappa$. However, the cutoff size κ limits to infinity as the innovation rate μ goes to zero, and κ can be large even for nonzero μ if the second moment of the distribution p_k is large [since $f''(1) = \sum_k k(k-1)p_k$].

If $p_k \propto k^{-\gamma}$ for large k with $2 < \gamma < 3$, then $f''(1)$ is infinite, and a different asymptotic analysis is required. In this case, we find, similar to [16], that as $n \to \infty$,

$$q_n(\infty) \sim \begin{cases} B n^{-(\gamma/\gamma-1)} & \text{if } \mu = 0\,, \\ C n^{-\gamma} & \text{if } \mu > 0, \end{cases} \tag{3}$$

with prefactors B and C given in [33]. Thus in the zero-innovation limit, the popularity distribution has a power-law exponent $\gamma/(\gamma - 1)$ that is smaller than the exponent γ of the out-degree distribution.

Figure 3(c) compares theory and simulation results for the model on the real Twitter network of [40,41]. The theory matches the simulation results rather well, despite the fact that this network is not treelike—indeed, 44% of links are reciprocal links—and does not have a homogeneous in-degree distribution, as assumed in the derivation of the theory. The accuracy of results from tree-based theories applied to real-world networks has been noted previously [42] and is examined further for this model in Sec. S4 of [33].

Conclusions.—We have used a simple model of meme diffusion to illustrate the phenomenon of competition-induced criticality. It is straightforward to generalize the basic model and the derivation of Eq. (2)—for example, by (i) increasing the capacity of screens to $c > 1$ memes, (ii) allowing followers to reject a meme tweeted to them with some probability so it does not appear on their screen, or (iii) permitting nodes to retweet a meme at most once—and to show that the CIC property is retained in the more general cases [33]. Despite their simplicity, we believe that the understanding of such analytically tractable models provides important insights on the origin of regularities observed in empirical data. For instance, our model does not include fat-tailed distributions of in-degrees, user activity levels, or response times [4,14,43]—these will be added in future work—but it can, nevertheless, produce fat-tailed popularity distributions.

We have also shown that the CIC model produces avalanches of popularity whose sizes have the same steady-state distributions as those found in sandpile models of self-organized criticality [16]. This is intriguing because the CIC mechanism is quite distinct from the sandpile paradigm: nodes do not have thresholds for triggering avalanches, for example, and the CIC popularity avalanches evolve on the same time scale as the general dynamics. We speculate that competition for limited resources may, therefore, play an important role in many other application areas where SOC-like phenomena have been identified in experiments or numerical simulations [18–24].

This Letter was partially funded by Science Foundation Ireland (Grants No. 11/PI/1026 and No. 09/SRC/E1780), the Engineering and Physical Sciences Research Council (MOLTEN, Grant No. EP/I016058/1) and by the FET-Proactive project PLEXMATH. We thank Peter Fennell for assistance with Fig. 3(c) and acknowledge helpful discussions with D. Cellai, S. Melnik, M. A. Porter, J-P Onnela, and F. Reed-Tsochas. We acknowledge the SFI/HEA Irish Centre for High-End Computing (ICHEC) for the provision of computational facilities, and the COSNET Lab for publishing the 15M dataset.

*james.gleeson@ul.ie

[1] R. A. Bentley, M. Earls, and M. J. O'Brien, *I'll Have What She's Having: Mapping Social Behavior* (MIT Press, Cambridge, MA, 2011).
[2] R. A. Bentley, M. W. Hahn, and S. J. Shennan, Proc. R. Soc. B **271**, 1443 (2004).
[3] J.-P. Onnela and F. Reed-Tsochas, Proc. Natl. Acad. Sci. U.S.A. **107**, 18375 (2010).
[4] K. Lerman, R. Ghosh, and T. Surachawala, arXiv:1202.3162.
[5] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts, in Proceedings of the Fourth ACM International Conference on Web Search and Data Mining (ACM, New York, 2011), pp. 65–74.
[6] R. Baños, J. Borge-Holthoefer, and Y. Moreno, arXiv:1303.4629.
[7] G. Szabo and B. A. Huberman, Commun. ACM **53**, 80 (2010).
[8] S. Redner, Eur. Phys. J. B **4**, 131 (1998).
[9] M. E. J. Newman, Contemp. Phys. **46**, 323 (2005).
[10] M. V. Simkin and V. P. Roychowdhury, Phys. Rep. **502**, 1 (2011).
[11] A. L. Barabási and R. Albert, Science **286**, 509 (1999).
[12] S. Zapperi, K. B. Lauritsen, and H. E. Stanley, Phys. Rev. Lett. **75**, 4071 (1995).

[13] P. Bak, *How Nature Works: The Science of Self-Organized Criticality* (Springer, New York, 1999).
[14] L. Weng, A. Flammini, A. Vespignani, and F. Menczer, Sci. Rep. **2**, 335 (2012).
[15] T. E. Harris, *The Theory of Branching Processes* (Dover, New York, 2002).
[16] K. I. Goh, D. S. Lee, B. Kahng, and D. Kim, Phys. Rev. Lett. **91**, 148701 (2003).
[17] P.-A. Noël, C. D. Brummitt, and R. M. D'Souza, Phys. Rev. Lett. **111**, 078701 (2013).
[18] R. E. Krider and C. B. Weinberg, Geogr. Anal. **29**, 16 (1997).
[19] A. Arenas, A. Díaz-Guilera, C. J. Pérez, and F. Vega-Redondo, J. Econ. Dyn. Control **26**, 2115 (2002).
[20] J. Sprott, Phys. Lett. A **325**, 329 (2004).
[21] M. E. J. Newman, Proc. R. Soc. B **263**, 1605 (1996).
[22] R. V. Solé and S. C. Manrubia, Phys. Rev. E **54**, R42 (1996).
[23] A. Haimovici, E. Tagliazucchi, P. Balenzuela, and D. R. Chialvo, Phys. Rev. Lett. **110**, 178101 (2013).
[24] M. Rubinov, O. Sporns, J.-P. Thivierge, and M. Breakspear, PLoS Comput. Biol. **7**, e1002038 (2011).
[25] R. A. Bentley, P. Ormerod, and M. Batty, Behav. Ecol. Sociobiol. **65**, 537 (2011).
[26] T. S. Evans and A. D. K. Plato, Phys. Rev. E **75**, 056101 (2007).
[27] C. Adami and J. Chu, Phys. Rev. E **66**, 011907 (2002).
[28] M. Beguerisse Díaz, M. A. Porter, and J.-P. Onnela, Chaos **20**, 043101 (2010).
[29] F. L. Forgerini, N. Crokidakis, S. N. Dorogovtsev, and J. F. F. Mendes, International Journal of Complex Systems in Science **1**, 141 (2011).
[30] B. Karrer and M. E. J. Newman, Phys. Rev. E **84**, 036106 (2011).
[31] P.-A. Noël, A. Allard, L. Hébert-Dufresne, V. Marceau, and L. J. Dubé, Phys. Rev. E **85**, 031118 (2012).
[32] J. C. Miller, Phys. Rev. E **87** 060801(R) (2013).
[33] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.112.048701 for details of mathematical derivations and extensions of the model.
[34] A. Barrat, M. Barthelemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, England, 2008).
[35] H. S. Wilf, *Generatingfunctionology* (A. K. Peters, Wellesley, MA, 2006).
[36] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).
[37] J. K. Cavers, IMA J. Appl. Math. **22**, 275 (1978).
[38] M. Marder, Phys. Rev. E **75**, 066103 (2007).
[39] Octave/Matlab codes for solving Eq. (2) and inverting the generating functions are available for download from http://www.ul.ie/sdcs/people/kevin-osullivan.
[40] J. Borge-Holthoefer *et al.*, PLoS One **6**, e23883 (2011).
[41] S. González-Bailón, J. Borge-Holthoefer, A. Rivero, and Y. Moreno, Sci. Rep. **1**, 197 (2011).
[42] J. P. Gleeson, S. Melnik, J. A. Ward, M. A. Porter, and P. J. Mucha, Phys. Rev. E **85**, 026106 (2012).
[43] J. L. Iribarren and E. Moro, Phys. Rev. E **84**, 046116 (2011).