Determinants of Meme Popularity

James P. Gleeson¹, Kevin P. O'Sullivan¹, Raquel A. Baños², and Yamir Moreno^{2,3}

¹MACSI, Department of Mathematics and Statistics, University of Limerick, Ireland ²Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), Universidad de

Zaragoza, Mariano Esquillor s/n, 50018 Zaragoza, Spain

³Department of Theoretical Physics, Faculty of Sciences, University of Zaragoza,

Zaragoza 50009, Spain

³Institute for Scientific Interchange (ISI), Turin, Italy

23 Jan 2015

Abstract

Online social media have greatly affected the way in which we communicate with each other. However, little is known about what are the fundamental mechanisms driving dynamical information flow in online social systems. Here, we introduce a generative model for online sharing behavior and analytically show, using techniques from mathematical population genetics, that competition between memes for the limited resource of user attention leads to a type of self-organized criticality, with heavy-tailed distributions of meme popularity: a few memes "go viral" but the majority become only moderately popular. The time-dependent solutions of the model are shown to fit empirical micro-blogging data on hashtag usage, and to predict novel scaling features of the data. The presented framework, in contrast to purely empirical studies or simulation-based models, clearly distinguishes the roles of two distinct factors affecting meme popularity: the memory time of users and the connectivity structure of the social network.

Recent advances in communication technologies and the emergence of social media have made it possible to communicate rapidly on a global scale. However, since we receive pieces of information from multiple sources, this has also made the information ecosystem highly competitive: in fact, users' influence and visibility are highly heterogeneous and memes or topics strive for users' attention in online social systems. Although several studies have described the dynamics of information flow in popular communication media [1–5], the main factors determining the observed patterns have not been identified and there is no theoretical framework that addresses this challenge. Indeed, given the potential for applications—e.g., having more efficient systems to spread information for safety and preparedness in the face of threats—a better understanding of how memes (ideas, hashtags, etc.) emerge and compete in online social networks is critical.

To address this problem, we develop a theoretical framework that describes how users choose among multiple sources of incoming information and affect the spreading of memes on a directed social network, like Twitter [1-3]. Our probabilistic model, in contrast to other studies [3-7] that



Figure 1: Schematic of the model. (A) Timeline of users' actions in a typical realization of the model. User A is followed by users B and C; arrows between nodes denote the direction of information transmission. Note that user B also follows many other users, and so his stream contains more memes than the streams of A or C. At time t_R^A , user A retweets a previously-seen meme (with probability $1-\mu$, given A is active). She chooses the red meme to retweet, by looking backwards in her stream a distance determined by the memory-time distribution Φ (only memes that A deemed "interesting" are shown in her stream). Her retweet of the red meme is accepted as "interesting" (and so inserted into their stream) by each follower of A with probability λ . At time t_R^C user C retweets the red meme to his followers, so further increasing the popularity of the red meme. At time t_I^A user A innovates (a probability μ event, given A is active) by inventing the new blue meme and broadcasting it to her followers. (B) Branching process representation (Sec. S2) of the popularities of the red meme and of the blue meme. Each retweet generates new branches of the process, as the meme is inserted into the streams of followers of the tweeting user.

are based on intensive computational simulations to fit to data, allows us to get analytical insights into the respective roles of the network degree distribution, the memory-time distribution of users, and the omnipresent competition between memes for the limited resource of user attention. Using this analysis, we fit the model to hashtag popularities extracted from micro-blogging data, and predict features of the time-dependent data.

Our model is as follows. In online communication platforms like Twitter, users follow (receive the broadcasts or "tweets" of) other users. In graph-theoretical terms, these relationships constitute directed links from the followed node (user) to the follower (Fig. 1) and the network composed of all users is characterized by an out-degree distribution p_k (p_k =probability that a randomly-chosen user has k followers). For simplicity, let us first assume that each user follows, on average, z others, where $z = \sum_k k p_k$ is the mean degree of the network; the general case in which this assumption is relaxed is addressed in the supplementary materials (SM). Additionally, each user has a "stream" that records all tweets she receives, time-stamped by their arrival time. We assume that only a fraction λ of the tweets received are deemed "interesting" by the user, and only the interesting tweets are considered for possible retweeting by that user. (Here we use the term "retweeting" in a general sense, to include any reuse of a previously-received meme such as a hashtag). The activity rate of a user—the average number of tweets that she sends per unit time—can depend on how well-connected the user is within the social network [3], but here we describe only the simplest case of homogeneous activity rates (for the general case, see SM), so that each user sends a tweet, on average, once per time unit (i.e., as a Poisson process with rate 1). When a user decides, at time t, to send a tweet, she has two options (see Fig. 1): with probability μ , the user innovates, i.e., invents a new meme, and tweets this new meme to all followers. The new meme appears in the user's own stream (it is automatically interesting to the originating user), and in the streams of all her followers, where it may be deemed interesting by each follower, independently, with probability λ . If not innovating (with probability $1 - \mu$), the user instead chooses a meme from her stream to retweet. The meme for retweeting is chosen by looking backwards in time an amount t_m determined by a draw from the memory-time distribution $\Phi(t_m)$, and finding the first interesting meme in her stream that arrived prior to the time $t-t_m$. The retweeted meme then appears in the streams of the user's followers (time-stamped as time t), but because it is a retweet, it does not appear a second time in the stream of the tweeting user. The popularity n of a meme is the number of times it has been tweeted or retweeted; this quantity depends on the age a of the meme (the time since it was first tweeted) [7].

The model as described is a "neutral model" [8, 9] in the sense that all memes have the same "fitness" [10]: no meme has an inherent advantage in terms of its attractiveness to users. Nevertheless, the competition between memes for the limited resource of user attention causes initial random fluctuations in popularities of memes to be amplified, and leads to popularity distributions with very heavy tails [11]: heavier, for example, than can be generated by models of preferential attachment or cumulative advantage type [12–16]. This "competition-induced criticality" was studied for a zero-memory ($\Phi(t_m) = \delta(t_m)$) version of this model in Ref. [17]; numerical simulation results for a closely related model were first reported in Ref. [3].

A branching process approximation for the model (SM), enables us to understand how the network structure (via the out-degree distribution p_k) and the users' memory-time distribution $(\Phi(t_m))$ affect the popularity distribution of memes. Defining $q_n(a)$ as the probability that a meme has popularity (total number of (re)tweets) n at age a, the branching process provides



Figure 2: Numerical simulations of the model, compared with analytical results. (A) Complementary cumulative distribution functions (CCDFs) for meme popularity at age a: numerical simulation results (black) on a network with scale-free out-degree distribution $(p_k \propto k^{-\gamma})$ for $k \ge 4$ with $\gamma = 2.5$, mean degree z = 11, $N = 10^5$ nodes), compared with asymptotic model result Eq. S1 (colored curves). The memory-time distribution is $\Phi = \text{Gamma}(k_G, \theta)$ with $k_G = 0.1$ and $\theta = 50$, so the mean memory time is $T = k_G \theta = 5$. Inset: As main, but for Poisson outdegree distribution p_k (z = 11) and gamma memory-time distribution with $k_G = 0.1$ and $\theta = 0.5$. (B) Fraction $q_1(a)$ of memes that are not retweeted by age a, on the scale-free network of (A) and for various memory-time distributions $\Phi(t_m)$ (red = exponential with mean T, blue/green = Gamma(0.1, 10T)), using Eq. S2. Dashed lines show the T = 5 cases; solid lines represent T = 1. (C) Mean popularity of memes of age a, for the same cases as in (B), and compared with Eq. 1 (using the numerical Laplace transform inversion described in Sec. S4); inset shows the large-abehavior. All panels have $\mu = 0.02$ and (except for green curves) $\lambda = 1$.

analytical expressions that determine the probability generating function (PGF) of the popularity distribution, $H(a; x) = \sum_{n=1}^{\infty} q_n(a)x^n$. In the small-innovation limit $\mu \to 0$, the model describes a critical branching process [18,19], for which power-law distributions of popularity (avalanche size) are expected [20–22]. Although the analytical expressions are somewhat complicated, we can see the impact of p_k and $\Phi(t_m)$ by focusing on certain features of the distribution. For example, the fraction of memes that are not retweeted at age a, i.e., those whose popularity remains at n = 1until at least age a, is given by $q_1(a) = \lim_{x\to 0} H(a; x)/x$. The large-age asymptotic behavior of this quantity can be written explicitly in terms of the network out-degree distribution and the memory-time distribution (Sec. S2.3). In particular, $q_1(a)$ limits to a non-zero value as the age a tends to infinity, capturing the observed fact that the competition for user attention results in a substantial fraction $q_1(\infty)$ of memes being ignored (never retweeted) subsequent to their birth [1,2,23], see Fig. 2B.

The expected (mean) popularity of memes of age a, $m(a) = \sum_n nq_n(a) = \partial H/\partial x|_{x=1}$, can also be found explicitly, up to a Laplace transform inversion:

$$\hat{m}(s) = \frac{1}{s} + \frac{1-\mu}{s} \frac{(\lambda z+1)\hat{\Phi}(s)}{\lambda z+\mu+s-(1-\mu)\lambda z\hat{\Phi}(s)},$$
(1)

where hats denote Laplace transforms. Strikingly, unlike q_1 , the mean popularity depends on the network degree distribution p_k only through the mean degree z, implying that the mean popularity is independent of the finer details of the network structure. The large-age, small- μ asymptotics of m(a) may be inferred from Eq. (1): this analysis (Sec. S2.5) is valid for ages a with $a \gg T$, where $T = \int_0^\infty t_m \Phi(t_m) dt_m$ is the mean memory time. We find that m(a) grows approximately linearly with age a: $m(a) \approx \frac{\lambda z+1}{T\lambda z+1}a$, until ages of order $\frac{T\lambda z+1}{\mu(\lambda z+1)}$, whereupon m(a) limits to its $a \to \infty$ value of $1/\mu$. If the memory-time distribution has significant probability mass at low values of t_m (e.g., the gamma distribution in Fig. 2) then the mean popularity initially grows faster-than-linearly with age (see Fig. 2C).

Notably, the infinite-age limits of both $q_1(a)$ and m(a) are independent of the memory-time distribution $\Phi(t_m)$. This property is, in fact, shared by the entire popularity distribution in the $a \to \infty$ limit, which implies that the large-*n* asymptotic behavior can be determined using similar methods as for the zero-memory case in Ref. [17]. We find that if the out-degree distribution p_k of the network has a finite second moment, then the popularity distribution is a truncated power law with exponent 3/2: $q_n(\infty) \sim A e^{-n/\kappa} n^{-3/2}$ as $n \to \infty$ (see Sec. S2.6 for explicit expressions for the constants A and κ ; note the exponential cutoff κ limits to ∞ as μ tends to 0). However, if the network has a heavy-tailed degree distribution, i.e., $p_k \propto k^{-\gamma}$ with $2 < \gamma < 3$, then the infiniteage popularity distribution is also power law: $q_n(\infty) \sim B n^{-\gamma/(\gamma-1)}$ as $n \to \infty$ if $\mu \approx 0$. Note the exponent of the power law, $\gamma/(\gamma - 1)$, lies between 1.5 and 2, indicating that the popularity distribution is extremely heavily skewed.

The approach of the popularity distribution to its infinite-age limit, i.e., the behavior for finite but large ages, is of particular interest for real-world social networks that have $p_k \sim D k^{-\gamma}$ for large k. We focus especially on the case where γ is close to 2, as found for the out-degree distribution of many real-world social networks. In this case, we show (Sec. S2.7) that for large n and $\mu \approx 0$, the probability $q_n(a)$ grows approximately linearly with age a for large a, with a slope that is proportional to the growth rate of the mean popularity m(a). This result implies that the



Figure 3: Comparison of the model with Twitter hashtags data. (A) CCDFs for popularity of hashtags at age a (at time a after their first appearance in the dataset). The model CCDFs (from Eq. S1) are multiplied by 10 for clarity. Model parameters are: $\lambda = 4.5 \times 10^{-4}$, $\mu = 0.033$, $k_G = 0.25$, $\theta = 500$, with one model time unit corresponding to 0.16 days (B) CCDFs at age a, each divided by the mean popularity at age a. The data shows a collapse onto a single curve that is closely matched by the model. (C) The mean popularity of hashtags of age a. (D) The fraction $q_1(a)$ of hashtags that are not retweeted by age a. Here, our basic model with homogeneous user activity rates does not fit well to the data (red curve), but allowing for heterogeneous user activity rates improves the fit (blue curve), without compromising the other matches with data (see Fig. S5).

ratio $q_n(a)/m(a)$ is independent of age a, at least for values of a within the range of validity of the approximations. As a consequence, we expect that if we rescale the age-a popularity distribution by dividing by its mean m(a), then the rescaled distributions for different ages should collapse onto a single curve.

We tested this prediction on a 1-year dataset comprised of the popularities of 1.4×10^5 hashtags related to the 2011 15M protest movement in Spain [24,25]. We randomly sampled 8.2×10^5 users on Twitter to determine the distribution p_k of number of followers, and found $p_k \sim D k^{-\gamma}$ for large k, with $\gamma = 2.13$ (Fig. S2). Although the model parameters cannot be directly calculated from the dataset (Sec. S1), by assuming that the memory time distribution for users is a gamma distribution [19, 26], $\Phi = \text{Gamma}(k_G, \theta)$, Eqs. 1 and S1 enabled us to find model parameters λ , μ , k_G and θ that provide a good fit to the age-dependent mean and the old-age $(a \to \infty)$ distribution of hashtag popularities. Figure 3B demonstrates that the model and data both exhibit the predicted collapse of age-dependent distributions when scaled by their mean (as fitted in Fig. 3C). We show in the SM (Sec. S3) that the fit of the model to the observed $q_1(a)$ (Fig. 3D) can be improved by including realistic dependence of user activity rates upon the number of followers [3,27].

In summary, despite its simplicity, the model matches the empirical popularity distribution of hashtags on Twitter remarkably well; this is consistent with random-copying models of human decision-making [28] where the quality of the product—here, the "interestingness" of the meme—is less important than the social influence of peers' decisions [29]. The generalization of the model (as shown in the SM) to incorporate (i) heterogeneous user activity rates and (ii) a joint distribution p_{jk} of the number of users followed j and the number of followers k, remains analytically tractable and confirms the robustness of our main finding: that competition between memes for the limited resource of user attention induces criticality in the vanishing-innovation limit, giving power-law popularity distributions and epochs of linear-in-time popularity growth. We believe that the analytical results and potential for fast fitting to data will render this a useful null model for further investigations of the entangled effects of memory, network structure, and competition on information spread through social networks [30].

Acknowledgements

The authors thank K. Lerman, J. Ward, P. Fennell, D. O'Sullivan and A. Faqeeh for helpful discussions. This work was supported by Science Foundation Ireland (grant numbers 11/PI/1026, 12/IA/1683, and 09/SRC/E1780, J.P.G. and K.P.O'S.) and by the European Commission through FET-Proactive projects PLEXMATH (FP7-ICT-2011-8; grant number 317614, J.P.G and Y.M.) and MULTIPLEX (FP7-ICT-2011-8; grant 317532, Y.M.). R.A.B. and Y.M. also acknowledge support from MINECO (Grant FIS2011-25167) and Comunidad de Aragón (Spain; grant FENOL). We acknowledge the SFI/HEA Irish Centre for High-End Computing (ICHEC) for the provision of computational facilities. The data used in this paper is available for download from www.ul.ie/gleesonj/twitter15M.

SUPPLEMENTARY MATERIALS

Contents

S1 Further information on Figures 2 and 3	8
S2 Derivation and analysis of model equations	10
S2.1 Derivation of governing equations	10
S2.1.1 Description of the generalized model	11
S2.1.2 Derivation of the PGF equations	12
S2.2 Criticality of the branching process	16
S2.3 An explicit expression for $q_1(a)$	17
S2.4 Distribution of response times	18
S2.5 Mean popularity	18
S2.5.1 Large-age asymptotics of mean popularity	20
S2.6 Infinite-age limit of popularity distribution	20
S2.7 Large-a, large-n asymptotics of popularity distribution	21
S2.7.1 Analysis of the data collapse in Fig. 3B	24
S3 Extension to heterogeneous activity rates	25
S4 Numerical inversion of Laplace transforms and PGFs	27

S1 Further information on Figures 2 and 3

For the numerical simulation results in Fig. 2 of the main text, we generate configuration-model directed networks with prescribed out-degree distribution p_k . Each one of N users (nodes) is assigned a random number k (drawn from the distribution p_k) of out-links (links to followers). The identities of the k followers are chosen uniformly at random from the set of all users; in the $N \to \infty$ limit, this gives a Poisson in-degree distribution p_j which, for sufficiently large z, gives similar results to using the in-degree distribution $p_j = \delta_{j,z}$, i.e., assuming every user follows exactly z others [17].

The large-*a*, large-*n*, $\mu = 0$ asymptotics of the popularity distribution are determined (see derivation in Sec. S2.7) from the Laplace transform of a probability generating function (PGF), which is given (under the assumptions of the main text— $p_j = \delta_{j,z}$ and homogeneous user activity rates—and for networks with the scale-free out-degree distribution $p_k \sim D k^{-\gamma}$ as $k \to \infty$) by

$$\hat{H}(s;x) = \frac{1}{s} \left[1 - \frac{\lambda z \left(s + \lambda z + \hat{\Phi}(s)\right) (\gamma - 1)(1 - x)\hat{\Phi}(s)}{(s + \lambda z) \left(s + \lambda z - \lambda z \hat{\Phi}(s) + (\gamma - 1)\lambda D^{\frac{1}{\gamma - 1}} \left[\Gamma(1 - \gamma)\right]^{\frac{1}{\gamma - 1}} (1 - x)^{\frac{\gamma - 2}{\gamma - 1}} \hat{\Phi}(s) \right)} \right],\tag{S1}$$

where Γ is the gamma function. Inversion of this transform, as described in Sec. S4, gives the colored curves plotted in Fig. 2A. The corresponding expression for the case where p_k has finite



Figure S1: As panels B and C of Fig. 2 of the main text, but for a network with Poisson out-degree distribution (mean degree z = 11). As in Fig. 2, both panels have $\mu = 0.02$ and (except for green curves) $\lambda = 1$.

second moment (e.g., a Poisson out-degree distribution, as in the inset of Fig. 2A) is given by Eq. (S60).

Under the assumptions of the main text, the fraction $q_1(a)$ of memes that have not been retweeted by age a has the following large-a asymptotic behavior (see Sec. S2.3 for derivation):

$$q_1(a) \sim \frac{\lambda z + \mu}{\lambda z + \mu + (1 - \mu)C(a)} \sum_{k=0}^{\infty} p_k \left[1 - \lambda + \lambda \frac{\lambda z + \mu}{\lambda z + \mu + (1 - \mu)C(a)} \right]^k,$$
(S2)

where C(a) is the cumulative distribution function for memory times: $C(a) = \int_0^a \Phi(t_m) dt_m$. The results in Figs. 2B and 2C are specific to networks with the scale-free out-degree distribution $p_k = D k^{-\gamma}$ for $k \ge 4$ (and $p_k = 0$ for k < 4), with exponent $\gamma = 2.5$. Figure S1 shows the corresponding results for networks with a Poisson out-degree distribution (e.g., Erdös-Rényi directed graphs) with mean degree z matching that of the scale-free networks in Fig. 2. The age-dependence of $q_1(a)$ in Fig. S1A is qualitatively similar to that of Fig. 2B, but note that the limiting value as $a \to \infty$ is different in the two cases: using Eq. (S2) (with $C(\infty) = 1$) we obtain $q_1(\infty) = 0.50$ for the scale-free case with $\lambda = 1$, whereas $q_1(\infty) = 0.37$ for the Poisson network. In contrast, the mean popularity shown in Fig. S1B is essentially identical to that of Fig. 2C: this is because the mean popularity given by Eq. (1) of the main text depends only on the mean degree z of the network, which is the same in both the scale-free and Poisson networks.

The data plotted in Fig. 3 gives the popularity of 1.4×10^5 hashtags related to the 2011 15M protest movement in Spain, that were tracked over the 1-year period from March 2011 to March 2012 [24, 25]. In Fig. 3 we use all hashtags for which we have at least 200 days of data; each curve in Fig. 3A shows the popularity distribution for all hashtags which have the same age (to the nearest day). The model parameter λ and the memory-time distribution $\Phi(t_m)$ cannot be directly estimated from the data because in cases where users receive multiple copies of the same meme (hashtag) prior to retweeting it, it is impossible to tell which of received memes "caused"

the retweet¹. Therefore, we instead use the analytical results of the model (Eqs. (1) and (S1)) to find parameter values that fit the model to the statistical characteristics of the data, see Fig. 3.

The data does, however, provide an upper bound on the value of the innovation probability μ . Recall that μ is defined as the probability that a tweeted meme (hashtag) is an innovation, i.e., that the hashtag has never before appeared in the system. Each innovation event thus increases by one the number of distinct hashtags that appear in the dataset, whereas a non-innovative (copying) tweet will instead increase the number of copies of a hashtag that is already present in the dataset. We can therefore calculate an upper bound on the empirical innovation probability from the ratio

$$\tilde{\mu} = \frac{\text{number of distinct hashtags used in the dataset}}{\text{total number of hashtags tweeted by users}} = \frac{322799}{5886837} = 0.055.$$
(S3)

Note this upper bound is consistent with the parameter value of $\mu = 0.033$ used in Fig. 3. The reason why Eq. (S3) gives an upper bound rather than an exact value for μ is the finite size of the dataset: the data collection started at a specific point in time and so any hashtags that are in fact copied from tweets received prior to the start date will be erroneously counted as "distinct hashtags" in the estimate, thus leading to an overestimate of the true innovation probability.

To sample the out-degree distribution p_k of the Twitter network, we randomly selected 8.2×10^5 Twitter user ids and recorded the number of followers k of each user. The measured mean number of followers is z = 703, but the distribution p_k is heavy-tailed. The complementary cumulative distribution function (CCDF) of the k values is shown in Fig. S2, along with the line $D/(\gamma-1)k^{1-\gamma}$ with D = 240 and $\gamma = 2.13$ that corresponds to an out-degree distribution with tail scaling as $p_k \sim D k^{-\gamma}$ as $k \to \infty$ [31]. We use these values of D, γ and z in Eq. (S1) to produce the model results in Fig. 3A and 3B (using numerical inversion of the Laplace transformed PGF, see Sec. S4 for details).

S2 Derivation and analysis of model equations

In this section we derive the equations for the branching process approximation of the model, and use asymptotic methods to understand how the results depend on the parameters of the model.

S2.1 Derivation of governing equations

We begin by considering a somewhat more general setting for the model than that described in the main text, and we derive the equations for the branching process approximation in this general case; the specialization to the case considered in the main text will then be straightforward. In particular, we allow here for a more general network structure, and for heterogeneous user activity rates, while in the main text we specialize to networks described by only their out-degree distribution p_k and we assume all users have equal activity rates.

¹Note that the branching process approximation remains accurate in cases where multiple copies of a meme are received at different (well-separated) times. The approximation also works well on networks that are non tree-like, including those that contain a substantial fraction of reciprocal links, see Sec. S4 of [17].



Figure S2: CCDF for the number of followers k of a random sample of 8.2×10^5 Twitter users. The straight line corresponds to an out-degree distribution with tail scaling as $p_k \sim D k^{-\gamma}$ as $k \to \infty$, with D = 240 and $\gamma = 2.13$ ($x_{\min} = 1.1 \times 10^4$, fitted as described in [31]).

S2.1.1 Description of the generalized model

The network structure is defined by the joint probability p_{jk} that a randomly-chosen node (user) has in-degree j (i.e., follows j other Twitter users) and out-degree k (i.e., has k followers), but the network is otherwise assumed to be maximally random (a configuration model directed network). The mean degree of the network is $z = \sum_{j,k} kp_{jk} = \sum_{j,k} jp_{jk}$. For the special case of the main text, we assume all users follow z others, so p_{jk} can be replaced with $\delta_{j,z}p_k$, where $\delta_{j,z}$ is the Kronecker delta and p_k is the out-degree distribution.

Each user has a "stream" that records all tweets received by the user, time-stamped by their arrival time. We assume that only a fraction λ of the tweets received are deemed "interesting" by the user, and only the interesting tweets are considered for possible retweeting by that user. The activity rate of a user—the average number of tweets that she sends per unit time, i.e., the rate of the Poisson process that describes her tweeting activity—is, in the general case, assumed to depend on her in-degree j and out-degree k (her "(j, k)-class" for short); this assumption is supported by empirical evidence from Twitter, see Fig. 6 of [27]. The user activity rates β_{jk} give the relative activity levels of users in the (j, k) class; the rates are normalized so that $\sum_{jk} \beta_{jk} p_{jk} = 1$. If there are N users in the network, this rate implies that an average of N tweets are sent in each model time unit. For the description in the main text (except in Fig. 3D), we specialize to the case where all user activity rates are equal: $\beta_{jk} = 1$.

When a user decides, at time t, to send a tweet, she has two options (refer to Fig. 1 of the main text): with probability μ , the user innovates, i.e., invents a new meme, and tweets this new meme to all followers. The new meme appears in the user's own stream (it is automatically interesting to the originating user), and in the streams of all her followers (where it may be deemed interesting



Figure S3: Schematic for the derivation of the PGF equations, see Sec. S2.1.2. (A) The stream of user A, showing only memes that were deemed interesting by user A; each color represents a different meme. At time t, user A decides to retweet a meme from the past, and looks back to time r, where she finds meme M (colored red). She sends this meme to her followers (not shown); each follower independently deems the meme interesting with probability λ . Also shown is a later retweet event, which also copies meme M. (B) The retweet tree for meme M, seeded at time τ . Each retweet by user A of meme M generates a new branch on this tree; each branch can also generate further retweets by followers of A, these subtrees are denoted by squares. (C) Schematic depiction of Eqs. (S9) and (S17).

by each follower, independently, with probability λ). If not innovating (with probability $1-\mu$), the user chooses a meme from her stream to retweet. The meme for retweeting is chosen by looking backwards in time an amount t_m determined by a draw from the memory-time distribution $\Phi(t_m)$, and finding the first interesting meme in her stream that arrived prior to the time $t - t_m$. The retweeted meme then appears in the streams of the user's followers (time-stamped as time t), but because it is a retweet, it does not appear a second time in the stream of the tweeting user. The popularity n(a) of a meme is the total number of times it has been tweeted or retweeted by age a, i.e., by a time a after its first appearance (when it was tweeted as an innovation).

S2.1.2 Derivation of the PGF equations

We define $G_{jk}(\tau, \Omega; x)$ as the probability generating function (PGF) for the size of the "retweet tree", as observed at time Ω , that grows from the retweeting of a meme that entered, at time $\tau \leq \Omega$, the stream of a (j, k)-class user, see Fig. S3B. To obtain an equation for G_{jk} , we consider the stream of a random (j, k)-class user (called "user A") with a meme M that entered the stream at time τ (either by innovation, or because it was received from a followed user and deemed interesting by A), see Fig. S3A.

The likelihood that meme M is retweeted in the future depends on how quickly other tweets enter the stream of user A. In fact, meme M can be considered to "occupy" the stream for a time interval ℓ stretching from τ until the time $\tau + \ell$ when the next interesting meme enters the stream of user A. New memes enter the stream as a Poisson process at the constant rate²

$$r_{jk} = j\overline{\beta}\lambda + \mu\beta_{jk},\tag{S4}$$

so the occupation time ℓ of meme M—the time it occupies the stream of user A—is an exponentially distributed random variable with PDF

$$P_{\rm occ}(\ell) = r_{jk} \exp\left(-r_{jk}\ell\right). \tag{S5}$$

We note in passing that the mean occupation time

$$\langle \ell \rangle = \int_0^\infty \ell P_{\rm occ}(\ell) \, d\ell = \frac{1}{j\overline{\beta}\lambda + \mu\beta_{jk}} \tag{S6}$$

is, for small innovation probabilities μ , inversely proportional to j, the number of users followed. Thus, a user who follows many others experiences tweets entering his stream at a higher rate than a lower-j user (compare the streams of users B and C in the schematic Fig. 1). Consequently, the high-j user is less likely to see (and so to retweet) a given meme than a low-j user. This aspect of the model clearly reflects empirical data, see Fig. 3 of [32] for example.

To determine the size of trees originating from meme M, we consider that trees observed at a time $\Omega \geq \tau$ must be created by the retweeting by user A, at some time(s) between τ and Ω , via looking back in her stream to a time r, where r lies between τ and $\min(\tau + \ell, \Omega)$ (i.e., r lies within the time interval where meme M occupies the stream). Let's consider a time interval of (small) length dr, centered at time r, and calculate the size of trees that are seeded by a retweet based on a lookback into this interval, from a time t, with t > r, see Fig. S3. In each dt interval centered at time t, a tree will be seeded with probability³

$$P_{\text{seed}} = (1 - \mu)\beta_{jk}\Phi(t - r)\,dr\,dt,\tag{S7}$$

and will grow to a tree with size distribution (at observation time Ω) generated by⁴

$$R_k(t,\Omega;x) = x \left[1 - \lambda + \lambda G(t,\Omega;x)\right]^k, \qquad (S8)$$

²User A follows j users, each of which is assumed to tweet at the average rate $\overline{\beta} = \sum_{jk} \frac{k}{z} \beta_{jk} p_{jk}$. Each meme sent by these j users is deemed interesting by A with probability λ , so the rate at which interesting memes enter the stream of user A is $j\overline{\beta}\lambda$. Moreover, user A innovates at a rate $\mu\beta_{jk}$, which gives the second term of Eq. (S4). If either an incoming tweet or an innovation event occurs, a new meme is inserted into the stream of user A, and the occupation time of meme M is ended.

³The β_{jk} factor is the probability that that user acts; the $1 - \mu$ factor is for the user copying (retweeting) rather than innovating, and $\Phi(t-r) dr dt$ is the probability that the copying is from the specified dr-interval to the dt-interval.

⁴There are k followers of user A, each of whom may deem the tweet "uninteresting" with probability $1 - \lambda$, or consider it "interesting"—and accept it into their stream—with probability λ . The factor of x counts the increase in popularity due to the tweet event.

where

$$G(t,\Omega;x) = \sum_{j,k} \frac{j}{z} p_{jk} G_{jk}(t,\Omega;x)$$
(S9)

is the PGF for the sizes of trees originating from the successful insertion at time t of a meme (that is deemed interesting) into the stream of a random follower.

To calculate the total size of the tree seeded by copying from the dr-interval, we must add the sizes of trees that are copied into all times t with t > r. Since each copying event is independent, the total tree size is generated by

$$J(r;x) = \prod_{t=r}^{\Omega} \left[1 - P_{\text{seed}} + P_{\text{seed}} R_k(t,\Omega;x)\right].$$
(S10)

Taking logarithms of both sides of this equation and expanding to first order in dt gives

$$\log J = \sum_{t=r}^{\Omega} \log \left[1 - (1-\mu)\beta_{jk} \Phi(t-r) \, dr \, dt (1-R_k(t,\Omega;x)) \right]$$

$$\approx -(1-\mu)\beta_{jk} \sum_{t=r}^{\Omega} \Phi(t-r) \, dr \, dt (1-R_k(t,\Omega;x))$$

$$\to -(1-\mu)\beta_{jk} \, dr \int_r^{\Omega} \Phi(t-r) (1-R_k(t,\Omega;x)) \, dt \quad \text{as } dt \to 0,$$
(S11)

so J(r; x) can be written as

$$J(r;x) = \exp\left[-(1-\mu)\beta_{jk} \, dr \int_{r}^{\Omega} \Phi(t-r)(1-R_{k}(t,\Omega;x)) \, dt\right].$$
 (S12)

Recall that J(r; x) is the PGF for trees seeded by copying from time r. To obtain the total size of all children trees of meme M, we must consider trees seeded at all possible times r from τ to the time min $(\tau + \ell, \Omega)$ that marks the end of the occupation of user A's stream by meme M. Each drtime interval again independently generates trees with sizes distributed according to Eq. (S12), so the PGF for the total size is found by multiplying together copies of the J(r; x) function for each dr time interval, thus:

$$P_{\text{size}}(\ell) = \prod_{r=\tau}^{\min(\tau+\ell,\Omega)} J(r;x)$$

$$= \exp\left[-(1-\mu)\beta_{jk} \sum_{r=\tau}^{\min(\tau+\ell,\Omega)} dr \int_{r}^{\Omega} \Phi(t-r)(1-R_{k}(t,\Omega;x)) dt\right]$$

$$\rightarrow \exp\left[-(1-\mu)\beta_{jk} \int_{\tau}^{\min(\tau+\ell,\Omega)} dr \int_{r}^{\Omega} d\tau \, \Phi(t-r)(1-R_{k}(t,\Omega;x))\right] \quad \text{as } dr \to 0.$$
(S13)

Combining probabilities, by integrating over all possible occupation times ℓ , gives

$$G_{jk}(\tau,\Omega;x) = \int_0^\infty P_{\rm occ}(\ell) P_{\rm size}(\ell) \, d\ell \tag{S14}$$

and combining Eqs. (S5), (S9) and (S13) yields an integral equation for G:

$$G(\tau,\Omega;x) = \sum_{jk} \frac{j}{z} p_{jk} \int_0^\infty d\ell \left(j\overline{\beta}\lambda + \mu\beta_{jk} \right) \exp\left[-\left(j\overline{\beta}\lambda + \mu\beta_{jk} \right) \ell \right] \times \\ \times \exp\left[-(1-\mu)\beta_{jk} \int_0^{\min(\tau+\ell,\Omega)} dr \int_r^\Omega dt \, \Phi(t-r)(1-x\left[1-\lambda+\lambda G(t,\Omega;x)\right]^k) \right].$$
(S15)

Next, we note that $G(\tau, \Omega; x)$ depends only on the difference $\Omega - \tau$ that defines the age of the meme M at the observation time Ω . We therefore write $G(\tau, \Omega; x) = G(\Omega - \tau; x)$, which enables the change of variables $a = \Omega - \tau$, $\tilde{r} = r - \tau$, $\tilde{\tau} = \Omega - \tau$ to yield

$$G(a;x) = \sum_{jk} \frac{j}{z} p_{jk} \int_0^\infty d\ell \left(j\overline{\beta}\lambda + \mu\beta_{jk} \right) \exp\left[-\left(j\overline{\beta}\lambda + \mu\beta_{jk} \right) \ell \right] \times \\ \times \exp\left[-(1-\mu)\beta_{jk} \int_0^{\min(\ell,a)} d\tilde{r} \int_0^{a-\tilde{r}} d\tilde{\tau} \, \Phi(a-\tilde{r}-\tilde{\tau})(1-x\left[1-\lambda+\lambda G(\tilde{\tau};x)\right]^k) \right].$$
(S16)

The popularity of a meme, as observed at time Ω , that is seeded by a single tweet (e.g., by an innovation) at time τ may be calculated in a similar way to the derivation of Eq. (S16); the generating function is of the form

$$H(\tau, \Omega; x) = \sum_{j,k} \beta_{jk} p_{jk} R_k(\tau, \Omega; x) G_{jk}(\tau, \Omega; x),$$
(S17)

where $\beta_{jk}p_{jk}$ represents the probability that the seed tweet originates from a (j, k)-class user, R_k is the PGF for the trees generated from the followers of the user, and G_{jk} is the PGF for the size of the retweet-tree of the meme (see Fig. S3C). Introducing the age a of the meme as before and defining $q_n(a)$ as the probability that an age-a meme has popularity n, we have the PGF

$$H(a;x) = \sum_{n=1}^{\infty} q_n(a)x^n,$$
(S18)

which is given by

$$H(a;x) = \sum_{jk} \beta_{jk} p_{jk} x \left[1 - \lambda + \lambda G(a;x)\right]^k \int_0^\infty d\ell \left(j\overline{\beta}\lambda + \mu\beta_{jk}\right) \exp\left[-\left(j\overline{\beta}\lambda + \mu\beta_{jk}\right)\ell\right] \times \exp\left[-(1-\mu)\beta_{jk} \int_0^{\min(\ell,a)} d\tilde{r} \int_0^{a-\tilde{r}} d\tilde{\tau} \Phi(a-\tilde{r}-\tilde{\tau})(1-x\left[1-\lambda + \lambda G(\tilde{\tau};x)\right]^k)\right].$$
(S19)

S2.2 Criticality of the branching process

A branching process may be classified by the expected (mean) number ξ of "children" of each "parent": if this number (called the "branching number") is less than 1, the process is *subcritical* and if ξ is greater than 1 the process is *supercritical*. *Critical* branching processes, with an average of exactly one child per parent, give rise to power-law distributions of tree-sizes and of durations of growth cascades [20, 22]. Here we demonstrate that the general process derived in Sec. S1 is a critical branching process in the limit of vanishing innovation $\mu \to 0$.

We identify the "parent" in the process as a meme that was accepted into the stream (i.e., deemed interesting) of a (j, k)-class user at time τ : see, for example, meme M in the stream of user A, as shown in Fig. S3. The "children" of this meme are the retweets of it that are accepted into the streams of the followers of A at any time $t > \tau$. The PGF for the number of children of meme M is derived by following the same steps as in Sec. S2.1.2, but replacing R_k by $(1 - \lambda + \lambda x)^k$: each power of x then counts a successful insertion of meme M into the stream of one of the k followers of A. The resulting PGF, for a meme of age a, is (cf. Eq. (S15))

$$K_{jk}(a;x) = \int_0^\infty d\ell P_{\text{occ}}(\ell) \times \\ \times \exp\left[-(1-\mu)\beta_{jk} \int_0^{\min(\ell,a)} d\tilde{r} \int_0^{a-\tilde{r}} d\tilde{\tau} \,\Phi(a-\tilde{r}-\tilde{\tau})(1-[1-\lambda+\lambda x]^k)\right] \\ = \int_0^\infty d\ell P_{\text{occ}}(\ell) \exp\left[-(1-\mu)\beta_{jk}(1-[1-\lambda+\lambda x]^k) \int_0^{\min(\ell,a)} C(a-\tilde{r})d\tilde{r}\right], \quad (S20)$$

where $C(t) = \int_0^t \Phi(t_m) dt_m$ is the cumulative distribution function for memory times. The expected (mean) number of children for a meme in the (j, k)-class stream is determined from the PGF in the usual way, by differentiating with respect to x and evaluating at x = 1, thus:

$$\xi_{jk} = \left. \frac{\partial K_{jk}}{\partial x} \right|_{x=1}.$$
(S21)

In the limit of large ages, $a \to \infty$, we use the fact that $C(\infty) = 1$ to obtain

$$\xi_{jk} \sim (1-\mu)\beta_{jk}\lambda k \int_0^\infty \ell P_{\text{occ}}(\ell) \, d\ell \quad \text{as } a \to \infty$$
$$= \frac{(1-\mu)\beta_{jk}\lambda k}{j\overline{\beta}\lambda + \mu\beta_{jk}}.$$
(S22)

Averaging over all (j, k) classes, the effective branching number ξ of the process is the expected number of children of a meme that is accepted into the stream of a random follower:

$$\xi = \sum_{j,k} \frac{j}{z} p_{jk} \xi_{jk}$$

$$\rightarrow \sum_{j,k} \frac{j}{z} p_{jk} \frac{\beta_{jk} \lambda k}{j\overline{\beta}\lambda} = 1 \quad \text{as } \mu \to 0$$
(S23)

(recall that $\overline{\beta} \equiv \sum_{j,k} \frac{k}{z} \beta_{jk} p_{jk}$). Thus, we have shown that the branching process underlying the model is critical when $\mu = 0$. The occupation time of a meme in a users' stream is due to the competition between neutralfitness memes for the limited resource of user attention; this competition ensures that the mean number of successful retweets (children) generated during the finite occupation time of the meme is precisely one, and so induces the power-law distributions of cascade sizes that are characteristic of critical branching processes [20, 22].

An explicit expression for $q_1(a)$ S2.3

The value $q_1(a)$ is the probability that a meme, once created via an innovation event, is not retweeted by the time it reaches age a. This probability may be calculated explicitly using Eq. (S19):

$$q_{1}(a) = \lim_{x \to 0} \frac{H(a;x)}{x}, = \sum_{j,k} \beta_{jk} p_{jk} \left[1 - \lambda + \lambda G(a;0)\right]^{k} \int_{0}^{\infty} d\ell P_{\text{occ}}(\ell) \exp\left[-(1-\mu)\beta_{jk} \int_{0}^{\min(\ell,a)} C(a-\tilde{r})d\tilde{r}\right],$$
(S24)

with G(a; 0) given, from Eq. (S16), by

$$G(a;0) = \sum_{j,k} \frac{j}{z} p_{jk} \int_0^\infty d\ell \, P_{\text{occ}}(\ell) \exp\left[-(1-\mu)\beta_{jk} \int_0^{\min(\ell,a)} C(a-\tilde{r})d\tilde{r}\right],\tag{S25}$$

and C(a) is the cumulative distribution function for memory times. If we consider the large-age limit, $a \to \infty$, than we can approximate the integral of the cumulative distribution function as

$$\int_{0}^{\min(\ell,a)} C(a-\tilde{r})d\tilde{r} \approx \ell C(a)$$
(S26)

and the integral over ℓ can be calculated to give the large-a approximation

$$q_1(a) \sim \sum_{jk} \beta_{jk} p_{jk} \frac{j\beta\lambda + \mu\beta_{jk}}{j\overline{\beta}\lambda + \mu\beta_{jk} + (1-\mu)\beta_{jk}C(a)} \left[1 - \lambda + \lambda G(a;0)\right]^k,$$
(S27)

with

$$G(a;0) \sim \sum_{jk} \frac{j}{z} p_{jk} \frac{j\overline{\beta}\lambda + \mu\beta_{jk}}{j\overline{\beta}\lambda + \mu\beta_{jk} + (1-\mu)\beta_{jk}C(a)}.$$
 (S28)

In the simplified case $p_{jk} = \delta_{j,z} p_k$ and $\beta_{jk} \equiv 1$, Eqs. (S27) and (S28) reduce to Eq. (S2), as used in Figs. 2B and 3D.

The $a = \infty$ limit of $q_1(a)$ gives the fraction of memes that are *never* retweeted, and so have popularity n = 1 forever⁵. The value of $q_1(\infty)$ is obtained from Eqs. (S27) and (S28) by setting

⁵Recall that the popularity n of a meme is set to 1 when it is first tweeted (i.e., at birth); subsequent retweets (if any) increase the value of n above 1.

C(a) to its $a \to \infty$ limit of 1. The approach of $q_1(a)$ towards the value $q_1(\infty)$ depends, through the CDF C(a), on the tail of the memory-time distribution Φ . If the distribution Φ is heavy-tailed, there is a non-negligible probability that a meme may be retweeted even if a very long time has elapsed since its birth.

S2.4 Distribution of response times

It is worth noting that all agents in the model have constant activity rates, so that the actions of each individual agent constitute a Poisson process. A Poisson process is characterized by an exponential distribution of inter-event times, where each event corresponds to an innovation or a retweeting action. This assumption is contrary to studies such as [19, 33-35], where heavy-tailed distributions of inter-event times are examined. Despite this, in our model the memory-time distribution $\Phi(t_m)$ directly influences the waiting times (or "response times") between the receipt of a specific meme, and the retweeting of it. Indeed, if $\Phi(t_m)$ is a heavy-tailed distribution, then a meme received by a given user at time τ will be retweeted by that user at a time t (with $t \gg \tau$) with probability proportional to⁶ $\Phi(t-\tau)$. Therefore, a heavy-tailed memory distribution gives rise to a heavy-tailed waiting-time distribution for individual memes, despite the fact that the activity of each individual user is described by a Poisson process (cf. the heavy-tailed waiting-time distributions found in empirical studies of email correspondence [36, 37]). It is clearly important to distinguish between the distributions of inter-event times (for actions of users) and of the waiting times experienced by individual memes: the model assumes each user has exponentiallydistributed inter-event times, but it can nevertheless produce heavy-tailed distributions of waiting times for memes to be retweeted.

In particular, if the memory-time distribution $\Phi(t_m)$ is a Gamma (k_G, θ) distribution [35] as used in Fig. 3 of the main text, i.e., $\Phi(t_m) = \frac{1}{\Gamma(k_G)\theta^{k_G}} t_m^{k_G-1} \exp(-t_m/\theta)$, then $\Phi(t_m)$ is approximately power-law for memory times t_m with $t_m \ll \theta$, with an exponential cutoff at larger times. The corresponding waiting-time distribution shows a similar scaling in this range; for the parameters used in Fig. 3 ($k_G = 0.25$, $\theta = 80$ days) the waiting-time distribution scales as $t_m^{-0.75}$ for $t_m \ll \theta$, similar to the slow decay noted in empirical response times for Twitter users in Fig. 5 of [32].

S2.5 Mean popularity

The age dependence of the mean popularity (i.e., the expected number of tweets/retweets for a meme of age a) is given by

$$m(a) = \sum_{n=1}^{\infty} n q_n(a) = \left. \frac{\partial H(a; x)}{\partial x} \right|_{x=1}.$$
 (S29)

⁶The exact relation depends on how long the meme occupies the stream of the user.

Differentiating (S19) and setting x = 1 yields an integral equation for m(a):

$$m(a) = \sum_{jk} \beta_{jk} p_{jk} \left\{ 1 + \lambda k \, m_G(a) + (1 - \mu) \beta_{jk} \int_0^\infty d\ell \, \left(j\overline{\beta}\lambda + \mu\beta_{jk} \right) \exp\left[- \left(j\overline{\beta}\lambda + \mu\beta_{jk} \right) \ell \right] \times \right. \\ \left. \times \int_0^{\min(\ell,a)} d\tilde{r} \, \int_0^{a - \tilde{r}} d\tilde{\tau} \, \Phi(a - \tilde{r} - \tilde{\tau}) \left[1 + \lambda k \, m_G(\tilde{\tau}) \right] \right\},$$
(S30)

where $m_G(a)$, defined by $m_G(a) = \frac{\partial G(a;x)}{\partial x}\Big|_{x=1}$, is the solution of the integral equation found by differentiating Eq. (S16):

$$m_{G}(a) = \sum_{jk} \frac{j}{z} p_{jk} \int_{0}^{\infty} d\ell \left(j\overline{\beta}\lambda + \mu\beta_{jk} \right) \exp\left[-\left(j\overline{\beta}\lambda + \mu\beta_{jk} \right) \ell \right] \times \\ \times (1-\mu)\beta_{jk} \int_{0}^{\min(\ell,a)} d\tilde{r} \int_{0}^{a-\tilde{r}} d\tilde{\tau} \Phi(a-\tilde{r}-\tilde{\tau}) \left[1 + \lambda k \, m_{G}(\tilde{\tau}) \right].$$
(S31)

The order of the time integrals may be swapped using the identity

$$\int_0^\infty d\ell \int_0^{\min(\ell,a)} d\tilde{r} = \int_0^a d\tilde{r} \int_{\tilde{r}}^\infty d\ell, \tag{S32}$$

and the resulting ℓ integral can be performed explicitly:

$$\int_{\tilde{r}}^{\infty} (j\overline{\beta}\lambda + \mu\beta_{jk})e^{-(j\overline{\beta}\lambda + \mu\beta_{jk})\ell}d\ell = e^{-(j\overline{\beta}\lambda + \mu\beta_{jk})\tilde{r}}.$$
(S33)

As a result, the expressions (S30) and (S31) can be written as double convolution integrals. Taking Laplace transforms, Eq. (S30) then becomes

$$\hat{m}(s) = \frac{1}{s} + z\overline{\beta}\lambda\hat{m}_G(s) + (1-\mu)\hat{\Phi}(s)\sum_{j,k}\beta_{jk}^2 p_{jk}\frac{\frac{1}{s} + \lambda k\hat{m}_G(s)}{j\overline{\beta}\lambda + \mu\beta_{jk} + s},$$
(S34)

where hats denote Laplace transforms, e.g.,

$$\hat{\Phi}(s) \equiv \int_0^\infty e^{-st} \Phi(t) dt, \qquad (S35)$$

and with $\hat{m}_G(s)$ given explicitly from the Laplace transform of Eq. (S31):

$$\hat{m}_G(s) = \frac{(1-\mu)\hat{\Phi}(s)\sum_{j,k}\frac{j}{z}p_{jk}\frac{\beta_{jk}}{j\overline{\beta}\lambda+\mu\beta_{jk}+s}}{s\left[1-(1-\mu)\lambda\hat{\Phi}(s)\sum_{j,k}\frac{j}{z}p_{jk}\frac{k\beta_{jk}}{j\overline{\beta}\lambda+\mu\beta_{jk}+s}\right]}.$$
(S36)

If we specialize now to the case considered in the main text, where $\beta_{jk} \equiv 1$ for all (j, k) classes, and $p_{jk} = \delta_{j,z} p_k$, we obtain the simpler expression

$$\hat{m}_G(s) = \frac{(1-\mu)\hat{\Phi}(s)\frac{1}{z\lambda+\mu+s}}{s\left[1-(1-\mu)\lambda\hat{\Phi}(s)\frac{z}{z\lambda+\mu+s}\right]}.$$
(S37)

Substituting for \hat{m}_G into the simplified version of Eq. (S34) yields Eq. (1) of the main text.

S2.5.1 Large-age asymptotics of mean popularity

To consider the large-age asymptotics of m(a), we replace $\hat{\Phi}(s)$ by the first two terms of its small-s expansion:

$$\hat{\Phi}(s) \approx 1 - sT$$
 as $s \to 0$, (S38)

where $T = \int_0^\infty t_m \Phi(t_m) dt_m$ is the mean memory time. We consider ages such that $a \gg T$ and using approximation (S38) in Eq. (S37) yields a rational function of s with inverse Laplace transform

$$m(a) \sim 1 + \frac{1-\mu}{\mu} \left(1 - \frac{1+T(\lambda z+1)}{1+T\lambda z(1-\mu)} e^{-\frac{\mu(\lambda z+1)}{1+T\lambda z(1-\mu)}a} \right) \quad \text{as } a \to \infty.$$
(S39)

Setting $a = \infty$, we immediately obtain the steady-state value of the mean popularity, $m(\infty) = 1/\mu$.

Although Eq. (S39) is a large-*a* asymptotic result, we may expand the exponential term about a = 0 provided that the argument of the exponential remains small. This is valid for ages *a* that obey the constraint

$$a \ll \frac{1 + T\lambda z(1 - \mu)}{\mu(\lambda z + 1)},\tag{S40}$$

which is consistent with the earlier assumption $a \gg T$ if the innovation probability μ is sufficiently small. Indeed, taking the $\mu \to 0$ limit of (S39), the function m(a) grows linearly with a for ages in this range:

$$m(a) \sim \frac{\lambda z + 1}{T\lambda z + 1}a.$$
(S41)

S2.6 Infinite-age limit of popularity distribution

In the infinite-age (steady-state) limit $a \to \infty$, we assume $G(a; x) \to G_{\infty}(x)$, independent of a, and use the fact that $\int_0^\infty \Phi(t) dt = 1$ in Eq. (S16) to obtain

$$G_{\infty}(x) = \sum_{jk} \frac{j}{z} p_{jk} \int_{0}^{\infty} d\ell \left(j\overline{\beta}\lambda + \mu\beta_{jk} \right) \exp\left[-\left(j\overline{\beta}\lambda + \mu\beta_{jk} \right) \ell \right] \times \exp\left[-(1-\mu)\beta_{jk}\ell(1-x\left[1-\lambda+\lambda G_{\infty}(x)\right]^{k}) \right].$$
(S42)

Calculating the ℓ integral then gives the equation satisfied by $G_{\infty}(x)$:

$$G_{\infty}(x) = \sum_{jk} \frac{j}{z} p_{jk} \frac{j\overline{\beta}\lambda + \mu\beta_{jk}}{j\overline{\beta}\lambda + \beta_{jk} - (1-\mu)\beta_{jk}x \left[1 - \lambda + \lambda G_{\infty}(x)\right]^{k}}.$$
 (S43)

Similarly, the infinite-age limit for H is given in terms of G_{∞} by

$$H_{\infty}(x) = \sum_{jk} \beta_{jk} p_{jk} \frac{\left(j\overline{\beta}\lambda + \mu\beta_{jk}\right) x \left[1 - \lambda + \lambda G_{\infty}(x)\right]^{k}}{j\overline{\beta}\lambda + \beta_{jk} - (1 - \mu)\beta_{jk} x \left[1 - \lambda + \lambda G_{\infty}(x)\right]^{k}}.$$
(S44)

Note that these steady-state equations are independent of the memory distribution function Φ . Accordingly, the asymptotic analysis approach used in [17] to obtain the large-*n* behavior of the popularity distribution $q_n(\infty)$ may also be applied here: this is based on writing x = 1 - w and $G_{\infty} = 1 - \phi(w)$ and analyzing the small-w, small- ϕ asymptotics of Eqs. (S43) and (S44). We refer to [17] for details, and here summarize the main results for the case $\beta_{jk} \equiv 1$, $p_{jk} = \delta_{j,z}p_k$ that is considered in the main text.

• Case 1: p_k has finite second moment

The large-n scaling of the popularity distribution is given by a power-law with exponential cutoff:

$$q_n(\infty) \sim A n^{-\frac{3}{2}} e^{-\frac{n}{\kappa}} \quad \text{as } n \to \infty,$$
 (S45)

where the prefactor A is⁷

$$A = \frac{z(\lambda z + 1)}{\lambda z + \mu} \left[2\pi \left(\frac{\langle k^2 \rangle (2 + \lambda z - \mu)}{\lambda z + \mu} - z \right) \right]^{-\frac{1}{2}}$$
(S46)

and the cutoff κ is

$$\kappa = \frac{2\lambda^2 (1-\mu)^2}{\mu^2 (\lambda z+1)^2} \left[\frac{\langle k^2 \rangle (2+\lambda z-\mu)}{\lambda z+\mu} - z \right].$$
(S47)

Note that κ is proportional to $1/\mu^2$ for small μ , so in the limit of vanishing innovation probability the exponential cutoff tends to infinity and the power-law part of the popularity distribution extends to all n.

• Case 2: $p_k \sim D k^{-\gamma}$ as $k \to \infty$, with γ between 2 and 3

Immediately taking the $\mu \to 0$ limit, we find in this case that the popularity distribution has a power-law form with exponent $\gamma/(\gamma - 1)$ lying between 3/2 and 2:

$$q_n(\infty) \sim B n^{-\frac{\gamma}{\gamma-1}}$$
 as $n \to \infty$ (S48)

with prefactor B given by

$$B = -(\lambda z + 1) \frac{(D\Gamma(1-\gamma))^{-\frac{1}{\gamma-1}}}{\lambda\Gamma\left(\frac{1}{1-\gamma}\right)} \left[(\lambda z)^2 \sum_{n=1}^{\infty} \frac{n^{\gamma-1}}{(\lambda z + 1)^{n+1}} \right]^{-\frac{1}{\gamma-1}}.$$
 (S49)

S2.7 Large-*a*, large-*n* asymptotics of popularity distribution

Next we consider how the popularity distribution $q_n(a)$ behaves for large, but finite, ages. We are particularly interested in the case where the out-degree distribution of the social network is scalefree, as we seek to understand the "parallel" CCDFs observed at various ages in the empirical data of Fig. 3A of the main text.

⁷The values of A, κ and B reported here are not identical to those reported in [17]; this is because of the approximations made in the analysis of [17]. However, the differences are of order $1/(\lambda z)$, and so are negligible in the case $\lambda z \gg 1$ that is considered in [17].

Restricting ourselves to the case $\beta_{jk} \equiv 1$, $p_{jk} = \delta_{j,z}p_k$ and taking the $\mu \to 0$ limit, Eq. (S16) becomes

$$G(a;x) = \sum_{k} p_{k} \int_{0}^{\infty} d\ell \,\lambda z e^{-\lambda z \ell} \times \\ \times \exp\left[-\int_{0}^{\min(\ell,a)} d\tilde{r} \int_{0}^{a-\tilde{r}} d\tilde{\tau} \,\Phi(a-\tilde{r}-\tilde{\tau})(1-x\left[1-\lambda+\lambda G(\tilde{\tau};x)\right]^{k})\right].$$
(S50)

Writing x = 1 - w and $G(a; x) = 1 - \phi(a; w)$, we note that the argument of the exponential function vanishes when w = 0 and $\phi = 0$, and so we consider the small-w, small- ϕ asymptotic behavior by expanding the exponential term to first order in its argument:

$$\phi(a;w) \approx \int_0^\infty d\ell \,\lambda z e^{-\lambda z\ell} \int_0^{\min(\ell,a)} d\tilde{r} \,\int_0^{a-\tilde{r}} d\tilde{\tau} \,\Phi(a-\tilde{r}-\tilde{\tau})(1-(1-w)\sum_k p_k \left[1-\lambda\phi(\tilde{\tau};w)\right]^k). \tag{S51}$$

We note that retaining only the first-order term in the expansion of the exponential is an approximation. We will estimate the accuracy of this "one-term expansion" by comparing the infinite-age limit determined under the approximation with the corresponding exact values as given in Sec. S2.6.

For the case of a scale-free out-degree distribution with $p_k \sim D k^{-\gamma}$ as $k \to \infty$, the asymptotic form of the summation term in Eq. (S51) is given by [17]

$$1 - (1 - w) \sum_{k} p_k \left[1 - \lambda \phi \right]^k \sim \lambda z \phi - C \phi^{\gamma - 1} + w + o(w, \phi) \quad \text{as } w \to 0, \phi \to 0, \tag{S52}$$

with the constant C given by $C = \lambda^{\gamma-1} D\Gamma(1-\gamma)$. Applying the integral-swapping trick of Eq. (S32) allows the right hand side of Eq. (S51) to be expressed as a double convolution integral. Laplace transforming then yields

$$\hat{\phi}(s;w) = \frac{1}{\lambda z + s} \hat{\Phi}(s) \mathcal{L} \left[\lambda z \phi - C \phi^{\gamma - 1} + w \right],$$
(S53)

where \mathcal{L} denotes the Laplace transform operation applied to the term in square brackets. In the $a \to \infty$ limit, this equation is satisfied by the steady-state solution

$$\phi(\infty; w) = C^{-\frac{1}{\gamma - 1}} w^{\frac{1}{\gamma - 1}}, \tag{S54}$$

as can be verified using the final value theorem for Laplace transforms. We note that the corresponding expression for $\phi(\infty; w)$ as calculated from the steady state Eq. (S43) has a additional multiplicative factor of $F(\lambda z, \gamma)$ that is absent in Eq. (S54), where the function $F(\zeta, \gamma)$ is defined by

$$F(\zeta,\gamma) = \left[\zeta^2 \sum_{n=1}^{\infty} \frac{n^{\gamma-1}}{(\zeta+1)^{n+1}}\right]^{-\frac{1}{\gamma-1}},$$
(S55)

see Fig. S4. If $\lambda z \gg 1$, then $F(\lambda z, \gamma) \approx 1$ and the one-term expansion gives results that are very close to the exact values (at least in the infinite-age limit $a \to \infty$). Moreover, even if λz is not



Figure S4: The function $F(\zeta, \gamma)$, as defined in Eq. (S55), for values of γ close to 2. The highlighted points are the parameter values that are relevant to Fig. 2 ($\lambda z = 11$, $\gamma = 2.5$) and to Fig. 3 ($\lambda z = 0.32$, $\gamma = 2.13$). In all cases of interest the values of F are close to 1, and so we conclude that the one-term expansion used in Eq. (S51) gives a good estimate of the exact steady-state solution.

large (e.g., $\lambda z = 0.32$ for the model fit in Fig. 3), the values of $F(\lambda z, \gamma)$ can still be close to unity if γ is sufficiently close to 2.

To consider small deviations from the steady state, we define g(a; w) by

$$\phi(a;w) = \phi(\infty;w) \left(1 - g(a;w)\right) \tag{S56}$$

with $g(a; w) \to 0$ as $a \to \infty$. Assuming that g is sufficiently small to allow the use of the linearizing approximation

$$(1-g)^{\gamma-1} \approx 1 - (\gamma - 1)g,$$
 (S57)

Eq. (S53) can be solved for the Laplace transform of g:

$$\hat{g}(s;w) = \frac{1}{s} \frac{s + \lambda z - \lambda z \hat{\Phi}(s)}{s + \lambda z - \lambda z \hat{\Phi}(s) + (\gamma - 1)C^{\frac{1}{\gamma - 1}} w^{\frac{\gamma - 2}{\gamma - 1}} \hat{\Phi}(s)}.$$
(S58)

The Laplace transform of ϕ then follows from Eq. (S56) and a similar asymptotic analysis of Eq. (S19) yields

$$\hat{H}(s;1-w) = \frac{1}{s} - \frac{\lambda z(s+\lambda z+\hat{\Phi}(s))}{s+\lambda z}\hat{\phi}(s;w)$$
(S59)

Substituting from Eqs. (S56) and (S58) results in Eq. (S1). Numerical inversion of the Laplace transform and of the PGF, as described in Sec. S4, give the results shown in Figs. 2 and 3 of the main text.

A similar analysis can be performed in the case where the out-degree distribution p_k has finite second moment. We again utilize a one-term expansion similar to Eq. (S51), but we can also retain a non-vanishing innovation probability μ in this case. The one-term expansion can be shown to be accurate when $\lambda z \gg 1$; this condition is obeyed in all relevant cases we examine. The resulting large-*a* asymptotics for the generating function H(a; x) are found by inverting the following Laplace transform:

$$\hat{H}(s;1-w) = \frac{1}{s} - \phi(\infty;w) \times \frac{(1-\mu)\lambda z(s+\lambda z+\mu+\hat{\Phi}(s))}{s(s+\lambda z+\mu)} \left[\frac{2(1-\mu)\frac{w}{\phi(\infty;w)}\hat{\Phi}(s) - \mu(\lambda z+1)\hat{\Phi}(s)}{s+\lambda z+\mu - (\lambda z(1+\mu)+2\mu)\hat{\Phi}(s) + 2(1-\mu)\frac{w}{\phi(\infty;w)}\hat{\Phi}(s)} \right],$$
(S60)

with $\phi(\infty; w)$ given by

$$\phi(\infty; w) = \frac{-\mu(\lambda z + 1) + \sqrt{\mu^2(\lambda z + 1)^2 + 2\lambda^2(1 - \mu)^2(\langle k^2 \rangle - z)w}}{\lambda^2(1 - \mu)(\langle k^2 \rangle - z)}.$$
 (S61)

Numerical inversion of the Laplace transform, as in Sec. S4, gives the asymptotic results shown in the inset of Fig. 2A.

S2.7.1 Analysis of the data collapse in Fig. 3B

The large-*n* asymptotics of the popularity distribution are found from the small-*w* expansion (with w = 1 - x) of h(a; x) = 1 - H(a; x), and in the $a \to \infty$ limit for the scale-free out-degree distribution we obtain from Eq. (S59) (using the final value theorem for Laplace transforms)

$$h(\infty; 1-w) \sim (\lambda z+1)C^{-\frac{1}{\gamma-1}}w^{\frac{1}{\gamma-1}}$$
 as $w \to 0.$ (S62)

Understanding the large-*a* approach to this steady state (i.e., the case where *a* is large but finite) is a difficult problem in asymptotic analysis, involving the double limits $n \to \infty$ and $a \to \infty$. However, some insight can be obtained by factoring the function *h* into a product of its infinite-age limit $h(\infty; x)$ and another function h_1 , with h_1 limiting to 1 as $a \to \infty$:

$$h(a;x) = h(\infty;x)h_1(a;x).$$
(S63)

Taking Laplace transforms gives

$$\hat{h}(s;x) = h(\infty;x)\hat{h}_1(s;x),$$
(S64)

where

$$\hat{h}_{1}(s;1-w) = \frac{\lambda z(s+\lambda z+\hat{\Phi}(s))}{s(\lambda z+1)(s+\lambda z)} \frac{(\gamma-1)\lambda D^{\frac{1}{\gamma-1}} \left[\Gamma(1-\gamma)\right]^{\frac{1}{\gamma-1}} w^{\frac{\gamma-2}{\gamma-1}} \hat{\Phi}(s)}{s+\lambda z-\lambda z \hat{\Phi}(s)+(\gamma-1)\lambda D^{\frac{1}{\gamma-1}} \left[\Gamma(1-\gamma)\right]^{\frac{1}{\gamma-1}} w^{\frac{\gamma-2}{\gamma-1}} \hat{\Phi}(s)}$$
(S65)

In particular, note that $\hat{h}_1(s; 1-w)$ depends on w only through the factor $w^{\frac{\gamma-2}{\gamma-1}}$. In the case where γ is very close to 2, the exponent $(\gamma - 2)/(\gamma - 1)$ of the w dependence is close to zero, and the dependence of h_1 on w is therefore very weak. It follows that the rate of approach of the corresponding distribution $q_n(a)$ to the steady state $q_n(\infty)$ does not show a strong dependence

on n, and the CCDFs for various ages appear almost parallel in the log-log plot of Fig. 3A of the main text (note $\gamma = 2.13$ in the Twitter network used in Fig. 3 of the main text, see Sec. S1).

As we saw in Sec. S2.5 for the large-age asymptotics of the mean popularity, the long-time behavior of the popularity distribution may be obtained by inserting the small-s approximation $\hat{\Phi}(s) \approx 1 - sT$ in Eq. (S1) and examining the linear (early-age) growth of the inverse transforms. The resulting popularity distributions $q_n(a)$ show (for large n) a regime of linear-in-age growth, and in the case where $\gamma \approx 2$, the rate of this growth depends only weakly on n. Since the mean popularity m(a) is also growing linearly during this age period (see Eq. (S41)), the division of the CCDFs at various ages by the corresponding mean m(a) leads to the collapse of the data onto the single curve that is seen in Fig. 3B.

S3 Extension to heterogeneous activity rates

To focus on understanding the combined effects of memory and out-degree distribution, most of our results thus far are specialized to the case of uniform user activity rates, $\beta_{jk} \equiv 1$. It is interesting, however, to examine the impact of heterogeneous activity rates upon the results we have obtained. To this end, we extend here to the case where the activity rate of a user depends on its out-degree k while retaining the assumption $p_{jk} = \delta_{j,z}p_k$, so that $\beta_{jk} = \beta_k$ (normalized so that $\sum_k \beta_k p_k = 1$ and with $\overline{\beta} = \sum_k \frac{k}{z} \beta_k p_k$).

The mean popularity is given in the general case by Eq. (S34). Repeating the asymptotic analysis of Eq. (S38) though to Eq. (S41) for the $\mu \to 0$ limit, we again find linear growth of m(a) with age a, with a slope that generalizes that found in Eq. (S41):

$$m(a) \sim \frac{\lambda z \overline{\beta} + \frac{\beta^2}{\overline{\beta}}}{T \lambda z \overline{\beta} + 1} a \quad \text{as } a \to \infty,$$
 (S66)

where we have introduced the notation $\overline{\beta^2} \equiv \sum_k \frac{k}{z} (\beta_k)^2 p_k$.

If we additionally assume that the user activity rates saturate to a constant level β_{∞} at very large k, so $\beta_k \to \beta_{\infty}$ as $k \to \infty$, then we can repeat the asymptotic approximations of Sec. S2.7 to determine a generalized version of Eq. (S1):

$$\hat{H}(s;x) \sim \frac{1}{s} - \frac{1}{s} \frac{\lambda z \overline{\beta} \left(s + \lambda z \overline{\beta} + \frac{\overline{\beta}^2}{\beta} \hat{\Phi}(s)\right) (\gamma - 1)(1 - x) \hat{\Phi}(s)}{(s + \lambda z \overline{\beta}) \left(s + \lambda z \overline{\beta} - \lambda z \overline{\beta} \hat{\Phi}(s) + \beta_{\infty}^{\frac{1}{\gamma - 1}} (\gamma - 1) \lambda D^{\frac{1}{\gamma - 1}} [\Gamma(1 - \gamma)]^{\frac{1}{\gamma - 1}} (1 - x)^{\frac{\gamma - 2}{\gamma - 1}} \hat{\Phi}(s)\right)}$$
(S67)

To demonstrate the effect of heterogeneous activity rates, we consider a model for β_k inspired by the data analysis shown in Fig. 6(a) of [27]. There, the average activity rate (as measured by the number of tweets by a user in a fixed time period) is found to grow approximately linearly with the number of followers k of that user, for k from 0 to about 100. Then, for k values from about 100 up to the maximum shown in the plot ($k = 10^3$), the activity rate grows as a more



Figure S5: As Fig. 3 of the main text, but including heterogeneous activity rates β_k given by Eq. (S68). Model parameters are: $\lambda = 5 \times 10^{-4}$, $\mu = 0.033$, $k_G = 0.25$, $\theta = 500$, with one model time unit corresponding to 0.18 days.

slowly increasing linear function of k. We model these characteristics (which are also seen in other studies, e.g., [38]), using a piecewise-linear and continuous function of k, assuming a saturation of activity at very high k, as follows:

$$\beta_k \propto \begin{cases} 0.35k & \text{if } k < 100, \\ 35 + 0.044(k - 100) & \text{if } 100 \le k < 10^4. \\ 470.6 & \text{if } k \ge 10^4, \end{cases}$$
(S68)

where the values are chosen to closely match the linear growth rates in Fig. 6(a) of [27]. Using this heterogeneous activity rate (with the constant of proportionality set by the condition $\sum_k \beta_k p_k = 1$), Fig. S5 shows results that correspond closely to the homogeneous-activity example of Fig. 3 of the main text. A comparison of panels D from both figures clearly shows that including heterogeneous activity rates allows a better fitting of the model to data for the fraction $q_1(a)$ of non-retweeted memes. However, the other results of the model (panels A, B and C of Fig. S5 compared to same panels in Fig. 3) are relatively unaffected by the activity rate, so that the good matches between model and data seen in Fig. 3 are not compromised by including heterogeneity in activity rates.

S4 Numerical inversion of Laplace transforms and PGFs

Many of our results for the popularity distribution $q_n(a)$ are expressed in terms of the corresponding PGF H(a; x). As in [17], we use the Fast Fourier Transform method of [39–41] to numerically invert the PGF at a fixed age a to produce, for example, the model distributions in Figs. 2 and 3; see Sec. S2 of [17] for further details and links to Octave/Matlab code for implementing the PGF inversion.

The results of the model for the age-dependence of several quantities are expressed in terms of Laplace transforms. To numerically invert the Laplace transforms we use the efficient Talbot algorithm [42], in its simplified version described in Sec. 6 of [43]. The Talbot algorithm is based on a numerical evaluation of the Bromwich (Laplace inversion) integral, using a cleverly-chosen deformation of the contour in the complex-*s* plane. The Laplace inversion of $\hat{H}(s;x)$ to obtain H(a;x) at a desired age *a*, for example, can be quickly computed using the $2M_L - 1$ weights γ_k and nodes δ_k defined by [44]

$$\delta_{0} = \frac{2M_{L}}{5}, \quad \delta_{k} = \frac{2k\pi}{5} (\cot(k\pi/M_{L}) + i) \quad \text{for } -M_{L} + 1 \le k \le M_{L} - 1,$$

$$\gamma_{0} = \frac{1}{2}e^{\delta_{0}},$$

$$\gamma_{k} = [1 + i(k\pi/M_{L})(1 + [\cot(k\pi/M_{L})^{2}]) - i\cot(k\pi/M_{L})]e^{\delta_{k}} \quad \text{for } -M_{L} + 1 \le k \le M_{L} - 1,$$

(S69)

(where $i = \sqrt{-1}$) by calculating the sum

$$H(a;x) = \frac{1}{5a} \left[\gamma_0 \hat{H}\left(\frac{\delta_0}{a};x\right) + \sum_{k=-M_L+1}^{M_L-1} \gamma_k \hat{H}\left(\frac{\delta_k}{a};x\right) \right].$$
(S70)

In practice, the precision of the Talbot algorithm is very high, and only relatively small values of M_L are required to obtain accurate results; we used $M_L = 25$ in the examples shown.

References

- E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts. Everyone's an influence: quantifying influence on Twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 65–74. ACM, 2011.
- [2] K. Lerman, R. Ghosh, and T. Surachawala. Social contagion: An empirical study of information spread on Digg and Twitter follower graphs. arXiv:1202.3162, 2012.
- [3] L Weng, A Flammini, A Vespignani, and F Menczer. Competition among memes in a world with limited attention. *Scientific Reports*, 2:335, 2012.
- [4] J. Cheng, L. Adamic, P. A. Dow, J. M. Kleinberg, and J. Leskovec. Can cascades be predicted? In *Proceedings of the 23rd international conference on World Wide Web*, pages 925–936. International World Wide Web Conferences Steering Committee, 2014.
- [5] J. P. Gleeson, D. Cellai, J.-P. Onnela, M. A. Porter, and F. Reed-Tsochas. A simple generative model of collective online behavior. *Proceedings of the National Academy of Sciences*, 111(29):10411–10415, 2014.
- [6] R. A. Bentley, P. Ormerod, and M. Batty. Evolving social influence in large populations. Behavioral Ecology and Sociobiology, 65(3):537–546, 2011.
- [7] M. V. Simkin and V. P. Roychowdhury. A mathematical theory of citing. Journal of the American Society for Information Science and Technology, 58(11):1661–1673, 2007.
- [8] O. A. Pinto and M. A. Muñoz. Quasi-neutral theory of epidemic outbreaks. *PloS One*, 6(7):e21946, 2011.
- [9] M. Kimura. The neutral theory of molecular evolution. Cambridge University Press, 1984.
- [10] G. Bianconi and A.-L. Barabási. Competition and multiscaling in evolving networks. EPL (Europhysics Letters), 54(4):436, 2001.
- [11] R. A. Bentley, M. W. Hahn, and S. J. Shennan. Random drift and culture change. Proceedings of the Royal Society of London. Series B: Biological Sciences, 271(1547):1443–1450, 2004.
- [12] M. V. Simkin and V. P. Roychowdhury. Re-inventing Willis. Physics Reports, 502(1):1–35, 2011.
- [13] H. A. Simon. On a class of skew distribution functions. *Biometrika*, 42:425–440, 1955.
- [14] D. de Solla Price. A general theory of bibliometric and other cumulative advantage processes. Journal of the American Society for Information Science, 27(5):292–306, 1976.

- [15] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. Science, 286(5439):509–512, 1999.
- [16] M. E. J. Newman. Power laws, Pareto distributions and Zipf's law. Contemporary Physics, 46(5):323–351, 2005.
- [17] J. P. Gleeson, J. A. Ward, K. P. O'Sullivan, and W. A. Lee. Competition-induced criticality in a model of meme popularity. *Physical Review Letters*, 112(4):048701, 2014.
- [18] T. E. Harris. The theory of branching processes. Dover Publications, 2002.
- [19] J. L. Iribarren and E. Moro. Impact of human activity patterns on the dynamics of information diffusion. *Physical Review Letters*, 103(3):38702, 2009.
- [20] K. I. Goh, D. S. Lee, B. Kahng, and D. Kim. Sandpile on scale-free networks. *Physical Review Letters*, 91(14):148701, 2003.
- [21] C. Adami and J. Chu. Critical and near-critical branching processes. *Physical Review E*, 66(1):011907, 2002.
- [22] S. Zapperi, K. B. Lauritsen, and H. E. Stanley. Self-organized branching processes: mean-field theory for avalanches. *Physical Review Letters*, 75(22):4071–4074, 1995.
- [23] S. Goel, D. J. Watts, and D. G. Goldstein. The structure of online diffusion networks. In Proceedings of the 13th ACM Conference on Electronic Commerce, pages 623–638. ACM, 2012.
- [24] J. Borge-Holthoefer, A. Rivero, I. García, E. Cauhé, A. Ferrer, D. Ferrer, D. Francos, D. Iñiguez, M. P. Pérez, G. Ruiz, F. Sanz, F. Serrano, C. Viñas, A. Tarancón, and Y. Moreno. Structural and dynamical patterns on online social networks: the Spanish May 15th movement as a case study. *PloS One*, 6(8):e23883, 2011.
- [25] S. González-Bailón, J. Borge-Holthoefer, A. Rivero, and Y. Moreno. The dynamics of protest recruitment through an online network. *Scientific Reports*, 1:197, 2011.
- [26] T. Hoffmann, M. A. Porter, and R. Lambiotte. Generalized master equations for non-Poisson dynamics on networks. *Physical Review E*, 86(4):046102, 2012.
- [27] N. O. Hodas, F. Kooti, and K. Lerman. Friendship paradox redux: your friends are more interesting than you. *ICWSM*, 13:8–10, 2013.
- [28] R. A. Bentley, M. Earls, and M. J. O'Brien. I'll Have What She's Having: Mapping Social Behavior. MIT Press, 2011.
- [29] M. J. Salganik, P. S. Dodds, and D. J. Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762):854–856, 2006.
- [30] L. Weng, F. Menczer, and Y.-Y. Ahn. Virality prediction and community structure in social networks. *Scientific Reports*, 3:2522, 2013.

- [31] A. Clauset, C. R. Shalazi, and M. E. J. Newman. Power-law distributions in empirical data. SIAM Review, 51(4):661–703, 2009.
- [32] N. O. Hodas and K. Lerman. How visibility and divided attention constrain social contagion. ASE/IEEE International Conference on Social Computing (SocialCom-2012), 2012.
- [33] A. Vazquez, B. Rácz, A. Lukács, and A.-L. Barabási. Impact of non-Poissonian activity patterns on spreading processes. *Physical Review Letters*, 98(15):158702, 2007.
- [34] B. Min, K.-I. Goh, and A. Vazquez. Spreading dynamics following bursty human activity patterns. *Physical Review E*, 83(3):036102, 2011.
- [35] J. L. Iribarren and E. Moro. Branching dynamics of viral information spreading. *Physical Review E*, 84(4):046116, 2011.
- [36] A.-L. Barabási. The origin of bursts and heavy tails in human dynamics. Nature, 435(7039):207–211, 2005.
- [37] R. D. Malmgren, D. B. Stouffer, A. E. Motter, and L. A. N. Amaral. A Poissonian explanation for heavy tails in e-mail communication. *Proceedings of the National Academy of Sciences*, 105(47):18153–18158, 2008.
- [38] S. A. Myers and J. Leskovec. The bursty dynamics of the Twitter information network. ArXiv:1403.2732, 2014.
- [39] J. K. Cavers. On the Fast Fourier Transform inversion of probability generating functions. IMA Journal of Applied Mathematics, 22(3):275–282, 1978.
- [40] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64(2):026118, 2001.
- [41] M. Marder. Dynamics of epidemics on random networks. Physical Review E, 75(6):066103, 2007.
- [42] A. Talbot. The accurate inversion of Laplace transforms. J. Inst. Math. Appl., 23:97–120, 1979.
- [43] J. Abate and W. Whitt. A unified framework for numerically inverting Laplace transforms. INFORMS Journal on Computing, 18(4):408–421, 2006.
- [44] J. Abate and P. P. Valkó. Multi-precision Laplace inversion. Int. J. Numer. Meth. Engng., 60:979–993, 2004.