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 & Statistics, University of Limerick, Ireland*
[2]*Centre for the Mathematics of Human Behaviour, Department of Mathematics and Statistics, University of Reading, Whiteknights, UK*

(Dated: 18 May 2013)

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.

PACS numbers: 89.65.-s, 05.65.+b, 89.75.Fb, 89.75.Hc

When humans 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 on Twitter [4, 5], or video views on YouTube [6], 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 2, 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 Yule-Simon type—as used to describe power-law degree distributions of networks, for example [7–10]—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 models [11]. This suggests 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 paper we present an analytically tractable model of selection behaviour, based on simplifying the model of Weng et. al [12] 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 [13]—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.

For clarity, we will phrase the model in terms of meme diffusion as in [12] but the same understanding of the basic mechanism—and the analytical techniques for time-dependent distributions—can also be applied to other popularity models, such as the random-copying models of [2, 14, 15]. The role of competition among items for limited resources has been examined from many viewpoints: see, for example, [16–18] and also related work on competing diseases [19–21]. The distribution of popularity increments (number of selections of an item in a small time interval) in Moran-type models has been obtained analytically [15]; however, our focus is on the (time-dependent) distributions of popularity accumulated over long timescales.

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 that edges are directed from nodes to their followers) with probability p_k; the k followers are chosen at random from the other $N-1$ nodes, giving a Poisson distribution of in-degrees. Each node has a *screen*, which holds the meme of current interest to that node. For simplicity, we assume here that each screen has capacity for only one meme, though this case is easily extended [31]. 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.

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 initally present on a small number of screens, with every other screen being empty, and with no innovation ($\mu = 0$). In a typical realization, the red and blue memes each spread quickly onto empty screens and their popularities both exhibit expo-

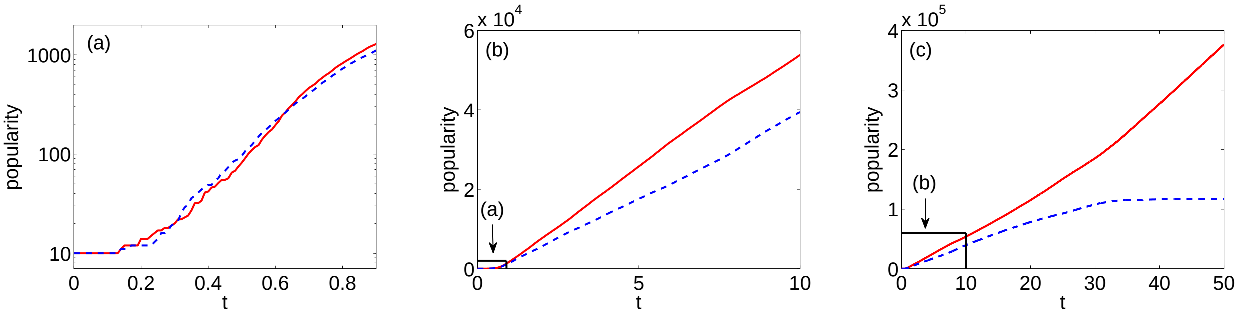


FIG. 1: Popularities $n_r(t)$ and $n_b(t)$ of red and blue memes in a single realization, starting from 10 screens each. Note the different timescales in each figure, and the fact that the vertical scale in (a) is logarithmic. The inset boxes in (b) and (c) show the area of the previous figures.

nential growth in time (Fig. 1(a)). Once most screens are occupied by either the blue or the red meme, the memes compete for space on screens, and both popularities grow approximately linearly with time (Fig. 1(b)). On longer timescales, one of the memes is eventually extinguished from all screens, and its popularity curve saturates (e.g., the blue meme in Fig. 1(c)).

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 $\sum_k kp_k$ 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

$$\frac{dr}{dt} = -zbr + zr(1-r), \qquad (1)$$

$$\frac{db}{dt} = -zbr + zb(1-b). \qquad (2)$$

The first term on the right-hand-side of (1), 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 non-red followers, the expected number of which is $z(1-r(t))$.

Equations (1) and (2) can be solved analytically: the fraction of non-empty 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 (SI) 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 $n_b(t)$, similar to early-stage growth of independent diseases [22]. 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) and (2) 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 $n_b(t)$ grows linearly in time, as in Fig. 1(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 can be understood as being induced by the competition between memes, in contrast to the exponential growth at earlier times (Fig. 1(a)) when the memes were not competing for the same resources [6, 18].

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 all screens (as in Fig. 1(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, using branching process theory [13].

Multiple competing memes: Now suppose that at $t = 0$ each node's screen is initialized with its own individual meme, so there are N distinct memes; the popularity of each meme is initialized to 1. There are no empty screens in the network—so we are in the highly-competitive regime—and the innovation probability μ

may be non-zero. 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 are interested in the distribution of popularity across the set of all N memes *initially* present in the system. We show that the model produces fat-tailed distributions of popularity, which are power-law 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, 23]).

The branching process description is strictly valid only when the number of screens occupied by an single meme is a small fraction of N, but we note that this is the case for long epochs of time when 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 [31].

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, chosen uniformly at random from all memes born at $t = 0$, has been tweeted n times by time a. This distribution can be represented via its probability generating function (PGF) [24, 25] $H(a, x)$, 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) [20, 26, 27]. 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 excess popularity is $n - 1$, the number of retweets of the meme subsequent to its initial seed tweet, so $G(a, x) = \sum_{n=0}^{\infty} q_{n+1}x^n$; the PGF for the popularity of age-a memes is then $H(a, x) = x\, G(a, x)$. In [31] we derive the following ordinary differential equation for $G(a, x)$, parameterized by x:

$$\frac{\partial G}{\partial a} = z + \mu - (z + 1)G + (1 - \mu)x\, G\, f(G). \qquad (3)$$

This equation is easily solved using standard numerical methods, starting from initial condition $G(0, x) = 1$. Some analysis is also possible [31]: 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 $(1 + \mu z)/\mu(z + 1)$. By expanding $G(a, x)$ into a truncated Taylor series about $x = 0$, the probabilities $q_n(a)$ for low n may be determined explicitly. For example, the probability that a meme's popularity never grows beyond its initial value of $n = 1$ is given by the large-a limit of q_1, which is $(z + \mu)/(z + 1)$, showing that (for $\mu \ll 1$ and $z \gg 1$) most memes do not go viral [28, 29] as they are never retweeted. The popularity distribution for larger values of n can be determined in a computationally efficient manner using FFTs [20, 26, 27]: our implementation [31] determines probabilities q_n for n values up to several thousand, shown as black curves in Figs. 2 and 3 [32]. The coloured symbols are the results of stochastic simulations of the model on networks of size $N = 10^5$; results for two independent realization of each network are shown, each giving a popularity distribution for the N memes present at $t = 0$. The match between theory and simulation is very good. Figure 2 shows the popularity distributions on networks where each node has exactly $z = 10$ followers. Figure 3 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 all cases the k followers of a given node are assigned at random, so the in-degree distributions are Poisson. We observe power-law popularity distributions $q_n \propto n^{-\alpha}$, with various exponents α, and with an exponential cut-off in Fig. 2.

The long-time (or "old-age", $a \to \infty$) asymptotics of the branching process are determined by analyzing the limiting solutions of Eq. (3) in the complex x-plane [31]; 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

$$q_n(\infty) \sim A\, n^{-\frac{3}{2}} e^{-\frac{n}{\kappa}} \quad \text{as } n \to \infty, \qquad (4)$$

where $\kappa = \frac{2(f''(1)+2z)}{\mu^2(z+1)^2}$ and $A = [2\pi(f''(1) + 2z)]^{-\frac{1}{2}}$. 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 non-zero 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 [23], that as $n \to \infty$,

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

with prefactors B and C given in [31]. Thus in the zero-innovation limit, the popularity distribution has a power-

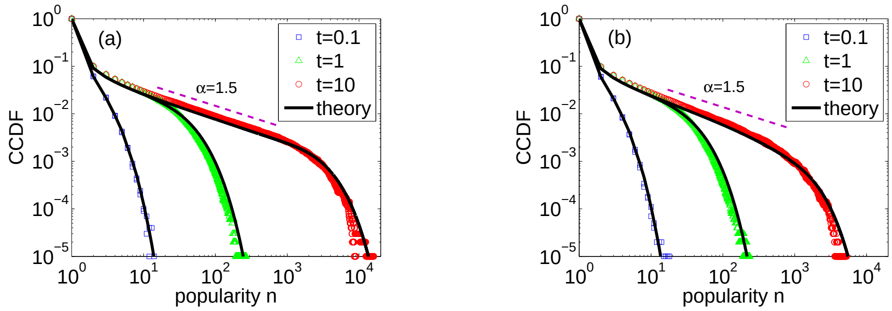


FIG. 2: Complementary cumulative distribution functions (CCDFs)—the fraction of memes with popularity $\geq n$—for numerical simulations (with $N = 10^5$), compared with the theory of Eq. (3). In this network every node has exactly $z = 10$ followers, $p_k = \delta_{k,10}$. In (a), the innovation parameter μ is zero; in (b), $\mu = 0.02$, causing a large-size cutoff as predicted by Eq. (4). In (b), only the popularities of the memes present at initialization are shown. Dashed lines correspond to CCDFs for power law popularity distributions $P_n \propto n^{-\alpha}$.

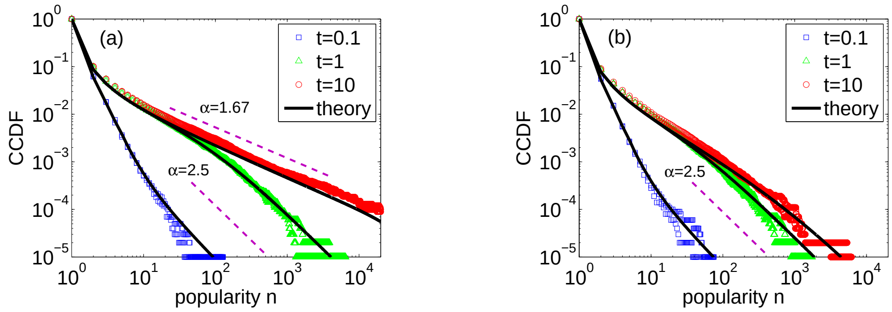


FIG. 3: As Fig. 2, but for a network with out-degree distribution $p_k \propto k^{-\gamma}$ for $k \geq 4$ with $\gamma = 2.5$ (mean degree $z = 10.6$). (a) $\mu = 0$, (b) $\mu = 0.1$. Note $\gamma/(\gamma - 1) = 5/3 \approx 1.67$ for this value of γ.

law exponent $\gamma/(\gamma-1)$ that is smaller than the exponent γ of the out-degree distribution. If $\mu > 0$, the large-n power-law exponent for the popularity distribution is the same as that of the out-degree distribution, but numerical results (Fig. 3(b)) show that for moderately large n the distribution still decreases more slowly that the out-degree distribution.

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. (3)—for example, by: (i) increasing the capacity of screens to $c > 1$ memes, with random choice among the c slots for tweeting or overwriting memes, and/or (ii) allowing followers to reject a meme tweeted to them with some probability so it does not appear on their screen—and to show that the CIC property is retained in the more general cases [31]. 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, 12, 30]—these will be added in future work—but it can nevertheless produce fat-tailed popularity distributions. These insights will guide the eventual creation of parsimonious but accurate models of human choice dynamics that can reproduce key characteristics of the rapidly expanding range of empirical data from online social networks.

Acknowledgements: This work was partially funded by Science Foundation Ireland (11/PI/1026 and 09/SRC/E1780), the Engineering and Physical Sciences Research Council (MOLTEN, EP/I016058/1) and by the FET-Proactive project PLEXMATH. Helpful discussions with Davide Cellai, Sergey Melnik, Peter Fennell, Mason Porter, J-P Onnela and Felix Reed-Tsochas are gratefully acknowledged. We acknowledge the SFI/HEA Irish Centre for High-End Computing (ICHEC) for the provision of computational facilities.

[1] R. A. Bentley, M. Earls, and M. J. O'Brien, *I'll Have What She's Having: Mapping Social Behavior* (MIT Press, 2011).

[2] R. A. Bentley, M. W. Hahn, and S. J. Shennan, Proceedings of the Royal Society of London. Series B: Biological Sciences **271**, 1443 (2004).

[3] J. P. Onnela and F. Reed-Tsochas, Proceedings of the National Academy of Sciences **107**, 18375 (2010).

[4] K. Lerman, R. Ghosh, and T. Surachawala, arXiv:1202.3162 (2012).
[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, 2011), pp. 65–74.
[6] G. Szabo and B. A. Huberman, Communications of the ACM **53**, 80 (2010).
[7] S. Redner, The European Physical Journal B-Condensed Matter and Complex Systems **4**, 131 (1998).
[8] M. E. J. Newman, Contemporary Physics **46**, 323 (2005).
[9] M. V. Simkin and V. P. Roychowdhury, Physics Reports **502**, 1 (2011).
[10] A. L. Barabási and R. Albert, Science **286**, 509 (1999).
[11] S. Zapperi, K. B. Lauritsen, and H. E. Stanley, Physical Review Letters **75**, 4071 (1995).
[12] L. Weng, A. Flammini, A. Vespignani, and F. Menczer, Scientific Reports **2** (2012).
[13] T. E. Harris, *The theory of branching processes* (Dover Publications, 2002).
[14] R. A. Bentley, P. Ormerod, and M. Batty, Behavioral Ecology and Sociobiology **65**, 537 (2011).
[15] T. S. Evans and A. D. K. Plato, Physical Review E **75**, 056101 (2007).
[16] C. Adami and J. Chu, Physical Review E **66**, 011907 (2002).
[17] M. Beguerisse Díaz, M. A. Porter, and J.-P. Onnela, Chaos **20**, 043101 (2010).
[18] F. L. Forgerini, N. Crokidakis, S. N. Dorogovtsev, and J. F. F. Mendes, Int. J. Complex Systems in Science **1**, 141 (2011).
[19] B. Karrer and M. E. J. Newman, Physical Review E **84**, 036106 (2011).
[20] P.-A. Noël, A. Allard, L. Hébert-Dufresne, V. Marceau, and L. J. Dubé, Physical Review E **85**, 031118 (2012).
[21] J. C. Miller, arXiv:1210.4511 (2012).
[22] A. Barrat, M. Barthlemy, and A. Vespignani, *Dynamical processes on complex networks* (Cambridge University Press, 2008).
[23] K. I. Goh, D. S. Lee, B. Kahng, and D. Kim, Physical Review Letters **91**, 148701 (2003).
[24] H. S. Wilf, *generatingfunctionology* (A K Peters/CRC Press, 2005).
[25] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Physical Review E **64**, 026118 (2001).
[26] J. K. Cavers, IMA Journal of Applied Mathematics **22**, 275 (1978).
[27] M. Marder, Physical Review E **75**, 066103 (2007).
[28] S. Goel, D. J. Watts, and D. G. Goldstein, in *Proceedings of the 13th ACM Conference on Electronic Commerce* (ACM, 2012), pp. 623–638.
[29] R. A. Baños, J. Borge-Holthoefer, and Y. Moreno, arXiv:1303.4629 (2013).
[30] J. L. Iribarren and E. Moro, Physical Review E **84**, 046116 (2011).
[31] See Supplementary Material accompanying this paper.
[32] Octave/Matlab codes for solving Eq. (3) and inverting the generating functions can be obtained from the authors upon request.

Supplementary Material for "Competition-induced criticality in a model of meme popularity"

S1 Derivation of Equation (3)

In the main text we introduced $G(a, x)$ as the probability generating function (PGF) of the excess popularity distribution:

$$G(a, x) = \sum_{n=0}^{\infty} q_{n+1}(a) x^n, \tag{S1}$$

where $q_n(a)$ is the probability of a randomly-chosen meme having popularity n at age a. It proves convenient here to also introduce $G^{(k)}(a, x)$, the PGF of the excess popularity distribution for a meme seeded by (i.e., first tweeted by) a node with out-degree k; we call this node's screen the *root screen* of the retweet-cascade tree. Using the out-degree distribution p_k of the network, we have the relation

$$G(a, x) = \sum_{k} p_k G^{(k)}(a, x). \tag{S2}$$

The meme on the root screen is the root of the cascade tree that results from the meme being tweeted and subsequently retweeted over a period of time, see Fig. S1. If we fix a time $t = \Omega$ as the observation time for the cascade sizes, then $G(a, x)$ is the PGF for the sizes (as observed at time Ω) of trees that are rooted at time $\Omega - a \equiv \tau$. We derive a relation between the PGF for the sizes of trees at age a (i.e., those rooted at time $\Omega - a$) and the PGF for tree sizes at age $a - \Delta t$ (i.e., those rooted at time $\Omega - a + \Delta t$), as follows.

Consider a meme on a given screen (call this screen S_1), at time $t = \tau :$ this is the root of the tree we call Tree(S_1), which has age a at the observation time $t = \Omega$; let k be the out-degree of the node with screen S_1. At the next time step $t = \tau + \Delta t$, there are four possible outcomes for this particular screen that contribute to the PGF $G^{(k)}(a, x)$, refer to Fig. S2:

- Outcome (a): the screen S_1 is overwritten by some other meme that is tweeted by another node. This terminates Tree(S_1)—setting the corresponding generating function to 1—as no future tweets can now result from the chosen root. Outcome (a) occurs with probability $z\Delta t$, since a node follows, on average, z other nodes, each of which is the tweeter with probability $1/N = \Delta t$. So outcome (a) contributes $z\Delta t$ (the probability of occurrence multiplied by the resulting generating function term) to the PGF $G^{(k)}(a, x)$. [Note that all terms of order $(\Delta t)^2$ and higher are ignored here and below, as these are negligible when we take the limit $\Delta t \to 0$.]

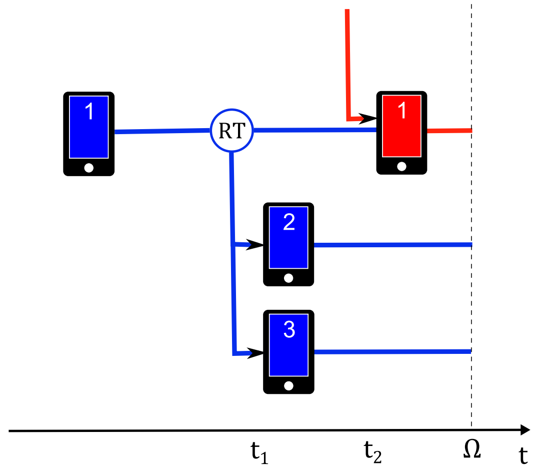
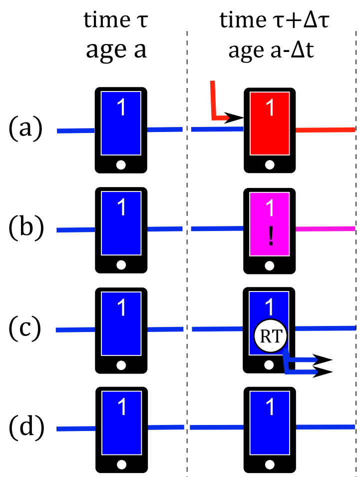


Figure S1: Schematic of the meme-diffusion model. Time runs horizontally and nodes of the network are listed vertically; the screen colour of each node indicates the meme it currently holds. At time t_1, node 1 retweets the blue meme to its followers (nodes 2 and 3). At time t_2, node 1's screen is overwritten by the red meme, which was tweeted by one of the nodes followed by node 1. At the observation time Ω, the data on meme retweets for all earlier times is gathered and the popularities of all memes are compared.



	Outcome for screen S_1	**Probability**	**Contribution to $G^{(k)}(a,x)$**
(a)	Overwritten	$z\,\Delta t$	1
(b)	Selected, innovates	$\mu\,\Delta t$	1
(c)	Selected, retweets	$(1-\mu)\,\Delta t$	$xG^{(k)}(a-\Delta t,x)[G(a-\Delta t,x)]^k$
(d)	Not selected, survives	$1-(z+1)\Delta t$	$G^{(k)}(a-\Delta t,x)$

Figure S2: Summary of the single-timestep outcomes that contribute to the PGF $G^{(k)}(a,x)$.

- Outcome (b): the screen S_1 is selected as the updater in the current time step (with probability Δt) and innovates (with probability μ), so terminating Tree(S_1). The contribution to $G^{(k)}(a,x)$ is thus $\mu\Delta t$.

- Outcome (c): the screen S_1 is selected for update (probability Δt) and retweets its meme (probability $1-\mu$). This (i) adds one to the size of Tree(S_1), whilst (ii) the branch on screen S_1 survives another time step, becoming the origin of a new tree rooted at $\tau+\Delta t$, which has age $a-\Delta t$ at the observation time. Moreover, (iii) screen S_1 is the parent of k new branches of Tree(S_1): each new branch acts as the root of a tree with PGF $G(a-\Delta t,x)$ (since the out-degrees of the newly-spawned roots are random). These effects (i)–(iii) lead to generating function contributions of x, $G^{(k)}(a-\Delta t,x)$, and $[G(a-\Delta t,x)]^k$, respectively, and since these occur simultaneously, the overall contribution of outcome (c) to $G^{(k)}(a,x)$ is $(1-\mu)\Delta t\, x\, G^{(k)}(a-\Delta t,x)[G(a-\Delta t,x)]^k$.

- Outcome (d): the survival of Tree(S_1), with none of the other outcomes (a)–(c) occurring: the probability of this is $1-(z\Delta t+\mu\Delta t+(1-\mu)\Delta t)=1-(z+1)\Delta t$, and the screen can then be considered as the origin of a new tree that is rooted at time $\tau+\Delta$, and so has age $a-\Delta$ at the observation time Ω. The overall contribution to $G^{(k)}(a,x)$ from this outcome is therefore $(1-(z+1)\Delta t)G^{(k)}(a-\Delta t,x)$.

Putting all four outcomes together gives an expression for $G^{(k)}(a,x)$, correct to first order in Δt:

$$G^{(k)}(a,x) = \underbrace{z\,\Delta t}_{\text{(a)}}+\underbrace{\mu\,\Delta t}_{\text{(b)}}+\underbrace{(1-\mu)\Delta t\, x\, G^{(k)}(a-\Delta t,x)\,[G(a-\Delta t,x)]^k}_{\text{(c)}}+\underbrace{(1-(z+1)\Delta t)G^{(k)}(a-\Delta t,x)}_{\text{(d)}}, \tag{S3}$$

and taking the limit $\Delta t\to 0$ yields an ordinary differential equation for $G^{(k)}(a,x)$, parameterized by x:

$$\frac{\partial G^{(k)}}{\partial a} = z+\mu-(z+1)G^{(k)}+(1-\mu)x\,G^{(k)}\,[G]^k. \tag{S4}$$

Averaging over the possible out-degrees of the root node—by multiplying by p_k and summing over all k—gives the following equation for $G(a,x)$:

$$\frac{\partial G}{\partial a} = z+\mu-(z+1)G+(1-\mu)x\sum_k p_k\,G^{(k)}\,[G]^k. \tag{S5}$$

Solving this equation for G requires also finding the functions $G^{(k)}$, for all k. However, if we make the following approximation

$$\sum_k p_k\left(G^{(k)}\,[G]^k\right)\approx\left(\sum_k p_k\,G^{(k)}\right)\left(\sum_k p_k\,[G]^k\right)=Gf(G), \tag{S6}$$

we obtain the single differential equation for $G(a,x)$ given by Eq. (3) of the main text. The simplifying moment-closure assumption (S6) will be examined in detail in further work; our numerical simulations indicate that it leads to quite accurate results for networks with reasonably large mean degree z (e.g. $z\approx 10$ as in Figs. 2 and 3 of the main text).

It is straightforward to generalize the derivation above to allow each node's screen to have capacity $c \geq 1$, meaning that the screen can simultaneously hold c memes: we then consider each screen to be composed of c distinct *screen-slots*. When retweeting, a node (user) chooses one of their c screen-slots at random to determine the meme that is transmitted to its followers; if the chosen screen-slot is empty then nothing happens. When a node innovates, or when it receives a meme from another node, the new meme is placed in a randomly-chosen screen-slot, overwriting any existing meme in that slot. An additional generalization is to allow for tweeted memes to be accepted onto followers' screens with probability $\lambda \leq 1$ (with $\lambda = 1$ giving the base case of the main text). Incorporating these generalizations into the derivation above leads to the following equation for $G(a, x)$:

$$c\frac{\partial G}{\partial a} = \lambda z + \mu - (\lambda z + 1)G + (1 - \mu)x\, G\, f(1 - \lambda + \lambda G), \tag{S7}$$

which reduces to Eq. (3) of the main text in the case $c = 1$ and $\lambda = 1$. We use this more general case throughout the Supplementary Material to demonstrate that the competition-induced criticality phenomenon is robust to changes in the model.

The mean popularity of age-a memes,

$$m(a) \equiv \sum_{n=1}^{\infty} n q_n(a) = 1 + \frac{\partial G}{\partial x}(a, 1), \tag{S8}$$

can be found found by differentiating Eq. (S7) with respect x to obtain the linear equation

$$c\frac{dm}{da} = 1 - \mu m - \mu\lambda z(m - 1), \tag{S9}$$

with $m(0) = 1$. The solution is

$$m(a) = \begin{cases} 1 + \frac{a}{c} & \text{if } \mu = 0 \\ \frac{1 + \lambda z \mu - (1-\mu) e^{-\frac{\mu(\lambda z + 1)}{c} a}}{\mu(\lambda z + 1)} & \text{if } \mu > 0. \end{cases}$$

Higher-order moments of the popularity distribution q_n can be obtained similarly by repeated differentiation of Eq. (S7). It is also possible to directly determine the probabilities $q_n(a)$ for small values of n, by expanding $G(a, x)$ as a Taylor series about $x = 0$. Setting $x = 0$ in Eq. (S7), for example, immediately yields a closed equation for q_1:

$$c\frac{dq_1}{da} = \lambda z + \mu - (\lambda z + 1)q_1, \tag{S10}$$

with solution

$$q_1(a) = \frac{\lambda z + \mu + (1 - \mu)e^{-\frac{\lambda z + 1}{c} a}}{\lambda z + 1}. \tag{S11}$$

Note that as $a \to \infty$, $q_1(a)$ approaches the value

$$\frac{\lambda z + \mu}{\lambda z + 1} = 1 - \frac{1 - \mu}{\lambda z + 1}. \tag{S12}$$

Thus, only a fraction $(1 - \mu)/(\lambda z + 1)$ of all memes grow in popularity: the majority of memes are forgotten before they are retweeted even once and so their popularity remains at $n = 1$ forever.

S2 Inverting PGFs using Fast Fourier Transforms

The probability $q_n(a)$ that a meme has popularity n at age a may be determined from the PGF $H(a,x)$ by repeated differentiation:

$$q_n(a) = \frac{1}{n!}\,\frac{d^n}{dx^n} H(a,x)\bigg|_{x=0}, \tag{S13}$$

where H is obtained as $xG(a,x)$ from the (numerical) solution of Eq. (S7). However, numerical differentiation is inaccurate for large values of n, so we invert the PGF using contour integration in the complex x-plane [1, 2]. The inversion integral is given by Cauchy's theorem

$$q_n(a) = \frac{1}{2\pi i}\oint_C H(a,x)x^{-(n+1)}\,dx, \tag{S14}$$

where all poles of $H(a,x)$ must lie outside the contour C; a common choice for C is the unit circle [2]. Writing $x = e^{i\theta}$ gives the form

$$q_n(a) = \frac{1}{2}\int_{-\pi}^{\pi} H(a, e^{i\theta})e^{-in\theta}d\theta, \tag{S15}$$

and numerical integration using the trapezoidal rule with M points yields the approximate formula

$$q_n(a) \approx \frac{1}{M}\sum_{m=0}^{M-1} H\left(a, e^{2\pi i m/M}\right) e^{-2\pi i n m/M}, \tag{S16}$$

which may be evaluated efficiently using standard FFT routines [1, 3]. Octave/Matlab code for implementing this inversion—and hence reproducing the theory curves of Figs. 2 and 3—is available from the authors upon request.

S3 Old-age asymptotics

The large-n asymptotic behaviour of $q_n(a)$ can be obtained in the limit $a \to \infty$ by asymptotic analysis of the solution of Eq. (S7): $G(\infty, x) = \lim_{a\to\infty} G(a,x)$. The following general result will prove useful (cf. Lemma 5.3.2 of Ref. [4]):

Lemma 1 *Let $\Phi(x) = \sum_{k=0}^{\infty} \pi_k x^k$ be the PGF for the distribution π_k, and suppose Φ has the following asymptotic series as $x \to 1$:*

$$\Phi(1-w) \sim \textit{analytic part} + \sum_{m=1}^{\infty} c_m w^{\beta_m} \qquad \textit{as } w \to 0, \tag{S17}$$

where $w = 1-x$ and $\beta_1 < \beta_2 < \ldots$ are positive, non-integer powers (note that the analytic part of Φ can be written as a power series in w with integer powers). Then the leading-order asymptotic behaviour of π_k is

$$\pi_k \sim \frac{c_1}{\Gamma(-\beta_1)} k^{-\beta_1 - 1} \qquad \textit{as } k \to \infty, \tag{S18}$$

where Γ is the Gamma function.

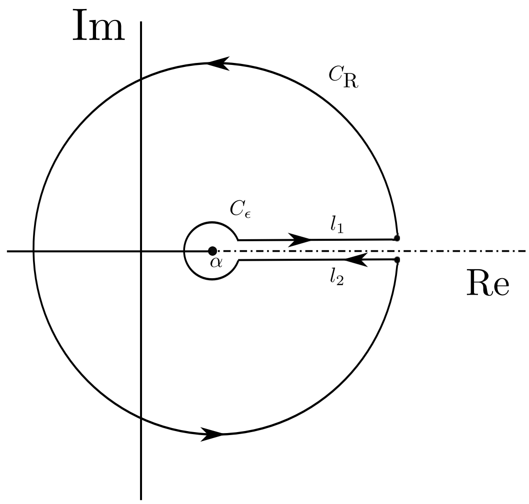


Figure S3: The contour C in the complex x-plane for the PGF inversion integral (S19). A branch cut extends from α to ∞.

To prove this result, we begin with the inversion integral

$$\pi_k = \frac{1}{2\pi i} \oint_C \Phi(x) x^{-k-1}\, dx, \tag{S19}$$

where the contour C can be deformed onto the contour $C_\epsilon \cup l_1 \cup C_R \cup l_2$ shown in Fig. S3. The point $\alpha = 1$ is a branch point, with the branch cut running from α to ∞. It is straightforward to show that the integrals along the circular contours C_ϵ and C_R limit to zero as $\epsilon \to 0$ and $R \to \infty$ [5], leaving

$$\pi_k = \frac{1}{2\pi i} \left[\int_{l_1} \Phi(x) x^{-k-1}\, dx + \int_{l_2} \Phi(x) x^{-k-1}\, dx \right] \tag{S20}$$

Along the rays l_1 and l_2 we make the substitution $x = e^\rho$ to obtain integrals whose asymptotic behaviour may be determined using Watson's Lemma [6]. The contributions from the analytic part of Φ to the l_1 integral and to the l_2 integral cancel each other. Along the branch cut, we write the leading-order non-analytic term of (S17) as

$$\begin{aligned}
c_1 w^{\beta_1} &= c_1 (1-x)^{\beta_1} \\
&= c_1 (1-e^\rho)^{\beta_1} \\
&\sim c_1 (-\rho)^{\beta_1} \qquad \text{as } \rho \to 0 \\
&= \begin{cases} c_1 e^{-\pi\beta_1 i} \rho^{\beta_1} & \text{as } \rho \to 0 \text{ along } l_1 \\ c_1 e^{\pi\beta_1 i} \rho^{\beta_1} & \text{as } \rho \to 0 \text{ along } l_2 \end{cases}
\end{aligned}$$

to obtain

$$\begin{aligned}
\pi_k &\sim \frac{1}{2\pi i} \left[\int_0^\infty c_1 e^{-\pi\beta_1 i} \rho^{\beta_1} e^{-k\rho} d\rho + \int_0^\infty c_1 e^{\pi\beta_1 i} \rho^{\beta_1} e^{-k\rho} d\rho \right] \qquad \text{as } k \to \infty \\
&= -\frac{c_1}{\pi} \sin(\pi\beta_1) \int_0^\infty \rho^{\beta_1} e^{-k\rho} d\rho. \qquad \text{(S21)}
\end{aligned}$$

The integral in this expression evaluates to $\Gamma(\beta_1 + 1)k^{-\beta_1 - 1}$. Using the Euler reflection formula $\pi/\sin(\pi\beta_1) = \Gamma(\beta_1)\Gamma(1 - \beta_1)$ and the Gamma function property $\Gamma(y + 1) = y\Gamma(y)$ completes the proof.

Turning now to the long-time (old-age) limit of Eq. (S7), $G(\infty, x)$ is the solution of

$$\lambda z + \mu - (\lambda z + 1)G + (1 - \mu)x\, G\, f(1 - \lambda + \lambda G) = 0. \tag{S22}$$

We seek an asymptotic series solution for $G(\infty, x)$ by inserting the expression

$$G(\infty, 1 - w) \sim \sum_m c_m w^{\beta_m} \tag{S23}$$

into Eq.(S22) to determine the successive values of β_m needed to balance the leading-order powers of $w = 1 - x$. Then we apply Lemma 1 to determine the asymptotic form of q_n for large n values.

- **Case 1: $f''(1)$ infinite, $\mu = 0$**
  An out-degree distribution with power-law tail $p_k \sim Dk^{-\gamma}$ with γ between 2 and 3 has a divergent second moment, so $f''(1)$ is infinite, and the behaviour of $f(x)$ near $x = 1$ is given by Lemma 1 as

  $$f(1 - w) \sim 1 - zw + D\Gamma(1 - \gamma)w^{\gamma - 1} \quad \text{as } w \to 0 \tag{S24}$$

  (recall $z = f'(1)$ is the mean degree). For $\mu = 0$, inserting expansion (S23) into Eq. (S22) then yields the leading-order asymptotic behaviour of G as

  $$G(\infty, 1 - w) \sim 1 - \frac{1}{\lambda}\left(D\Gamma(1 - \gamma)\right)^{-\frac{1}{\gamma-1}} w^{\frac{1}{\gamma-1}} \quad \text{as } w \to 0, \tag{S25}$$

  and the popularity distribution asymptotics follow from Lemma 1:

  $$q_n \sim Bn^{-\frac{\gamma}{\gamma-1}} \quad \text{as } n \to \infty, \tag{S26}$$

  with prefactor

  $$B = -\frac{\left(D\Gamma(1 - \gamma)\right)^{-\frac{1}{\gamma-1}}}{\lambda\Gamma\left(\frac{1}{1-\gamma}\right)}. \tag{S27}$$

  The network used in Fig. 3 of the main text, for example, has out-degree distribution $p_k = Dk^{-\gamma}$ for $k \geq 4$, with $p_k = 0$ for $k < 4$: the absence of degrees less than 4 ensures that the mean degree z is reasonably large ($z = 10.6$ for $\gamma = 2.5$), as assumed in the theory. The normalization constant D for this case is

  $$D = \frac{1}{\zeta(\gamma) - 1 - 2^{-\gamma} - 3^{-\gamma}}, \tag{S28}$$

  where ζ is the Riemann zeta function, allowing the prefactor B to be calculated explicitly from (S27).

- **Case 2: $f''(1)$ infinite, $0 < \mu < 1$**
  When f is given by Eq. (S24) and $\mu > 0$, the leading-order behaviour of G is found to have the form

$$G(\infty, 1-w) \sim 1 - \frac{1-\mu}{\mu(\lambda z+1)} w + D\Gamma(1-\gamma)\lambda^{\gamma-1} \left[\frac{1-\mu}{\mu(\lambda z+1)}\right]^{\gamma} w^{\gamma-1} \quad \text{as } w \to 0 \quad \text{(S29)}$$

  and the first non-analytic term gives the asymptotic form of the popularity distribution from Lemma 1:

$$q_n \sim C n^{-\gamma} \quad \text{as } n \to \infty. \quad \text{(S30)}$$

  The prefactor of the popularity distribution power-law is explicitly related to the prefactor of the degree distribution by

$$C = \frac{D}{\lambda} \left(\frac{\lambda(1-\mu)}{\mu(\lambda z+1)}\right)^{\gamma}. \quad \text{(S31)}$$

- **Case 3: $f''(1)$ finite, $\mu = 0$**
  If $f''(1)$ is finite (so the network out-degree distribution p_k has finite second moment) and $\mu = 0$, we expand $f(1-\lambda+\lambda G)$ to second-order terms in the small parameter $\phi = 1 - G$, and find the leading-order asymptotics of $G(\infty, 1-w)$ to be

$$G(\infty, 1-w) \sim 1 - \left(\lambda z + \frac{1}{2}\lambda^2 f''(1)\right)^{-\frac{1}{2}} w^{\frac{1}{2}} \quad \text{as } w \to 0. \quad \text{(S32)}$$

  Using Lemma 1, we conclude that

$$q_n(\infty) \sim A\, n^{-\frac{3}{2}} \quad \text{as } n \to \infty, \quad \text{(S33)}$$

  where the prefactor is

$$A = \left[2\pi\lambda\left(\lambda f''(1) + 2z\right)\right]^{-\frac{1}{2}}. \quad \text{(S34)}$$

- **Case 4: $f''(1)$ finite, $0 < \mu \ll 1$**
  Finally, we consider the case where $\mu > 0$ but $f''(1)$ is finite. In this case, $G(\infty, x)$ is analytic at $x = 1$, indicating that the popularity distribution does not have a power-law tail, i.e., Lemma 1 does not apply. However, if we write $G(\infty, 1-w) = 1 - \phi(w)$ with $|\phi| \ll 1$ and expand Eq. (S22) for w near 0, retaining terms of orders w, ϕ and ϕ^2, but neglecting terms of order $w\phi$, we obtain a quadratic equation for ϕ with solution

$$\phi(w) = \frac{-\mu(\lambda z+1) + \sqrt{\mu^2(\lambda z+1)^2 + 2\lambda(1-\mu)^2\left(\lambda f''(1) + 2z\right) w}}{(1-\mu)\lambda\left(\lambda f''(1)+2z\right)}. \quad \text{(S35)}$$

  Note this solution scales as $\phi = O\left(w^{\frac{1}{2}}\right)$ as $w \to 0$, which is consistent with the choice of retained terms in the scaling analysis of Eq. (S22). The solution (S35) has a branch point in the complex x-plane at (recall $w = 1 - x$):

$$\alpha = 1 + \frac{\mu^2(\lambda z+1)^2}{2\lambda(1-\mu)^2\lambda\left(\lambda f''(1)+2z\right)} > 1. \quad \text{(S36)}$$

Integrating along the branch cut in the complex x-plane (in a very similar fashion to the proof of Lemma 1) enables us to find the large-n asymptotic form of the popularity distribution as

$$q_n(\infty) \sim A\, n^{-\frac{3}{2}} e^{-\frac{n}{\kappa}} \quad \text{as } n \to \infty \tag{S37}$$

with A given by Eq. (S34) and the large-n cutoff by

$$\kappa = \frac{2\lambda(\lambda f''(1) + 2z)}{\mu^2(\lambda z + 1)^2}, \tag{S38}$$

and where we have assumed $\mu \ll 1$ to simplify the results. When $\lambda = 1$, Eq. (S37) reduces to Eq. (4) of the main text.

References

[1] J. K. Cavers. On the Fast Fourier Transform inversion of probability generating functions. *IMA Journal of Applied Mathematics*, 22(3):275–282, 1978.

[2] 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.

[3] M. Marder. Dynamics of epidemics on random networks. *Physical Review E*, 75(6):066103, 2007.

[4] H. S. Wilf. *generatingfunctionology.* A K Peters/CRC Press, 2005.

[5] J. P. Keener. *Principles of applied mathematics: transformation and approximation.* Addison Wesley, Advanced Book Program, 1988.

[6] C. M. Bender and S. A. Orszag. *Advanced mathematical methods for scientists and engineers I: Asymptotic methods and perturbation theory*, volume 1. Springer Verlag, 1999.